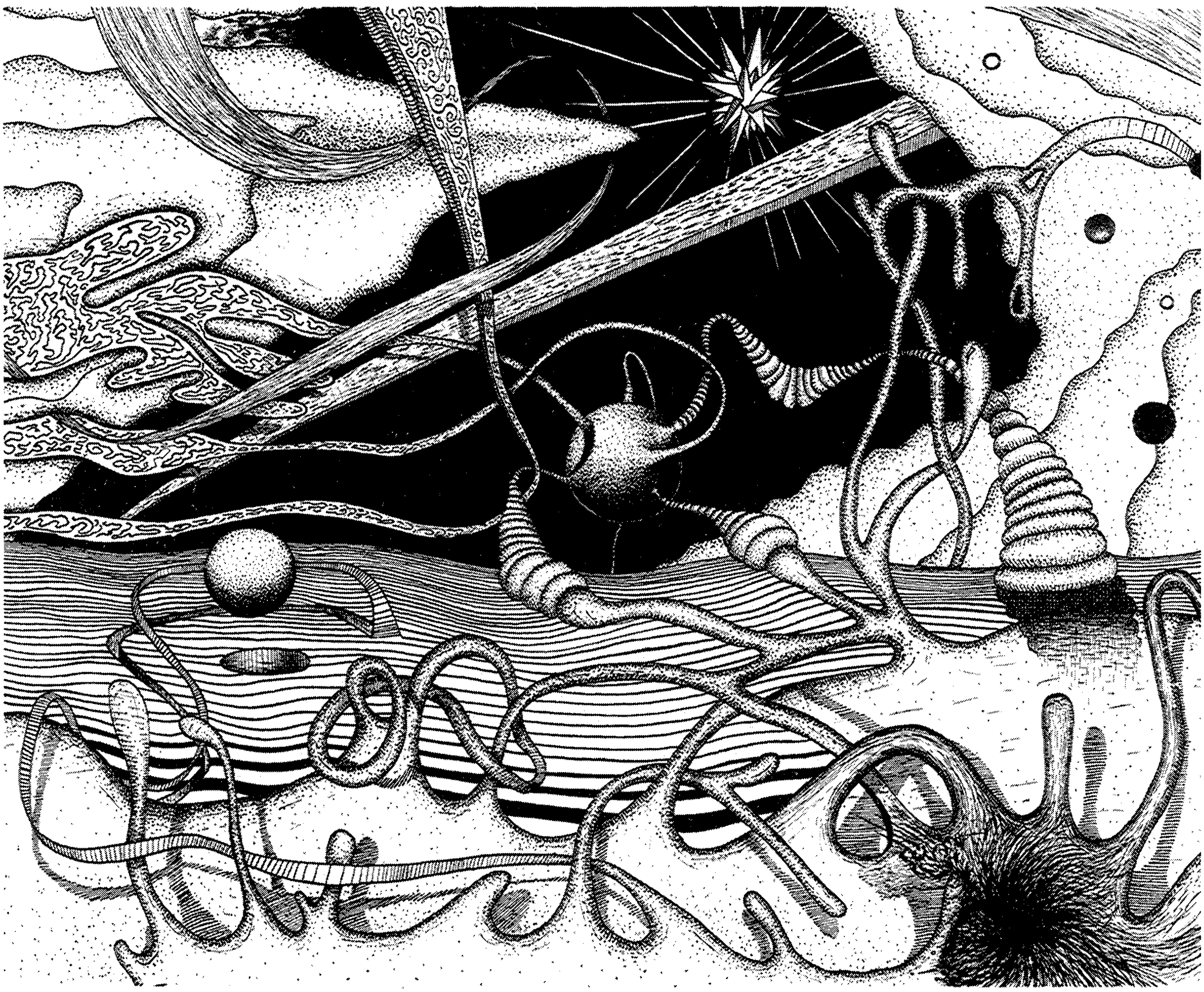
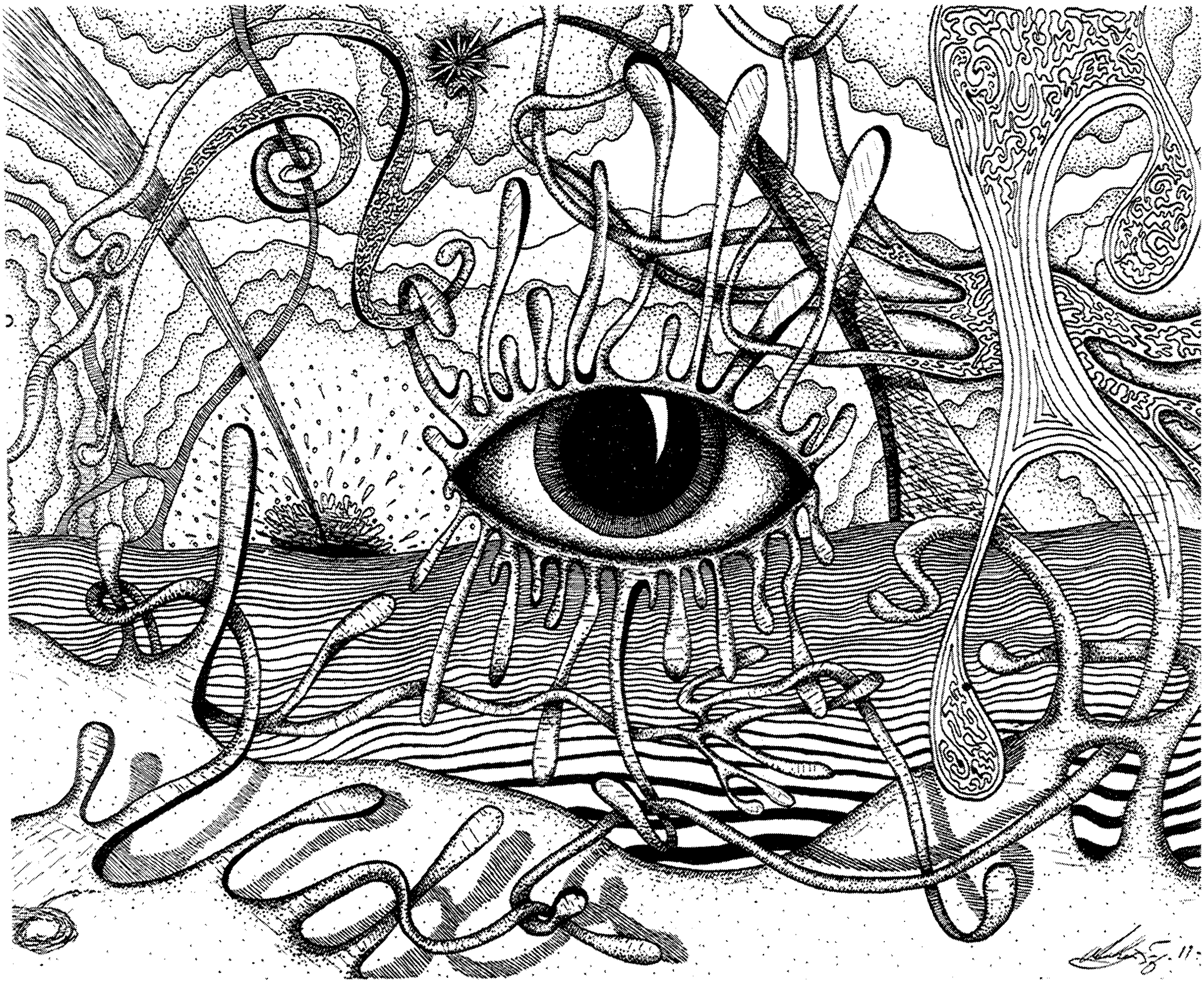




BEER

MANAGING MODERN COMPLEXITY

Stafford Beer [1]

I. Threat Systems

I speak to you today against a background of seemingly ungovernable crisis which it is impossible to ignore. We are met to discuss the future, but we must know the context from which we begin.

The business of forecasting is fraught with many traps; it often seems ascientific. But the perspicuous detection of inexorable trends can be a matter of good science. There is a reality to observe and to measure, a reality in which a dead man is a corpse and not a statistic. There is a reality, too, with which to experiment; a reality that does not come in parcels labelled for the attention of appropriate officials. The very stuff of this reality is *complexity*. The elements of our society ever more richly interact: the more this happens, the more participation is invoked, the more the streams of data flow . . . the more complex does society become.

Handling complexity seems to be the major problem of the age, in the way that handling material substance offered challenge to our forefathers. Computers are the tools we have to use, and their effective use must be directed by a science competent to handle the organization of large, complex, probabilistic systems. This is the science of cybernetics, the science of communications and control.

The central thesis of cybernetics might be expressed thus: that there are natural laws governing the behaviour of large interactive systems—in the flesh, in the metal, in the social and economic fabric. These laws have to do with self-regulation and self-organization. They constitute the "management principle" by which systems grow and are stable, learn and adjust, adapt and evolve. These seemingly diverse systems are one, in cybernetic eyes, because they manifest viable behaviour—which is to say behaviour conducive to survival.

In my opinion, the most important fact which a quarter of a century's worth of cybernetics has revealed is that this behaviour is governed by the dynamic structure of the system, rather than by special events occurring within it or by the particular values taken up by even its major variables. "Structure" means the way in which the parts of a whole are inter-related; and here it includes both the feedback loops by which systems regulate themselves and also the conditional probability mechanisms by which systems learn and organize themselves. "Dynamic" relates to the speeds at which communication is effected within the system, and especially to the relative lags with which messages are promulgated, overtake each other, and combine to form new patterns. Dynamic structure generates outcomes.

Therefore I say that what will happen to mankind in its battle with complexity will be determined neither by particular innovation nor by isolated achievement at some unknown future date. Hence the attempted prediction of such things is not to the point. Outcomes are latent in the dynamic structure of the systems we have or may adopt: they will inexorably emerge.

At present, the most obtrusive outcome of the system we have is a gross instability of institutional relationships and of the economy. This cannot last. The society we have known will either collapse, or it will be overthrown. In either case a new kind of society will emerge, with new modes of control; and the risk is that it will be a society which no one actually chose, and which we probably will not like. I shall argue that we must use our science to detect the latent outcomes which will one day characterize the future of mankind. And let us so engineer our systems that their latent outcomes suit our social purpose. It is true that the outcomes cannot be fully determined, because there is noise (or shall we call it free will?) in the system. But a systemic design taking due account of cybernetic laws may be expected to produce behaviour which is predictable in terms of the overriding social need for stability.

Thanks to the growth of complexity, which is very much a function of the growth in data-handling capacity and of the information explosion, society has outgrown the dynamic regulating capacity of its own hallowed structure. History did not design that structure to cope with such complexity, and a cybernetically grotesque machinery is a result. It is from this standpoint that I ask you to look again at the environmental crises from which our view of the future must necessarily start.

The thermonuclear threat is a computable threat, and one which computably grows—although we act as if we were inured to it. The various pollution threats—by pesticides, by noise, by sewage, by carcinogenic urban air—were and remain systemically predictable. None of these things happened by chance, by accident, or by the wrath of God. We have run ourselves into these problems by failing to calculate the predictable consequences of the systems civilization has underwritten. The same seems to me to be true, though less obviously so, of the various forms of societary crisis which run alongside the environmental crises. Problems of race, problems of poverty, problems of overpopulation: all these are quantifiable aspects of computable systems. It has taken social upheaval and threatening violence to draw them to our proper attention; it has taken a major revolt of the young to motivate any kind of rethinking.

The risk which faces us today is the probability that society will yet refuse to study the systemic generators of human doom, and will disregard the cybernetic capability which already exists competent to bring these many but inter-related forms of crisis under governance.

There are two reasons for this fear. First of all, our culture does not take kindly to the notion that it nurtures the seeds of its own destruction. Instead of studying the systemic reality in which outcomes are latent, it prefers the technique of prognostication. Small wonder: by using such wholly non-systemic devices as the Delphi technique, we may predict a possible millennium for our comfort. But the Delphi technique is aptly named: its pronouncements are shrouded in ambiguity—because they take no account of the systemic context. Meanwhile, the systems we have already started, which we nourish and

---

[1] Development Director, International Publishing Corporation, and visiting Professor of Cybernetics in the Business School of Manchester University, Great Britain.

foster, are grinding society to powder. It might sound macabre to suggest that computers will finish the job of turning this planet into a paradise after human life has been extinguished. But that vision is little more macabre than the situation we already have, when we sit in the comfort of affluent homes and cause satellites to transmit to us live pictures of children starving to death and human beings being blown to pieces.

The second reason for my pessimism is that technology now seems to be leading humanity by the nose. We appear to have no sense of priorities where our problems are concerned; we do what is technologically easy—and we do it regardless of cost. For example, the problem people have of transporting themselves from one remote place to another really exists between homes or offices and international airports. But the problem we continuously solve is the non-existent problem of moving between those airports. It is easier to go from Mach 1 to Mach 2 than to tackle the genuine problem. Perhaps it was also easier to go to the moon than to face up to what is happening in the street outside.

Thus I direct myself and you to the claim that cybernation is about the regulation of society, and that this is what computers are for. Perhaps this opening is a surprise. Would it not have been *easier* for all of us to plunge into the technology of computation, to prattle on happily about nanoseconds and massive data banks, to wonder at the explosion of knowledge and the impending marvels of data storage and retrieval by holograms and photochromic tubes, rather than to tell the truth about cybernation? What did you really expect? The fact is that most of the problems we stand ready to consider are bogus problems. They are generated by theories about technological progress, and theories about the way society works. Theory is often the only reality countenanced by our culture.

The reality is that we are elements in a vast and almost ungovernable social system generating outcomes that happen to us. We come sprightly to conference, dragging lead-heavy bones, to talk about machines that matter only if they can help us men. Our fat is suffused with insecticide, but we are avid to decide what it will be like to take our newspaper out of the back of the television set. The expansion of knowledge will yet save the world, shall I not tell you, coughing through the carcinogens—and assuming that my plane was not hijacked and that I was not "mugged" on the way * * *.

I am fighting for a way through to your real ears. That is exactly to say that I am trying to differentiate, in you, between data and information. Data are a whole lot of meaningful patterns. We can generate data indefinitely; we can exchange data forever; we can store data, retrieve data, and file them away. All this is great fun, maybe useful, maybe lucrative. But we have to ask why. The purpose is regulation. And that means translating data into information. *Information is what changes us.* My purpose too is to effect change—to impart information, not data.

Data, I want to say to you, are an excrescence. Data are the very latest kind of pollution. We are not going to do anything at all about the management of information and knowledge towards the regulation of society as long as we think in data-processing terms. That is technologically easy. It is what the computer companies and the telecommunication interests would like us to do. Data are assuredly the great new marketable commodities of the nineteen-seventies. But, let me repeat, data of themselves have no value.

What has value is the machinery to transform data into information, and the machinery by which that information may be used to innervate society. Society has become a complex organism, and it needs a nervous system. Managing the development of informational science and technology is all about this task. There is no other message than this.

2. Bases of Argument

The technological capabilities on the availability of which my arguments will be based already exist. There is not really a significant element of prognosis about them. There is however one proviso to this: it derives from a logical trap to which I will shortly draw attention. But first, here are some fundamental propositions.

First Proposition.—We can now automate whatever we can exactly specify.

Second Proposition.—Most (possibly all) ostensibly human prerogatives for inferential, judgmental, learned and adaptive behaviour can be exactly specified—at least with respect to finite contexts.

To extend the second proposition to intuitional and creative behaviour poses grave difficulties of definition, and invariably invites emotional uproar. But we may at least stand by this weaker statement.

Third Proposition.—Within specified frameworks, much ostensibly intuitional and creative human behaviour can be indistinguishably imitated by machine.

Fourth Proposition.—Distance is technically irrelevant.

All this means that purposive systems can now be created to undertake any kind of purpose at all. We know how to design those systems, and how to innervate them with data streams. And so society would appear to be confronted by a problem of choice: what activities should actually be automated? But I shall argue that this question is largely illusory.

First of all, there is the logical trap. This is of the sort called by logicians a fallacy of addition. We may do *any* of the things we can do; it does not follow that we may do *all* the things we can do. In the present state of the art, that is to say, we shall rapidly exhaust our reserves of skill. So here is the proviso about technological capability. My own belief is that we shall have to embody a great deal of basic software in special purpose hardware, and that we shall need to automate the creation of special software itself. I think that computer science will break through the barrier of human programing, and move to an era when programs are written by machines under general human surveillance. This will in turn lead to programs which modify themselves in the light of experience. Then we shall be near the realization of the machine being more intelligent than its designer, which von Neumann envisaged and showed mathematically possible more than twenty years ago. There is no need for more than this one paragraph of such modest guesswork—because after *that* it may well be too late to do what ought to be done right now. At any rate, this is the only technological barrier which I can identify.

Then we revert to the spurious problem of choice. Why should not responsible authorities choose between desirable and undesirable systems for handling knowledge and information? The answer is that in neither the private nor the public sector of a free society is there a sufficient concentration of power to do so. If, for example, mammoth publishing interests decide (as they may) not to mobilize the resources of electronics adequately in the dissemination of knowledge, then it is open to electronic interests to become the publishers of the future. It is also open to the information handling community itself to embark on entrepreneurial activity at the expense of both these industries. In the public sector, it is certainly open to central government, through its grant-awarding agencies in particular, to encourage or discourage particular applications of cybernation. But it will be very difficult to inhibit developments which are of themselves economically viable in the way that (for example) space exploration would be inhibited without central funding.

And here we perhaps identify the basic nature of the problem which cybernetic systems set out to solve. Throughout history until this time the problem was to acquire sufficient information to generate effective change. The individual wishing to become expert in some field of knowledge had to buy information expensively; the government wishing to understand even the rudiments of the structure of its society had to buy information through the census. And so we have gone on, paying more and more money for data acquisition—on the assumption that data constitute information. But we have already said that data become information only at the point when we ourselves are changed. It is self-evident that our capacity to be changed, whether we are an individual seeking private knowledge or a government seeking understanding of society, is strictly finite. In conditions of data paucity, almost all data acquired can be transformed into information—and used to procure effective change. But in conditions when the supply of data far outruns this metabolic capability, most data are literally worthless. Yet we pay more and more for these worthless data because that is the established order of things.

The fact is that quite recently the sign of the informational problem changed from plus to minus. The problem is no longer about acquiring data, which are generated as a by-product of every modern undertaking. The problem is about informational overload. The private citizen seeking knowledge is inundated by information which is virtually free. Yet the publishing industry responds in the old mode—by selling him yet more. The firm continues to buy expensive market research, because that is what it has always done, oblivious of the fact that transactions of every kind can now be electronically monitored, so that data are in glut. Its problem too is one of procuring adaptive behaviour, and no longer at all one of "finding the facts". As for government, there is really no dearth of societary information either; there is instead a problem of organizing information—across departmental boundaries and in time.

Institutions, firms and (thanks to television) private citizens today receive critical information very quickly indeed; the aggregate picture at federal level is slow by comparison to materialize. To put the point the other way round, then, the body politic has wildly overactive reflexes. In the body physiologic this is the condition of *clonus*—it is a symptom of spasticity. If we live, as I suspect, in a spastic society it is because of clonic response. And by the expectations of these arguments, the clonus will get worse.

Thus I argue that the problem of information management is now a problem of filtering and refining a massive overload—for all of us, whether citizens, firms, institutions or governments. We might well say that it is a problem not so much of data acquisition as of right storage; not so much of storage as of fast retrieval; not so much of retrieval as of proper selection; not so much of selection as of identifying wants; not so much of knowing wants as of recognizing needs—and the needs are precisely the requirements of systemic equilibria.

This almost tabular account of the matter ostensively defines another cybernetic truth. In any controlled system, there must be an hierarchic array of sub-systems, in which both the values and the structure of any one sub-system are set by a logically superior system. That is to say that one cannot discuss the purposive nature of a system in its own language, but only in a higher order language. There are potent reasons for this in theoretical logic, just as there are potent practical issues in terms of the need systematically to reduce the informational overload by a system of filters. These filters are necessarily arranged hierarchically, in a way which matches the hierarchy of logical systems.

Thus I introduce the concept of *meta*system: a system which stands over and beyond a logically inferior system, and one which is competent to handle that lower system's logic. Please note that metasystems are logically superior, and not necessarily more senior or more highly endowed with status or privilege. Please note also that in an hierarchy of systems there will be several orders of "metal". Let us take a moment of time to illustrate these points, since the concept of metasystem plays an important role in what I have to say.

Consider for example a school, in which each of a hundred teachers adequately controls and instructs a roomful of pupils. The roomful is in each case made up of several sets of pupils. Now each set of pupils is in fact pursuing a course of instruction which takes it from one room, one association of sets, and one teacher, to another room, another set of sets, and another teacher. If we consider the totality of rooms, holding their pupils and teachers, as sub-systems of the school (for this is indeed the organizational format we observe on a visit) there is no way of knowing or discussing *in such terms* the educative process as it affects all the pupils. To do this we shall need to find the metasystem which organizes all the groupings and ensures that they mesh together. This metasystem is the timetable, in terms of which the course followed by a particular pupil stands revealed. This is a logically superior system; but we do not expect the teacher in his room to treat the timetable as some kind of ju-ju. On the contrary; but if he wishes the timetable altered, he will perforce raise the issue in metasystemic terms. It is simply no good to say "this is my class, and I will take it at another time."

Furthermore: if the state wished to discuss the total process of education for all its high schools in relation to nursery schooling on the one hand and to university education on the other, then a new metasystem logically beyond the first metasystem would be required. And in this case the question whether the second metasystem is not only logically but also constitutionally superior would arise. It would be discussed in those familiar terms about autonomy, about professional integrity, about bureaucratic interference, about sub-optimization, about synergy. . . . Such discussions would be less boring if we could get the logic right first.

The Esoteric Box

Let us now retrieve the argument that the development of purposive automated systems involves a spurious problem of choice. For, we argued, there is no method in a free society whereby such choice could be implemented. I would like to examine this argument in more detail, with a view to uncovering certain mechanisms which are germane to the issue before us. The objective now is to try, like good scientists, to determine the basic parameters of the problem at some level of abstraction which facilitates understanding. Were we to fail in this endeavour to stand back and to generalize, we should conclude with long lists of possible systems, in hundreds of possible contexts, with long lists of possible dangers attaching to each. Then we should achieve no useful insights at all.

Firstly, what is the entity which will in practice develop systems of knowledge and information? It is some kind of social institution: perhaps a firm, perhaps a profession, perhaps a social service * * *. Whatever it is, it is surely an identifiable entity, with certain recognizable characteristics. I call it an *esoteric box*. What is going on inside this box is an established order of things: things accepted as mores of the box, things professional, things historical, and so on. There is a complex arrangement of sub-systems, a strange set of relationships between people of standing inside the box, and a recondite way of behaving. These features—their complexity and unintelligibility to the outsider—justify the box's adjective "esoteric". Admission to the box's activity cannot be gained without the appropriate passport. But the box is not a closed system, it is part of society; it certainly has inputs and outputs. Even so it is internally and autonomously self-organizing and self-regulating. And although the box *processes* whatever it exists to affect (and this is often people), that which is processed does not change the box at all. The box goes on: it is very powerfully organized to maintain its own internal stability, and therefore its survival as an integral institution.

I have elsewhere sought to show that the esoteric box, the identifiable social institution, is a strongly robust system in equilibrium. If we try to influence its behaviour by changing variables which apparently affect it, it responds neither by collapsing nor by a violent reaction. It simply shifts the internal position of equilibrium very slightly, thereby offsetting the environmental change that has occurred. (In the model from physical chemistry that I have used to study these boxes, this behaviour would be an instance of the operation of Le Chatelier's principle.)

Now if it is an esoteric box which is going to develop an information system directed to cybernetic ends, its primary objective will be to enhance its own performance and chance of survival—it will not attend first to the performance and survival of society at large. Equally, the box will be highly resistant to efforts made to constrain its freedom to do so. There seem to be only two mechanisms available to a free society seeking to influence an autonomous institution in any case. The first is to facilitate some modes of development and to inhibit others by the provision of incentives and inhibitors from outside. I mean by this the awarding or withholding of grants, tax concessions, public campaigns, and so on. Every esoteric box has its own feedback mechanisms: what the state can do is to change the gain on the relevant amplifiers. But because of the high internal stability of the box, we must expect this kind of control device to operate in a cumbersome and generally inefficient way. The other device available is legislative. The main trouble here lies in the identification of what is antisocial. Most advances in human welfare have paid a price in the infringement of personal liberty: whether that price is seen as reasonable or as a fundamental deprivation of human rights will often be a matter of interpretation. But I shall in any case assume that wise government will interact with the authorities in any esoteric box to achieve acceptable codes of behaviour. What really concerns us in this situation is what happens at the metasystemic level.

The fact is that esoteric boxes interact. Any major facet of public policy, such as health, education, the manipulation of credit, security, and balance of payments and so forth, involves at least a string and possibly a complex network of interacting esoteric boxes. Now just as the esoteric box itself is seen as something extremely stable and survival-worthy, so the system which links the boxes is typically tenuous and unstable. It is not itself an institution, not itself a higher order esoteric box. It is simply an assemblage of esoteric boxes, and it does not constitute a proper metasystem at all. It is in this fact that the threat to society really lies; it is here that we shall seek the important scientific generalizations.

Consider education, for example. There are, to speak arbitrarily, four major esoteric boxes involved in this facet of society. There is the system of compulsory schooling; there is the university system; there is the post-experience career-oriented system sponsored by industry; there is the free market in adult education. All four of these esoteric boxes may be sub-divided, almost endlessly; but we are seeking to move our thoughts in the opposite direction—to identify the commonality of these systems and to examine their interactions. If we take health as our example, we shall find a similar situation. There is an esoteric box labelled general medical practice, and another called hospitals; there is a public health box labelled sanitation; there is a market-oriented box dealing in pharmaceuticals; there is a market-place for medical information which belongs to publishing.

In short, we may take any facet of social policy and find the strings and networks of highly stable esoteric boxes which between them make a composite but not integrated impact on the individual citizen. We may do this for security, discovering esoteric boxes for the police, esoteric boxes for fire protection, and esoteric boxes for insurance—not to mention the esoteric boxes which are the armed services themselves. We may do the same thing for the movement of goods, discovering esoteric boxes for every method of transport. We may do it for the movement of money, detecting esoteric boxes for emolument and social benefit, for taxation, for credit * * *.

Then the question arises, why are those strings and networks as unstable as they appear to be? If there is no genuine metasystem, why has one not grown up? Was there never a stabilizing structure of any kind? I think that there was a metasystemic structure of a very remarkable kind, but that it has been abandoned. We have thereby lost the meta-controls which made the composite systems of esoteric boxes viable. If this be true, no wonder we need assiduously to design replacements.

First, there was the structure of society's "external skeleton": the religious, legal and moral framework. Into this hooked the structure of the "internal skeleton": there were indeed formal bonds linking social institutions themselves. Younger people seem to be systematically abandoning the values of the external system, so that it ceases to be relevant to any control process dependent on negative feedback. Given that almost fifty percent of the population of the United States is now under twenty-five years of age, the revolt of youth is destroying metasystems whose stabilizing value they do not understand is a serious matter indeed. The young have more power in society than ever before: purchasing power, and the power that derives from not being afraid of inherited norms. Most of them are not taking technology for granted. Many of them are questioning established values in terms which their elders do not understand. Some have already begun smashing up computer installations. As to the internal system, changes in technology are moving the interfaces between the esoteric boxes representing established institutions—and they are not responding. Instead of evolving by adaptation, these boxes are putting up the shutters and seeking to maintain themselves as integral systems while the context changes around them. This will not work.

Thus the strings and networks are unstable, and the metasystems are missing. Rather than attempt the exhaustive enumeration of these composite systems let us try to state the features they share in terms of knowledge, information and control. They seem to me to be the following:

Characteristics of Strings and Networks of Esoteric Boxes

(i) In all cases some esoteric boxes in the system are part of the public sector and some part of the private sector.

(ii) In all cases the esoteric boxes are generating, and (inefficiently) passing between themselves, knowledge about the world in which they operate.

(iii) In all cases they are also generating, either as primary or as spin-off data, knowledge about the individual citizen which they rarely interchange.

(iv) In all cases the very forces which produce stability within the esoteric boxes themselves conduce to instability between the boxes.

(v) In all cases, what constitutes the improved management of knowledge *within* the esoteric box has to do with the rapid matching of sets of possible courses to sets of actual conditions, and the rapid correction of mismatches by feedback governors.

(vi) In all cases what would count as an improvement in the management of information *between* esoteric boxes, and therefore an embodiment of the metasystem concerned, would be an integral information network and a mutual trade-off in knowledge—both of the world and of the citizen.

If this list of six points correctly states the position, it behooves us to elucidate them further.

3. Elucidation of Systemic Characteristics

We begin this elucidation by developing a generalization about the management of information within the esoteric box. This is an explanation of point (v) in the foregoing *List*.

Whatever we are looking at at any given moment in time will be found to represent a complex state of affairs Call this total situation the *initial condition*. For example, a patient entering a health system has an initial condition; so has a pupil in any educational situation. The first step taken by a professional in reviewing this initial condition is to try and characterize it with a name. In the case of health, this name is the diagnosis (*diabetis*="he needs more insulin than he has got"). In the case of the educational condition, we may name a state of ignorance relative to some need (*advanced physics*="he needs more physics than he has got"). This naming process may be very inefficient, as for instance when we name the complicated economic status of a citizen within the economy as: *credit*=$100. And even in medical diagnosis, for instance in psychiatric medicine or in prophylactic medicine, the name may not be very much help. Then why do we go through the naming process?

The answer to this is surely that the brain is a coding device. We are not cerebrally organized to hold in our heads large wedges of information about complicated states of affairs. Having examined the complexity of the initial condition, we seek to encapsulate it in a name—which can later be used to retrieve at least the critical attributes of the situation so named. Next, we use this name in our search of courses of action from which to select a treatment of the initial condition. Thus the very mention of a medical diagnosis selects in the mind of the physician a subset of the whole set of human therapies which relates to the name, and from this subset one therapy will be selected and applied. Similarly, "advanced physics" selects a subset of courses from all possible education courses, and from that subset one course will be recommended. The credit rating simply selects one figure from a small number of possible figures to be applied as a ceiling on purchasing power.

Depending on the seriousness of the situation, as measured perhaps by its "professional" content, this naming filter is a more or less elaborate tool for making the system work. A higher professional content can be injected into the process by a more elaborate taxonomy of names, and also by iterating the process of selection. Thus, having made a diagnosis and selected a possible therapy, the physician will go back through the name filter to the actual initial condition, and verify the treatment in every particular. In most social situations, however, this iteration is far too expensive to undertake. And for that reason, many of the responses which social systems make to the initial condition are crude indeed.

The first general capability of automation within such a system is to abandon the naming filter. For computers *can* hold large wedges of information. The computer is faced with the problem of matching one complex profile (the initial condition) with another—probably less complex—profile of possible courses of action. Far from simply automating the human professional component in the system, then, the automatic system should much improve upon it, especially if it is organized to interrogate the subject in order to fill out details of the initial condition which it perceives to be relevant. Moreover, as its model of the system it handles is enriched and improved by experience, it becomes possible in principle for a preliminary choice of action to be iteratively simulated. Then the likely effects of choosing this action, and in particular the vulnerability of this strategy to unknown factors or a range of possible futures, may rapidly be estimated before any indication of choice is given at all. Next again, if the automated system is geared to invigilate the actual process of applying the course to the initial condition, so that the subject's response is continuously monitored, then corrective action against any mismatching or systemic oscillation may be continuously taken. And of course it will be taken on the basis of the total richness of possible interaction between the two sets (states of the subject and possible treatments) rather than through the exiguous filtering channels of the naming which have hitherto been used with so little finesse.

In all of this we find key applications of another fundamental cybernetic principle: Ashby's Law of Requisite Variety. Variety is the cybernetic measure of complexity. It is explicitly the possible number of states of a system. The Law says that the variety of a given situation can be managed adequately only by control mechanisms having at least as great a capacity to generate variety themselves. Names typically do not do this: they are archetypes of variety reducers. Indeed, in most socio-economic situations of our age, we seek to obey Ashby's law by *reducing* the variety of the real world, necessarily in a somewhat artificial way, as with naming. As I said earlier, this leads us to manage low-variety theories about the economy, because we can handle those, rather than to manage the high-variety economy itself. A much more satisfactory method of handling the problem is to *increase* the variety of the system doing the judging, managing or controlling—by automating the "professional" component. The second method is now technologically open as we saw in the last section. Allied to fast feedback, whether through simulations of the total system or through the invigilation of actual results, the whole mechanism permits a much more refined and much speedier convergence on a stable outcome.

By looking at this mechanism in its relevant detail, we simultaneously lay bare the major threat to privacy of which everyone who has ever contemplated these matters is already aware. As we seek better control of situations by confronting variety with variety within the system, we lose the anonymity which used to cloak the identity of an individual by the use of a name. This is quite clearly seen in the simplest case of all—the name of the citizen as normally understood. My name identifies me from among the rest of us here; but it undertakes to disclose no more information than this primary selection. Yet the more effectively any esoteric box handles my case, then the higher the variety it disposes as a measure of my own variety; therefore the more risky to my personal integrity does the whole process become. Here is the person rawly exposed. Because in higher variety, within the professional system appropriate to any esoteric box, I am saying that the better the system, both from the point of view of the social institution concerned and therefore from my own as its patient or pupil or client in any other way, *ipso facto* the more potentially damaging to me is that system. Am I psychologically ill? The medical system will know. Am I educationally inadequate to my job? The educational system will know. Where was I at the time the murder was committed? The credit system knows when and where I bought petrol that night * * *.

This analysis successfully generalizes the problem of privacy, and also says a great deal about the reasons why esoteric boxes are under such pressure to withdraw into themselves—instead of collaborating in metasystemic management systems (see Point (vi) in the *List*).

As to privacy: It is all too possible that the computer will sweep forward to destroy privacy and freedom of choice without our really knowing that this is happening—much as the motor car has swept forward, poisoning us and inexorably changing the quality of life. Consider two major mechanisms which might bring this about.

First, there is the question of a man's credibility as a citizen. When a man is too well documented, electronically buttoned-up, in what sense can he make a new start? How can he restore his credit, once it is lost? How will he persuade the machine to emulate his own God-given capability to *forget*? A man is to himself as to others a complex package of information. In behavioural terms at least, his vital statistics, his knowledge, his actions and his emotional response as well—all may be catalogued and stored. By the criteria of information theory, then, my electronic image in the machine may be more real than I am. It is rounded and retrievable. Above all, it is a high-variety image—higher very likely than the image of me in the minds of my own friends. The behaviour of the image is predictable in statistical terms. Probably I am not. But the strength of the machine image is its pragmatic validity. There is no confusion here, no ambiguity, no loss of history, no rationalization. I am a mess; and I don't know what to do. The machine knows better—in statistical terms. Thus is my reality less real than my mirror image in the store. That fact diminishes me.

Second of the threats to my reality, there is the likelihood of my manipulation on a scale which is also frightening. Overt advertising has already taken us to the brink of what seems to be tolerable in this respect. But at least we are conscious of the risk—we may note the Freudian images of the ad-man cult, and the importunity of slogans which are akin to physiological conditioning. We may thus protect our personalities. But the computer's machinations are covert. A long-term record of my purchases should enable a computer to devise a mailing shot at me which is virtually irresistible.

As to involution: We earlier made the assumption that esoteric boxes themselves will engage in dialogue with their own clients and with governments to protect the citizen in this threatening situation. The important thing is not so much that this ought to happen as that it will certainly happen. For if it is vital to the social institution to remain integral, and if it is the proclivity of that esoteric box to be highly stable, then integrity and stability will be supported and reinforced by the highest ethical codes where professions are concerned, and by commercial self-interest where they are not. Each esoteric box will identify its own vested interest in solving these problems; and in solving them it will increase its own stability and survival power. Then these systems will become more involuted, and yet more esoteric; they will become more stable, and more resistant to change; in many cases it will be literally impossible to assess the information they contain without a special electronic key.

As the solutions begin to emerge from the studies which institutions are already making, it can be expected that legislative force will be asked for the implementation of any provisions which repeatedly occur as proposing matters of principle. For example, it already looks likely that legislation will be sought to permit the citizen access to his own computer files, or at least to permit him the knowledge that an entry has been made therein. Even so, there will be many difficulties for legislators, and especially difficulties of definition. After all, many records have been kept in the past, records made up with quill pens, of which the citizen had no intimate knowledge—and in cases of national security, or even of high-grade employment, perhaps no knowledge at all.

But the point for which we are reaching here really concerns the missing metasystems for the regulation and stabilization of strings and networks of esoteric boxes. If the inexorable trend is toward involution, and toward the isolation of information within the box, then the interchange of information between esoteric boxes becomes less and less likely (see Point (iii) in the *List*). Institutions will not dare to move towards the creation of metasystems, because this would breach confidentiality. As for the legislators, how can they possibly launch bills at one parliamentary sitting intending to keep information inside the box (for the reasons adduced), and then launch bills at the next sitting aimed at better management on the strings-and-network level? For the requirements of the second legislation would be to assemble information more economically for metasystemic purposes, to enrich the understanding of social needs by synthesizing information within higher-order models of the economy, and in general to seek modes of control which would necessarily diminish participation at lower societary levels to the point of total incomprehension as to what was going on.

This is a king-size dilemma. It has already been encountered in a relatively mild form by government bureaus of statistics, all of whom operate under legislation which guarantees the privacy of the individual firm by statistical aggregation. But in situations where large firms dominate sparsely populated localities, real skill may be needed to avoid betrayal of this rule by sheer accident. And perhaps in avoiding such risks the efficacy of the network will be sensibly reduced. The extension of the problem down to the rights of the individual, and up to meta-metasystems, and across to include the whole gamut of socio-economic behaviour, is a daunting prospect. But the difficulty is real; it will not go away.

So here is the meaning of Point (iv) in the *List of Characteristics* we set out to elucidate. Strings and networks of esoteric boxes will become less and less cohesive; and the metasystems they represent will become more and more unstable. These are the inexorable trends, and this is the basic reason why (I unhappily suggest) society is falling apart.

The Blurring of Interfaces

We have been seeking to elucidate the meaning of the four final points of the six statements made in the *List* which ended Section 2. It is time to revert to the first two of those six points. For in our recent discussion we have concerned ourselves primarily with information about the citizen as a product of either public or professional social institutions. But the argument of Points (i) and (ii) was that every facet of social control shared in the public and private sector, shared too in knowledge of the world as well as in knowledge of the citizen. Then let us begin a fresh analysis, beginning with the missing pieces of the puzzle, and see where that leads.

We want to talk in the first place, then, about knowledge of the world, and its dissemination as an entrepreneurial activity to anyone needing knowledge. This whole process began and continued historically in a very distinctive way. There were people—individuals by name—in the time of the ancient Greeks in whom reposed such knowledge as there was. Those wishing to acquire knowledge did so on a personal basis and at great expense. This often meant journeying to sit at the feet of an Aristotle and to learn from him. We might call the process *custom-built publishing*. We should note that it was a very high-variety process (the cybernetic analysis of a dialogue demonstrates to perfection Ashby's Law of Requisite Variety). And we should note finally that the effectiveness of the process relied on a relative paucity of knowledge compared with the capacity of the human brain and the calls on its time. For nearly two thousand years this situation prevailed. Although writing and its tools were developed, any piece of writing was still custom-built. One's copy of any text was a personal copy, bearing unique imperfections, omissions and additions. Then, five hundred years ago, came printing—a process which remains almost unchanged to this day as the accepted principle of permanent imaging.

It was the invention of printing that procured the first qualitative change in the management of information and knowledge for mankind. In achieving the massive dissemination of knowledge Gutenberg and Caxton also destroyed its custom-built character. In mitigation, the publishing industry (as it has become) developed an activity called editing. This critical occupation fulfills almost exactly the same function as naming or diagnosis in our earlier model. It constitutes a crossover point between a high-variety set of information on the one hand and a high-variety set of clients on the other; it selects subsets from each, and attempts to match them. Insofar as the matching succeeds, there is a marketable product. This may be defined as an edited publication, identifying a sufficient number of clients satisfied with the editorial process as between them to pay for the cost of publishing and printing (with of course a profit margin for all concerned).

The steady development of this whole marketing operation has led, like all other recent developments in the dissemination of human knowledge, to the informational overload mentioned before. Publishers continue to issue more and more printed material, relying specifically on their editorial skill in identifying market subsets willing to pay the price. But increasingly the process depends on mythology. It is easy enough to demonstrate that in fact the overload threshold has already been passed, and that (as we said) the sign has now changed on the stream of data input. No professional man can possibly read more than a fraction of what he would like to read or feels he should read. In some professions, current trends when extrapolated show that the whole population of professionals will shortly be employed in preparing abstracts of papers—whereupon no authors will be left to write them. This shows that insofar as people continue to purchase new publications they are not driven to do so by any residual capacity to convert data into information (meaning: what changes us). One may entertain various theories about the motives which do drive them. Such theories range from feelings of guilt and a sense of threat at one end of the spectrum to a pious belief that the editing process is (hopefully) converging on *my* special interests at the other. However this may be, no professional man can now cope effectively with the material he is expected to buy; and most would agree that they buy more than they can cope with.

Various mechanisms may operate to put an end to this situation, perhaps quite suddenly and dramatically. Which mechanisms will operate depends on which motives turn out to be most significant. For example, insofar as many publications depend on their advertising revenue for survival, then when the advertisers become aware that their advertisements are not even seen (because the journals are not opened) they may suddenly and disastrously withdraw support en masse. But the more profound threat to the established mores of the whole industry derives from the likelihood that someone will give convincing entrepreneurial effect to the unrecognized but inexorable trends of the situation. These are quite simply that professional people have a need for less and not more information, and therefore—in the long run—are going to *pay* for less and less, and to refuse to pay for more and more. The publishing industry and government itself continue to regard data as equivalent to information. The metasystem in which this issue can alone be sensibly contemplated, will shortly recognize that any one client is overloaded by any one editor who provides for the needs of a coterie, however small, intended to cover his costs.

There is then an inexorable requirement for a return to custom-built publishing directed at individuals, whether private citizens or cabinet ministers. This service must be economically viable, once the necessity for it is generally recognized, because it meets a need which cannot much longer be ignored. Moreover, the new technology is able to supply it. We shall use the power of computers to undertake an editing process on behalf of the only editor who any longer counts—the client himself. It matters not whether the information reaches that client on a computer terminal, or in a custom-built personalized print. Economics and personal preference will decide that issue. What does matter is the inevitable reversion to the age-old principle of publishing based on the finite capacity of the brain to assimilate data, and to convert them to the information which changes the brain's condition. And in all of this we may note the mechanisms already uncovered in this paper: especially obedience to the Law of Requisite Variety, and the vitality of the principle of adaptive feedback.

I here repeat that this kind of prognosis is not to my mind a matter of forecasting, but the detection of an inexorable requirement. There is no need to extend the argument to publication in the field of leisure, important though this is, because the considerations are much more difficult—and I think longer term. But the field of professional publishing, which includes knowledge about the whole of science and technology, and includes knowledge about everything that government may do, is sufficiently significant in itself. Both areas may be treated as their own esoteric boxes. In both cases there has to be a high variety of exposure of the client to the system, and there has to be fast adaptive feedback. If you will allow that this is possible, then we reach a new dimension of concern in the field of socio-economic management.

We know by now, as a matter of principle, that the increased effectiveness of the service provided inside an esoteric box increases the vulnerability of its clients to intimate relevations—because of Ashby's Law. The case of both commercial and gove nmental publishing to professional individuals offers no exception. In exactly the same degree, and by exactly the same mechanisms, that custom-built publishing becomes effective at all, so does the increasingly well-served client become a target of exploitation. Insofar, that is, as a particular product of either commercial or governmental publishing is especially meant for me, valuable to me, valued by me * * *. So far is it irresistible to me. We encountered this point before.

There is no problem here so long as we continue to speak of professional publishing by reputable publishers (and governments) itself. The matter for concern is of course that if such a system works for this purpose it will work for other and nefarious purposes too. If we can encode an individual's interests and susceptibilities on the basis of feedback which he supplies, if we can converge on a model of the individual of higher variety than the model he has of himself, then we have exactly the situation inside the automated system which was observed to be such a threat in more protected contexts. I think that marketing people will come to use this technique to increase the relatively tiny response to a mailing shot which exists today to a response in the order of ninety percent. All this is to say that the conditioning loop exercised upon the individual will be closed. Then we have provided a perfect physiological system for the marketing of anything we like—not then just genuine knowledge, but perhaps "political truth" or "the ineluctable necessity to act against the elected government". Here indeed is a serious threat to society.

Now we can see how the first three points in the *List of Characteristics* about the behaviour of esoteric boxes are indivisible from the last three points. Knowledge of the world and knowledge of the citizen are indissolubly united in systems of the kind we must expect; private and public interests moreover are inseparably involved in each. Then the interfaces between these four major components of information systems become hopelessly blurred. We shall not be able to legislate to keep what is indivisible divided. These arguments are based on realities manifested by situations which cannot be controlled at their own level without interference on a totalitarian scale in the rights and autonomy of our social institutions, the esoteric boxes.

4. Metasystem Management

The jigsaw puzzle is complete. We have looked closely at the emergent picture of interacting social institutions, exemplified as esoteric boxes. They are stable, involuted, resistant to change. Their interaction is embodied in strings and networks of complex connectivity, exemplified as metasystems. These are unstable, mercurial, existing more in concept than reality. The problems of information management that assail the boxes will be solved, if with the greatest difficulty. These solutions will themselves inexorably increase the metasystemic instability, threatening to blow society apart.

If all this offers an effective generalization of the problem of data pollution, and if we are to see any possibility of its solution in terms of good cybernetics, practice is needed in applying the model here envisaged. Let us then look at two levels of application, as widely separated as possible, to see how readily the systems concerned map onto each other, and what may be the commonalities of acceptable metasystemic controls.

First Example: at Hearth and Home

One plausible development of existing capabilities in informational science looks like this.

It is already possible to transmit textual material and the instructions for printing it into a television receiver—during a normal broadcast, and without interfering in any way with the broadcast itself. This is done by utilizing some of the enormous channel capacity available and not used by the flying spot defining the picture. For example, the spot has a "flyback" period, when it returns from the end of one scan to the beginning of the next. One line of scanning on a TV screen contains approximately six hundred bits of information. The flyback takes five lines to return, and is thus capable of carrying three thousand bits of information. If sixty frames are scanned every second (this would be fifty in Great Britain) there is spare capacity to transmit 180,000 bits a second of other information while the broadcast itself is going on. We know how to produce hard copy from the television set, using this input information. If we wish to produce a column of print six inches wide, with excellent resolution at a hundred lines per inch, we need 600 x 100 bits of information to produce an inch of text. It follows from all this that we have a capacity to produce three inches of text every second without interfering with the television broadcast.

Newspapers can be produced in the home like this, as is well known, and experiments continue. But newspapers are not custom-built; they belong to the informational overload. This overload is due to be met by custom-built publishing. Then apply the existing technology to the new publishing concept and see what happens.

Suppose that there are twenty buttons on the side of the television set which can be pressed by the viewer. The broadcaster invites the viewer to participate in a "personal response program". He shows the viewer two pictures, and asks him to press the first button if he prefers the first picture to the second—otherwise not. He then asks a question, and says that the second button should be pressed if the answer is yes—otherwise not. And so on. By the time the viewer has pressed or not pressed all twenty buttons he has identified himself in a high-variety way. For there are 2^{20} ways in which the set of buttons may be pressed, and that means enough patterns to distinguish between more than a million individuals (where each offers a separate pattern). As to privacy, the viewer is at home and alone with his set. So no one knows which buttons he presses (or do they?).

Having completed this exercise, the broadcaster suggests that the viewer should press his "print" button. The television set will then print out, from the vast amount of information being carried on the flyback, a piece of print which is determined by the particular pattern set on the twenty buttons. After all, if the sponsor hires one minute of flyback time at the end of his advertisement, he may *transmit* no less than a hundred and eighty inches of text. The "computer program" set up on the twenty buttons *selects* (say) six inches of this available text, and the apparatus prints it. This means that the individual concerned receives a very highly directed message. By the arguments used earlier, the viewer is likely to find this message irresistible. For example, the old lady sitting in one house reads "this product is especially suitable for old ladies", while the young man next door reads "this product is especially suitable for young men". (One needs little knowledge of the advertising world to recognize this example as remarkably naive.) Moreover, because the TV set is in a particular location, and can be preprogramed with that information, the custom-built message and advertisement could well include instructions as to which local supplier will make what special reduction for immediate compliance with the suggestion to purchase. Again, this example is offered for display purposes only: the opportunities are hair-raising. Suppose for instance that the apparatus is able to store previous sets of responses * * *.

The viewer lifts the telephone in order to place his order—or perhaps he simply presses a new button on his set labelled "yes". The supplier now has an order, and his system (for he is his own esoteric box) must immediately check the credit-worthiness of the customer. If by this time we have reached the cashless society, it could well be that the whole transaction is finalized and the viewer's bank account debited within the millisecond.

This is all entertaining, and something like it will very likely happen. Now consider the esoteric boxes on whose integrity and security he relies, but which he may by now himself have violated. The information he betrays might well include—

his medical status,
his educational status,
his intimate psychological situation,
the family context (i.e. someone else's privacy is breached),
the employment context (i.e. commercial security may be breached),
the economy's view of his credit,
the state of his bank balance,
his religious outlook.
his political outlook,
his social attitudes at large * * *.

Twenty bits, a variety of a million, every time: here is an inexhaustible source of metasystemic information available to anyone who sets out to acquire it. And from this information could be synthesized a new account of society and of the economy, orders of magnitude more powerful and valuable *and* threatening than any we have hitherto known or countenanced. With this unthinking violation of privacy goes the betrayal of all the mechanisms for protection and security to both the individual and the state which the esoteric boxes themselves have sought to guarantee. And with it go also the distinctions between public and private information, knowledge of the citizen and knowledge of the world.

Second Example: in World Economics

Undertaking now the largest possible change in the scale of this thinking, and leaping over a staggering array of other plausible examples large and small, we turn to the future of mankind itself and the stability of world economics.

A consensus of opinion might define an economy as the observable, quantifiable aspect of the social metasystem. The metric of economics appears to offer the only *lingua franca* which enables us to talk in figures about strings and networks of esoteric boxes—for typically these have no other commensurable denominator. But it seems to me most important to observe that this circumstance has let us into jejune descriptions of the social weal—which are obsessionally treated as *merely* economic. Surely no-one can believe that the total state of the world with all its pressures, ethnic, religious, liebensraum-oriented, power-geared, and all its problems of military, societary and environmental crisis, can adequately be discussed in terms of econometric models. Input-output analysis tells us something about the connectivity of esoteric boxes; cash flows say a little about their dynamic interrelationships; but we may discuss fiscal and monetary policy until we are purple without touching on the major causes of even economic disequilibria, still less of social dysfunction. This contention is relevant at both the national and the international level.

Having criticized the metric that is used and the models that are adopted, I may readily claim that the networks linking social institutions at this level are the most tenuous yet discussed. This underlines the fact that major political entities—states and nations—are the ultimately complete exemplifications of the esoteric box. They answer both to the definition of this term and to the behaviourial analysis of its operation. I shall risk as your foreign guest a remark about this which I hope will not be regarded as a solecism. We have just entered the decade in which the founding bicentenary of a remarkable interactive network will be celebrated in your country, the metasystem for which is perfectly exemplified in a federal constitution and its law. Is it not fair to say that there are esoteric boxes within this system, some of which are whole States while others are social institutions of other kinds, which maintain to this day those characteristics of the integral, stable, change-resistant box which we have taken much trouble to elucidate? And if there is cause for alarm about national instability, then surely it is metasystemic in nature. Correctives are hard to apply, for reasons we have also uncovered: they lead to involution and even exacerbate the problems.

At the level of world affairs, the case is far more strong. The sovereign nation is the ultimately esoteric box; the interconnective networks between nations are like so much spun silk. All the mechanisms described here clearly operate, and they too are clearly involutionary. The problems and threats are the same, but they are writ large. Just as we may identify spurious metasystems purporting to link the esoteric boxes of our own social institutions, so there are spurious international metasystems. All approaches to world government, from the League of Nations onwards, and including market-oriented consortia, *speak* metalinguistically but do not *operate* metasystemically. This is why I call them spurious. Hence it is in the network of world economies that we find the ultimately inadequate description and the ultimately incompetent management of the ultimately unstable metasystem.

Perhaps the nearest approach to genuinely stable organizations of this kind are the multinational companies. They represent linkages of esoteric boxes, beyond doubt; they certainly have identifiable metasystems. Even so, the cohesive forces required to make them survival-worthy barely emerge—given a potentially hostile environment. Do we have adequate management mores and philosophy, company lore, or international law, to underwrite their responsible self-regulation? It is a serious question, bearing in mind that these companies are in a sense the emergent *nations* of the next few decades. I mean by this that the gross product of some mushrooming companies already exceeds the gross national product of the smaller historic nations—for whom tradition, constitution, legal precedent and other long-standing regulators provide the cybernetic grounds of stability.

The vision of a small but historic nation in revolt is bad enough. The explosion of knowledge among people whose intellectual horizons are thereby expended and burst, the extension of personal vulnerability and loss of security through the uncontrolled spread of informational networks, and the political threats let loose by all of this, could turn such revolt into a societary crisis for that nation of unexampled magnitude. It would have to rely very heavily on the propensity to stability of the esoteric box to contain the situation. But what if instability such as this were to assail a multi-national company of greater size than this nation, a company that is not itself truly an esoteric box but a network existing at the metasystemic level without a metasystem. This would be a leviathan greatly to be feared, a leviathan obscenely polluted by its own data which it found itself powerless to metabolize.

Outcomes for Action

Action is required. The form of this action is a matter for you rather than for me. My endeavour has been to penetrate the immense complexity of the information management problem, in search of a scientific generalization. This I have tried to define in fairly plain English, to describe and to exemplify. The objective was to aid your endeavour to decide right action.

I suggested before that the problem is to manage complexity itself, complexity considered as the very stuff and substance of modern society. In the end, when all the computers have crushed their numbers to the last intransigent bit, the unquenched spirit of man takes final responsibility for life or death. Even so, this spirit necessarily operates—for ordinary folk and senators alike—through the medium of the human brain. This is one computer among many larger and faster (if so far less flexible) computers.

The cerebral computer is no more than a three-pound electrochemical device, slightly alkaline, which runs ten thousand million logical elements on the power of glucose at twenty-five watts. Its ability to discriminate is somewhat less than people imagine when they think of the human being in mystical terms as suffused by the divine afflatus. We can in general discriminate on a variety-scale of about nine. To understand an average is our métier. If we then judge that something is slightly, considerably, much or hugely better or worse than the mean, we have done as much discriminatory computing as we can normally manage. That scale of nine points is an output of roughly 3.2 bits of information.

Improvement in requisite variety is possible, since we enrich the dimensions of our comprehension by inter-relating several scales of discrimination. Even so, our human capability is geared at this general level. So when data processing systems offer us millions of bits of data, we dare not believe in a mythical metabolic process which could convert these data into information within our personal ken. There are ineluctable limits to the assimilation of knowledge, set by the finite size of the sugar-furnace in our heads. These facts to my mind determine the sorts of action which count as both feasible and effective. I have refined my ideas about this to offer a final set of specific postulates.

1. We may reasonably assume that esoteric boxes can take care of themselves, since that is what they are for.

2. They can be aided: their actions can be facilitated or inhibited by government. Any intervention, however, interferes with autonomy, denies participation, and may prove ineffectual (by Le Chatelier's principle).

3. Then legislation directed *into* particular boxes is unlikely to be much help. In any case, there is probably no time to tackle the problem at this level.

4. Then the *focus of attention* should always be at metasystemic level. This is the locale of societary instability; here then reside the massive threats.

5. First, the relevant metasystem must be identified, and in some sense institutionalized. Otherwise, who is to act or who can be held accountable? This primary task can be undertaken only by those holding the constitutional mandate.

6. The purpose of a metasystemic social institution is precisely, and only, to embody the *nerve-centre* for metasystemic affairs. Its function is precisely, and only, to identify situations of dangerous and therefore explosive instability, and to identify trajectories leading to stability.

7. The recommended methodology is the construction of metamodels, continuously innervated by data effectively filtered through a cybernetically designed hierarchy of systems.

8. The implementation of conclusions might be vested in the metasystemic social institution; if it is, however, there will be problems about autonomy (see 2).

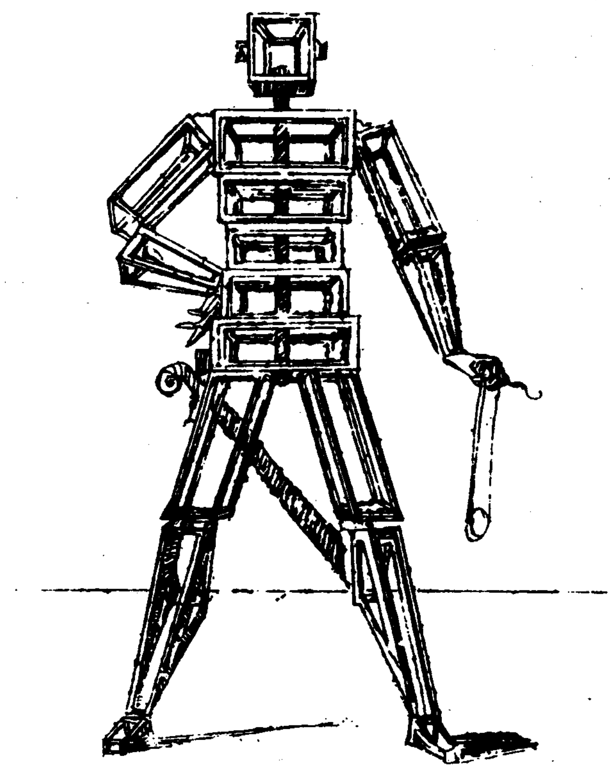
9. Insofar as legislation may be needed, the need can be pinpointed by these means. Directing either legislation or central executive action at strings and networks in the absence of metamodels is likely to increase instability rather than reduce it.

10. The kinds of model needed operate necessarily at a high level of abstraction: this makes almost everyone impatient. Consciously identify, then, the barrier to progress as anti-intellectualism.

Some metasystemic institutions already exist. The World Health Organization and the Food and Agriculture Organization are examples at international level, as are several international economic bodies; government departments handling whole sets of esoteric boxes are examples at national level. The questions I leave with you are these: Have such institutions been correctly identified? Do they at all map onto the *dynamic structure* of viable systems as understood by cybernetics? What other such institutions would be required in a stable society?

As to the warning in Point 10 about anti-intellectualism, it seems that the arguments used here would themselves predict this self-defeating syndrome in a society newly faced with the need to manage overwhelming complexity. If the brain is eclipsed in terms of variety by the computer milieu, it may itself revolt. Then panic-stricken attempts at the highest and most responsible level to quell forces that are not understood are as dangerous as the irresponsible cavorting of hooligans. One may already detect at either end of the scale of social responsibility a response equivalent to laying about with the jawbone of an ass.

The alternative is to *design* a stable society, and to treat our complexity-control capability through computers as offering a nervous system for the body politic. This involves the deployment of a political science to new ends, in the recognition that our difficulties have gone beyond anything that can be grasped by a slogan. We should recognize a cybernetic issue for what it is. But when the unthinkable is already happening it is indeed difficult to think about, and we are robbed of our semantic strength.

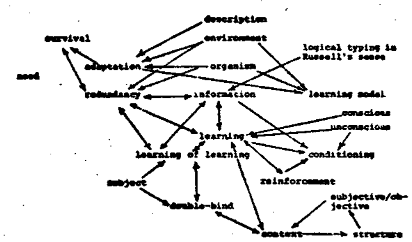


CONDITIONING

In the early phases of any science it is common that binary distinctions are applied to polymorphic classes of phenomena. Examples: "conditioning"/"learning", "analogue"/"digital", etc. The classification, learning/conditioning, is applied to the various contexts in which mammals are changed by experience, the ways in which the information is stored, etc., as follows:

Conditioning	Learning
The context of change is such that the subject does not control the coming of reinforcement. Reinforcement (so-called "unconditioned stimulus" is contingent upon the stimulus (so-called "conditioned stimulus") and lapse of time.	Reinforcement is at least partly contingent upon subject's actions.
The actions in response to stimulus are commonly innervated by autonomic nervous system.	Actions innervated by medullated nerves.
Actions carried out by unstriped muscle.	Actions carried out by striped ("voluntary") muscles.
The subject is unconscious of the process of change. Unconscious of the stimuli. Unconscious of the actions. Has no insight. No voluntary control.	Conscious. Conscious. Conscious. Has insight. Has control.
Animal or human subject obeys Russian cultural premises.	Subject obeys WASP premises.

There is no reason to believe that this Procrustean dichotomy can be applied to invertetrates - but vertebrates and especially mammals with their high degree of telencephalization do show some linkage or correlation between the various criteria.



Personally, I prefer to use the word "learning" for all input of information to an organism - whether conscious or unconscious, whether in the form of report or injunction, and of whatever logical type. [G.B.]

ADAPTATION

Imagine a description of an organism-and-its-environment. This description will consist of an internally interlocking tangle of descriptive propositions. Divide this tangle according to taste into two parts, "organism" and "environment". Now "adapted" and "adaptation" (the latter being the process of becoming the former) are words which refer to <u>certain sorts of redundancy</u> across the arbitrary dividing line.

What sorts of redundancy?

It is so, that some constellations of descriptive propositions endure longer than others, and <u>ergo</u> some systems endure longer than others. In particular, certain enduring constellations and systems owe their endurance to certain sorts of redundancy across the organismic/environmental boundary.

These instances of redundancy are called "adaptations".

Example I

O/E line divides 1, 4, 6/2, 3, 5

Description of O.	Description of E.
1. O is edible.	2. E contains predators which eat O's.
4. O is green.	3. Predators have color vision.
6. O has habit of sitting on green patches.	5. E contains green patches.

Example II

O/E line: 1, 5, 6, 8/2, 2, 3, 4, 7

1. O is sensitive to cold.	2. E contains wood.
	3. Wood is non-conductor (wind and temperature).
5. O is carpenter.	4. Weather is beastly.
6. O makes houses.	7. E now contains houses.
8. O lives in houses.	
9. O becomes more sensitive to weather.	

Example III

O/E line: 1, 4, 5/2, 3,6

1. O's muscles deteriorate if not exercised.	2. E contains other O's.
4. O responds competitively, engaging in conflict.	3. Other O's challenge O.
5. O's muscles thereby exercised.	6. Muscles of other O's likewise.

Note that each example starts with a proposition defining "weakness" or "need". It is not, of course, suggested that this proposition represents anything more than a justification for calling what follows an "adaptation".

In no case is there unique determination between need and adaptation. The need could have been "satisfied" in a million other ways. And, in many cases, the "need" is retrospective mythology. When automobiles were first invented they were unnecessary and the roads even were unsuited to them. Since then the environment has been adapted to the machines and people nowadays cannot (or think they could not) live without automobiles.

Moreover it seems generally true that the connection between need and adaptation is indirect. For example:

Example IV

1. O "needs" oxygen.	2. E contains oxygen in air.
3. O has respiration mechanism.	4. When O stops breathing, CO_2 accumulates in E (in blood-stream and alveoli).
5. O responds to CO_2 with increased respiration.	

All statements by evolutionary theorists equating the "need", currently satisfied by a given adaptation, with some part of the evolutionary determination of the adaptation are suspect. (Lamarckians and social reformers, please note).

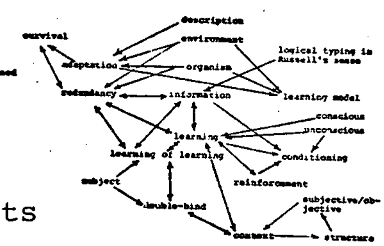
We have (I think) no satisfactory phrasing of the relation between components of the description and survival. The latter term is usually supposed to refer to the continuance of the organism under discussion. It could alternatively refer to the continuance of the (described) environment or to the continuance of the relation (description of patterns of interaction, etc.) between organism and environment. In this latter usage (which I prefer) an "adaptation" would be characterized by the ongoing truth of the propositions describing it.

In "co-evolution" (e.g., of horses and grassy plains), progressive (orthogenetic) changes in both organisms combine to maintain unchanged the relationship between them. Today horses are adapted to eat grass, and grass is adapted to make turf only if appropriately grazed, trampled, and manured by ungulates. The "adaptation" is mutual.

Of course (how could it be otherwise!) the descriptive propositions, whose ongoing truth is preserved by adaptive changes (e.g., in the dentition of the horses), are much more abstract - more general and of higher logical type - than the propositions which are altered by the evolutionary change.

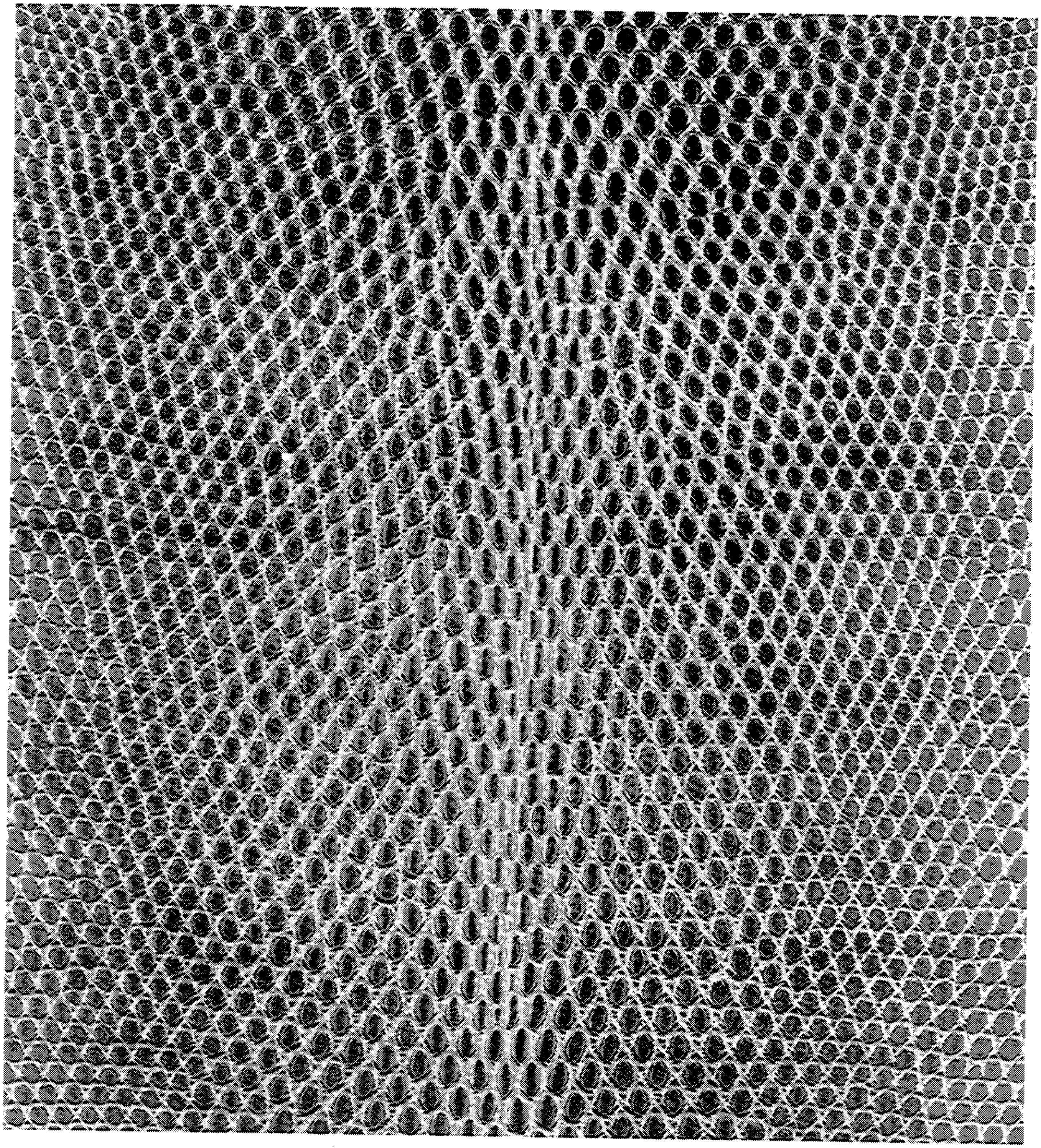
Returning to the definition of "those sorts of redundancy which are to be called adaptations", the matter can now be seen as requiring mapping (sc. definition) onto distinctions between descriptive propositions of higher or lower logical type and the resolution of problems of how redundancy in time (i.e., survival) is related to redundancy in space (or in synchronic description).

Two aspects of this problem can be mentioned here: first, it is clear that, *ceteris paribus*, there is a primary expectation that the redundant will outlast the random. The redundant is already at an intersection of the complex network of determinisms. Further increase in degree of redundancy will simply be an increase in the number of threads of causation meeting at the point of redundancy - until finally the multiply determined proposition becomes "hard programmed" and cannot be changed without disrupting all the converging threads.



Second, the characteristics of cybernetic circuits discussed in many other parts of this glossary will be relevant to the question of redundancy through time. [G.B.]

THE FUNDAMENTAL PURPOSE OF A SYSTEM SHOULD NOT BE JEOPARDIZED NOR ITS FUNDAMENTAL OBJECTIVES SIGNIFICANTLY COMPROMISED, IN ORDER TO ACCOMODATE EVENTS OF EXTREMELY LOW PROBABILITY



RETICULUM

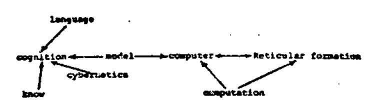
A network of connections in which unique pathways may or may not be specifiable. [S.B.]

RETICULAR FORMATION

The Reticular Formation of a vertebrate animal is a "net" of neurons, as the word indicates, in the spinal cord and brain stem. The core of this formation is credited with determining from moment to moment the class of thing that the animal will do. McCulloch's list of those classes follows:

1. sleep
2. eat
3. drink
4. fight
5. flee
6. hunt
7. search (or explore)
8. urinate
9. defecate
10. groom
11. mate
12. give birth (or lay eggs)
13. mother the young (including suckling the young, or hatching, retrieving, perineal licking, etc.)
14. build or locate nest
15. special innate forms of behavior such as migrate, hibernate, gnaw, hoard, depending on the species

The core of the reticular formation is one of the "computational domains" or "computers" identified by McCulloch in the vertebrate central nervous system. Prof. William Kilmer devised a model of the reticular core which operates in the large general purpose computer at Draper Laboratory. [L.S.]

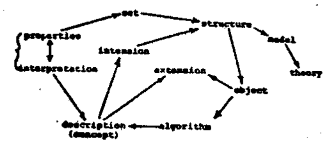


STRUCTURE

Structure comes from the Latin word struere: to build, and refers to what is built, and to the way the components of what is built are put together. The structure of a system is the set of components and relations between components making up the unity. [K.L.W.]

STRUCTURE

Structure within model theory is a set endowed with a maximal set of relations* (properties*) with respect to the predicate symbols of the theory to be interpreted in the structure. Thus a structure is described by a maximal intension and hence may be looked upon as a concrete object, suitable for modelling of theories. [L.L.]

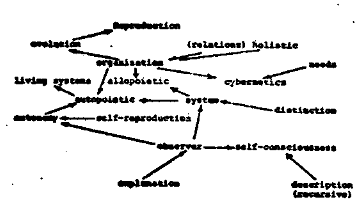


ORGANIZATION

Organization comes from the Greek word organon: instrument, and refers to the function of components in the constitution of a whole. The organization of an entity or system is the set of relations that the observer specifies as defining the entity. [K.L.W.]

ORGANIZATION

All of the relations between the components of a system, and which define the system as such. In contradistinction to structure, the actual relations holding between the components of a given system, organization refers to classes of relations which can be realized by several possible structures. Organization is truly the subject matter of a system science in a broad sense. Its central connotation, that of a holistic view of what is seen, is age-old, but only recently understood as a discipline in itself. Briefly: the whole is certainly more than its parts, it is the parts and its organization. [F.V.]

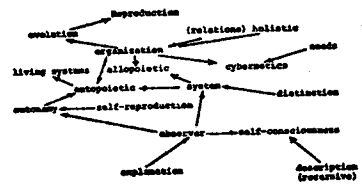


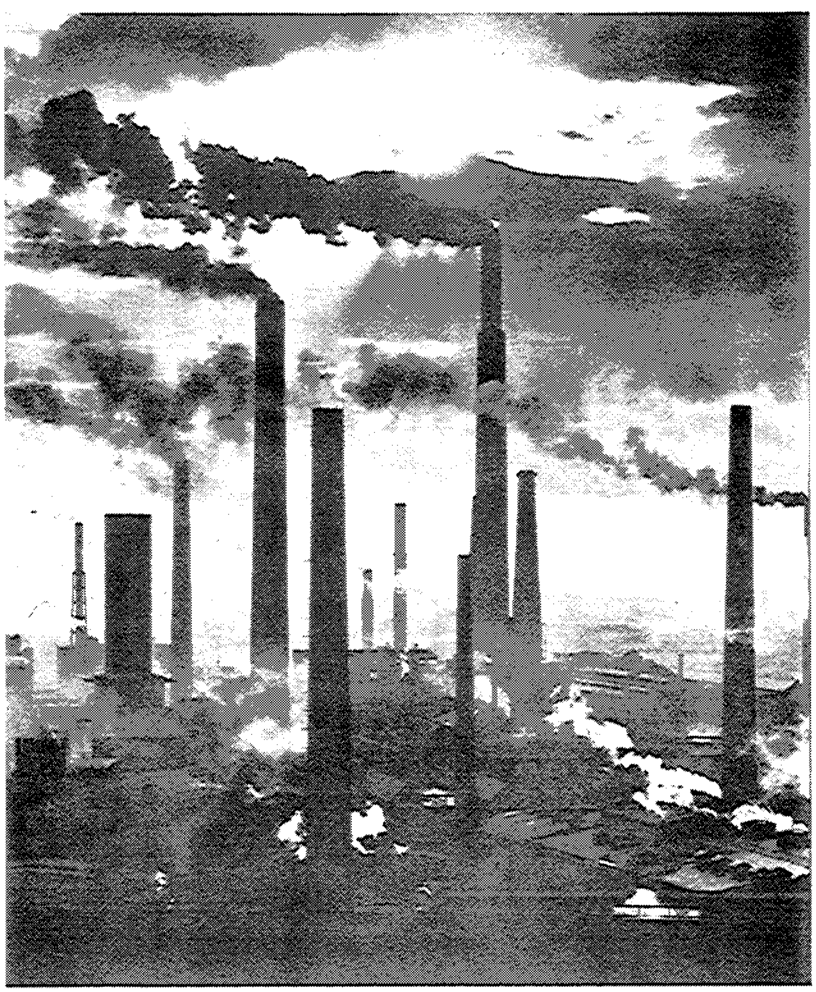
ALLOPOIETIC

An allopoietic system is one in which the product of operation of the system is different from the system itself. In a man-made system such as a car, there is a concatenation of processes which specifies organization and yet does not produce the components of the car since the components of the car are produced by processes which are independent of the car and its operation. [K.L.W.]

ALLOPOIETIC

A class of organization. Allo/autopoietic descriptions can only apply when the system considered is defined by productions, i.e. when the relations between the components that integrate it, are productions. Whenever the productions are not productions of precisely the components which integrate the system (in whichever space) the system is allopoietic. The vast majority of systems commonly studied are allopoietic. [F.V.]





ASHBY

THE SELF-REPRODUCING SYSTEM*

W. ROSS ASHBY

High among the interesting phenomena of organization shown by life is that of reproduction. We are naturally led to ask: How can a system reproduce itself? And we go headlong into a semantic trap unless we proceed cautiously. In fact, the answer to the question, "How does the living organism reproduce itself?" is "It doesn't."

No organism reproduces *itself*. The only thing that ever has had such a claim made for it was the phoenix, of which we are told that there was only one, that it laid just one egg in its life, and that out of this egg came itself. What then *actually* happens when ordinary living organisms reproduce? We can describe the events with sufficient accuracy for our purpose here by saying:

(1) There is a matrix (a womb, a decaying piece of meat, a bacteriological culture tube perhaps).
(2) Into it is introduced a form (an ovum, a fly's egg, a bacterium perhaps).
(3) A complex dynamic interaction occurs between the two (in which the form may be quite lost).
(4) Eventually the process generates more forms, somewhat like the original one.

In this process we must notice the fundamental part played by the matrix. There is no question here of the ovum reproducing *itself*. What we see is the interaction between one small part of the whole and the remainder of the whole. Thus the outcome is a function of the *interaction* between two systems. The same is true of other forms. The bacterium needs a surrounding matrix which will supply oxygen and food and accept the excretion of CO_2, *etc.* An *interaction* between the two then occurs such that forms somewhat resembling the initial bacterium eventually appear.

So, before we start to consider the question of the self-reproducing system we must recognize that *no organism is self-reproducing*. Further, we would do well to appreciate that Rosen [2] has recently shown that the idea of a self-reproducing automaton is logically self-contradictory. He uses an argument formally identical with that used by me [1] to show that a self-organizing system is, strictly, impossible. In each case the idea of a *self*-acting machine implies that a mapping must be able to alter itself—*i.e.*, that it is within its own domain. Mathematics and logic can do nothing with such a concept. It is in the same class as the fantasy that can see a man getting behind himself and pushing himself along.

I make these remarks, not in order to confuse or to obstruct, but simply to make sure, by clearing away sources of confusion, that we do really find the right approach to our topic. Though the adjective "self-reproducing" is highly objectionable semantically and logically, it does of course refer to a highly interesting process that we know well, even if we sometimes use inappropriate words to describe it.

I propose, then, to consider the question re-formulated thus:

A given system is such that, if there occurs within it a certain form (or property or pattern or recognizable quality generally), then a dynamic process occurs, involving the whole system, of such a nature that eventually we can recognize, in the system, further forms (or properties or patterns or qualities) closely similar to the original.

I ask what we can say about such systems.

CAN A MACHINE DO IT?

Having got the question into its proper form, we can now turn to the question whether a machine can possibly be self-reproducing. In a sense the question is pointless, because we know today that all questions of the type "Can a machine do it?" are to be answered "Yes." Nevertheless, as we are considering self-reproduction, a good deal more remains to be said in regard to the more practical details of the process. Our question then is: Does there exist a mechanism such that it acts like the matrix mentioned, in that, given a "form," the two together lead eventually to the production of other forms resembling the first?

*The work on which this paper is based was supported by the Office of Naval Research, Contract N 62558-2404.

I propose to answer the question largely by a display of actual examples, leaving the examples to speak for themselves.

The first example I would like to give is a formal demonstration in computer-like terms showing the possibility. Let us suppose a computer has only ten stores, numbered 0 to 9, each containing a two-digit decimal number, such as 72, 50, 07, or perhaps 00. The "laws" of this little world are as follows: Suppose it has just acted on store S−1. It moves to store S, takes the two digits in it, a and b say, multiplies them together, adds on 5 *and the store-number S*, takes the right-hand digit of the result, c say, and then writes the original two digits, a and b, into store c. It then moves on to the next store and repeats the process; and so on indefinitely.

At first sight, this "law" might seem to give just a muddle of numbers. At store No. 3 say, with 17 in the store, it multiplies together 1 and 7, adds 5 to the product, getting 12, adds the store number 3, getting 15, takes the right-hand digit, getting 5, and puts 17 into store 5. It then goes on to its next store, which is No. 4. There seems to be little remarkable in this process. On the other hand, a 28 in a store has a peculiar property. Suppose it is in store 7. $2 \times 8 = 16$, $16 + 5 = 21$, $21 + 7 = 28$, 28 gives 8, so 28 goes into store 8. When we work out the next step we find that 28 goes again into store 9, and so on into store after store. Thus, *once a 28 turns up in the store it spreads until it inhabits all the stores.* Thus the machine, with its program, is a dynamic matrix such that, if a "28" gets into it, the mutual interaction will lead to the production of more 28's. In this matrix, the 28 can be said to be self-reproducing.

The example just given is a formal demonstration of a process that meets the definition, but we can easily find examples that are more commonplace and more like what we find in the real world. Suppose, for instance, we have a number of nearly assembled screw drivers that lack only one screw for their completion. We also have many of the necessary screws. If now a single complete screw driver is provided, it can proceed to make more screw drivers. Thus we have again the basic situation of the matrix in which if one form is supplied a process is generated that results in the production of other examples of the same form.

On this example, the reader may object that a great deal of prefabrication has been postulated. This is true, of course, but it does not invalidate the argument, because the amount of prefabrication that occurs can vary over the widest limits without becoming atypical; and some prefabrication *has* to be allowed. After all, the living things that reproduce do not start as a gaseous mixture of raw elements.

(The same scale of "degrees of prefabrication" sometimes confuses the issue when a model maker claims that he has "made it all himself." This phrase cannot be taken in any absolute sense. If it were to be taken literally, the model maker would first have to make all the screws that he used, but before that he must have made the metal rods from which the screws were produced, then he must have found the ores out of which the metal was made, and so on. As there is practically no limit to this going backward, the rule that a model maker "must make it all himself" must be accompanied by some essentially arbitrary line stating how much prefabrication is allowed.)

The two examples given so far showed only reproduction at one step. Living organisms repeat reproduction: fathers breed sons, who breed grandsons, who breed great-grandsons, and so on. This possibility of extended reproduction simply depends on the scale of the *matrix*. It can be present or absent without appreciably affecting the fundamentals of the process.

FURTHER EXAMPLES

The subject of self-reproduction is usually discussed on far too restricted a basis of facts. These tend to be on the one hand simply the living organisms, and on the other hand machines of the most rudimentary type, such as the watch and the motor car. In order to give our consideration more range, let us consider some further examples. Those I give below will be found to be sometimes unorthodox but every one of them, I claim, does accord with the basic definition—that the bringing together of the first form and matrix leads to the production of later forms similar to the first.

Example 3. A factory cannot start producing because the power is not switched on. The only thing that can switch the power on is a spanner (wrench) of a certain type. The factory's job is to produce spanners of that type.

Example 4. A machine that vibrates very heavily when it is switched on can be started by a switch that is very easily thrown on by vibration. Such a system, if at rest and then given a heavy vibration, is liable to go on producing further heavy vibrations. Thus the form "vibration," in this matrix, is self-reproducing.

Example 5. Two countries, A and B, were at war. B discovered that country A was a dictatorship so intense that every document bearing the dictator's initials (X.Y.Z.) had to be obeyed. Country B took advantage of this and ruined A's administration by bombing A with pieces of paper bearing the message: "Make ten copies of this sheet, with the initials, and send to your associates. X.Y.Z." In such a matrix, such a form is self-reproducing.

Example 6. A number of chameleons are watching one another, each affected by the colors it sees around it. Should one chameleon go dark it will increase the probability of "darkness" appearing around it. In this matrix, the property "darkness" tends to be self-reproducing.

Example 7. In a computer, if the order 0101010 should mean "type 0101010 into five other stores taken at random," then in this matrix the form 0101010 is self-reproducing.

Example 8. A computer has single digit decimal numbers in its various stores. It is programed so that it picks out a pair of numbers at random, multiplies them together, and puts the right-hand digit into the first store. In this condition, as any zero forces another zero to be stored, the zero is self-reproducing.

Example 9. Around any unstable equilibrium, any unit of deviation is apt to be self-reproducing as the trajectory moves further and further away from the point of unstable equilibrium. Thus, if a river in a flat valley happens to be straight, the occurrence of one meander tends to lead to the production of yet other meanders. Thus in this matrix the form "meander" is self-reproducing.

Example 10. A similar example occurs when a ripple occurs in a soft roadway. Under the repeated impact of wheels, the appearance of one tends to lead to the appearance of others. In this matrix, "ripple" is self-reproducing.

Example 11. (Due to Dr. Beurle) A cow prefers to tread down into a hole rather than up onto a ridge. So, if cows go along a path repeatedly, a hollow at one point tends to be followed by excessive wear at one cow's pace further on, and thus by a second hollow. And this tends to be followed by yet another at one pace further on. Thus, in this matrix, "hollow" is self-reproducing.

Example 12. Well known in chemistry is the phenomenon of "autocatalysis." In this class is the dissociation of ethyl acetate (in water) into acetic acid and alcohol. Here, of course, the dissociation is occurring steadily in any case, but the first dissociation that produces the acid increases the *rate* of the later dissociations. So, in this matrix, the appearance of one molecule of acetic acid tends to encourage the appearance of further molecules of the same type.

Example 13. In the previous example the form has been a material entity, but the form may equally well be a pattern. All that is necessary is that the entity, whatever it is, shall be unambiguously recognizable. In a supersaturated solution, for instance, the molecular arrangement that one calls "crystalline" is self-reproducing, in the sense that in this matrix, the introduction of one crystalline form leads to the production of further similar forms.

Example 14. With a community of sufficiently credulous type as matrix, the introduction of one "chain letter" is likely to lead to the production of further such forms.

Example 15. In another community of suitable type as matrix, one person taking up a particular hobby (as form) is likely to be followed by the hobby being taken up by other people.

Example 16. Finally, I can mention the fact that the occurrence of one yawn is likely to be followed by further occurrences of similar forms. In this matrix, the form "yawn" is self-reproducing.

REPRODUCTION AS A SPECIALIZED ADAPTATION

After these examples we can now approach the subject more realistically. To see more clearly how special this process of reproduction is, we should appreciate that reproduction is not something that belongs to living organisms by some miraculous linkage, but is simply a specialized means of adaptation of a specialized class of disturbances. The point is that the terrestrial environments that organisms have faced since the dawn of creation have certain specialized properties that are not easily noticed until one contrasts them with the completely nonspecialized processes that can exist inside a computer. Chief among these terrestrial properties is the extremely common rule that if two things are far apart they tend to have practically no effect on one another. Doubtless there are exceptions, but this rule holds over the majority of events. What this means is that when disturbances or dangers come to an organism, they tend to strike locally. Perhaps the clearest example would be seen if the earth had no atmosphere so that the organisms on it were subject to a continuous rain of small shotlike particles traveling at very high speeds. Under such a rain the threat by each particle is local, so that a living form much increases its chance of survival if replicates of the form are made and dispersed. The rule of course is of extremely wide applicability. Banks that may have a fire at one place make copies of their records and disperse them. If a computing machine were liable to sudden faults occurring at random places, there would be advantage in copying off important numbers at various stages in the calculation so as to have dispersed replicates. Thus, the process of reproduction should be seen in its proper relation to other complex dynamic processes as simply a specialized form of adaptation against a special class of disturbances. It is all that and nothing more. Should the disturbances not be localized there is no advantage in reproduction. Suppose, for instance, that the only threat to a species was the arrival of a new virus, that was either overwhelmingly lethal or merely slightly disturbing. Under such conditions the species would gain nothing by having many distinct individuals. The same phenomenon can be seen in industry. If an industry is affected by economic circumstances or by new laws, so that either all the companies in it survive, or all fail, then there is no advantage in the multiplicity of companies; a monopoly can be as well adapted as a multiplicity of small companies.

FUNDAMENTAL THEORY

After this survey we have at least reached a point where we can see "reproduction" in its proper nature in relation to the logic of mechanism. We see it simply as an adaptation to a particular class of disturbances. This means that it is at once subject to the theoretical formulations that Sommerhoff [3] has displayed so decisively. The fact that it is an adaptation means that we are dealing essentially with an invariant of some dynamic process. This means that we can get a new start, appropriate to the new logic of mechanism, that will on the one hand display its inner logic clearly, and on the other hand state the process in a form ready to be taken over by machine programing or in any related process. We start then with the fundamental concept that the dynamic process is properly defined by first naming the set S of states of the system and then the mapping f of that set into itself which corresponds to the dynamic drive of the system. *Reproduction is then one of the invariants that holds over the compound of this system and a set of disturbances that act locally.* If then f is such that some parts within the whole are affected individually, "reproduction" is simply a process by which these parts are invariant under the change-inducing actions of the dynamic drive f.

It must be emphasized that reproduction, though seeming a sharply defined process in living organisms, is really a concept of such generality that precise definition is necessary in all cases if it is to be clear what we are speaking of. Thus, in a sense every state of equilibrium reproduces itself; for if $f(x) = x$, then the processes f of the machine so act on x that at a moment later we have x again. This is exactly the case of the phoenix. It is also "self-reproduction" of a type so basic as to be uninteresting, but this is merely the beginning. It serves as a warning to remind us that processes of self-reproduction can occur, in generalized dynamic systems, in generalized forms that *far exceed in variety and conceptual content anything seen in the biological world.* Because they are nonbiological the biologist will hesitate to call them reproducing, but the logician, having given the definition and being forced to stick to it, can find no reason for denying the title to them. What we have in general is a set of parts, over some few of which a property P is indentifiable. This property P, if the concept is to be useful, must be meaningful at various *places* over the system. Then we show that "self-reproduction of P" holds in this system if along any trajectory the occurrence of P is followed, at the states later in the trajectory, by their having larger values for the variable "number of P's present."

It should be noted that because self-reproduction is an adaptation, which demands (as Sommerhoff has shown) a relation between organism and environment, and because the property P must be countable in its occurrences over the system, we *must* be dealing with a system that is seen as composed of parts. I mention this because an important new development in the study of dynamics consists of treating systems actually as a whole, the parts being nowhere considered. This new approach cannot be used in the study of reproduction because, as I have just said, the concept of reproduction *demands* that we consider the system as composed of parts.

The new point of view which sees reproduction simply as a property that may hold over a trajectory at once shows the proper position of an interesting extension of the concept. Reproduction, as I said, is a form of invariant. In general, invariants are either a state of equilibrium or a cycle. So far, we have considered only the equilibria, but an equally important consideration is the cycle. Here we reach the case that would have to be described by saying that A reproduces B, then B reproduces C, and then C reproduces *A*. Such a cycle is of course extremely common in the biological world. Not only are there the quite complicated cycles of forms through the egg, pupa, imago, and so on that the insects go through, there is of course also the simple fact that human reproduction itself goes regularly round the cycle: ovum, infant, child, adult, ovum, and so on.

A further clarification of the theory of the subject can be made. Let us define "reproduction" as occurring when the occurrence of a property increases the probability that that property will again occur elsewhere; this of course is positive reproduction. We can just as easily consider "negative" reproduction, when the occurrence of a property *decreases* the probability that the property will occur elsewhere. Examples of this do not appear to be common. We can of course at once invent such a system on a general-purpose computer; such "negative reproduction" would occur if, say, the instruction 00000 were to mean "replace all zeroes by ones." I have found so far only one example in real systems—namely, if, under electrodeposition, a whisker of metal grows toward the electrode, the chance of another whisker growing nearby is diminished. Thus "whiskers" have a negative net reproduction.

This observation gives us a clear lead on the question: Will self-reproducing forms be common or rare in large dynamic systems? The *negatively* self-reproducing forms clearly have little tendency to be obtrusive—they are automatically self-eliminating. Quite otherwise is it with the positively self-reproducing forms; for now, if the system contains a single form that is *positively* self-reproducing, that form will press forward toward full occupation of the system.

Suppose now we make the natural assumption that the larger the system, if assembled partly at random, the larger will be the number of forms possible within it. Add to this the fact that if any *one* is self-reproducing, then self-reproducing forms will fill the system, and we see that there is good reason to support the statement that *all sufficiently large systems will become filled with self-reproducing forms.*

This fact may well dominate the design of large self-organizing systems, forcing the designer to devote much attention to the question: "What self-reproducing forms are likely to develop in my system?" just as designers of dynamic systems today have to devote much attention to the prevention of simple instabilities.

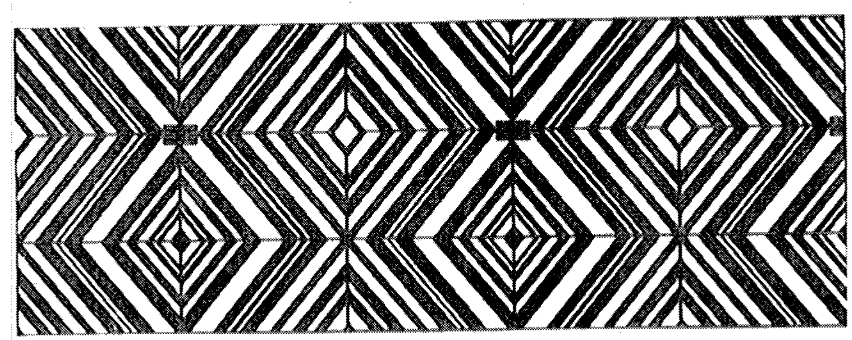
SUMMARY

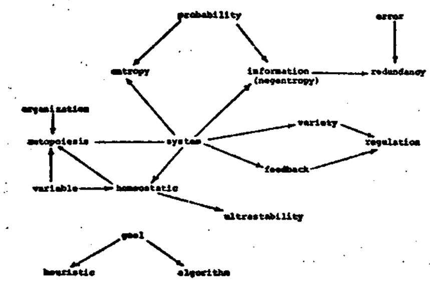
Reproduction has, in the past, usually been thought of as exclusively biological, and as requiring very special conditions for its achievement. The truth is quite otherwise: it is a phenomenon of the widest range, tending to occur in all dynamic systems, if sufficiently complex.

The brain may well use this tendency (for self-reproducing forms to occur) as part of its normal higher processes. The designer of large self-organizing systems will encounter the property as a major factor, as soon as he designs systems that are really large and self-organizing.

REFERENCES

1. W. Ross Ashby, "Principles of the Self-organizing System," Symposium on Self-organizing Systems, University of Illinois, June 7-10, 1960, Pergamon Press, 1962.
2. R. Rosen, "On a Logical Paradox Implicit in the Notion of a Self-reproducing Automaton," Bull. Math. Biophysics, Vol. 21, pp. 387-394, 1959.
3. G. Sommerhoff, "Analytical Biology," Oxford University Press, London, 1950.



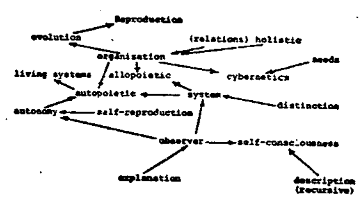


AUTOPOIESIS

In cybernetics, this is a special case of homeostasis in which the critical variable of the system that is held constant is that system's own organization. The term is due to Humberto Maturana. [S.B.]

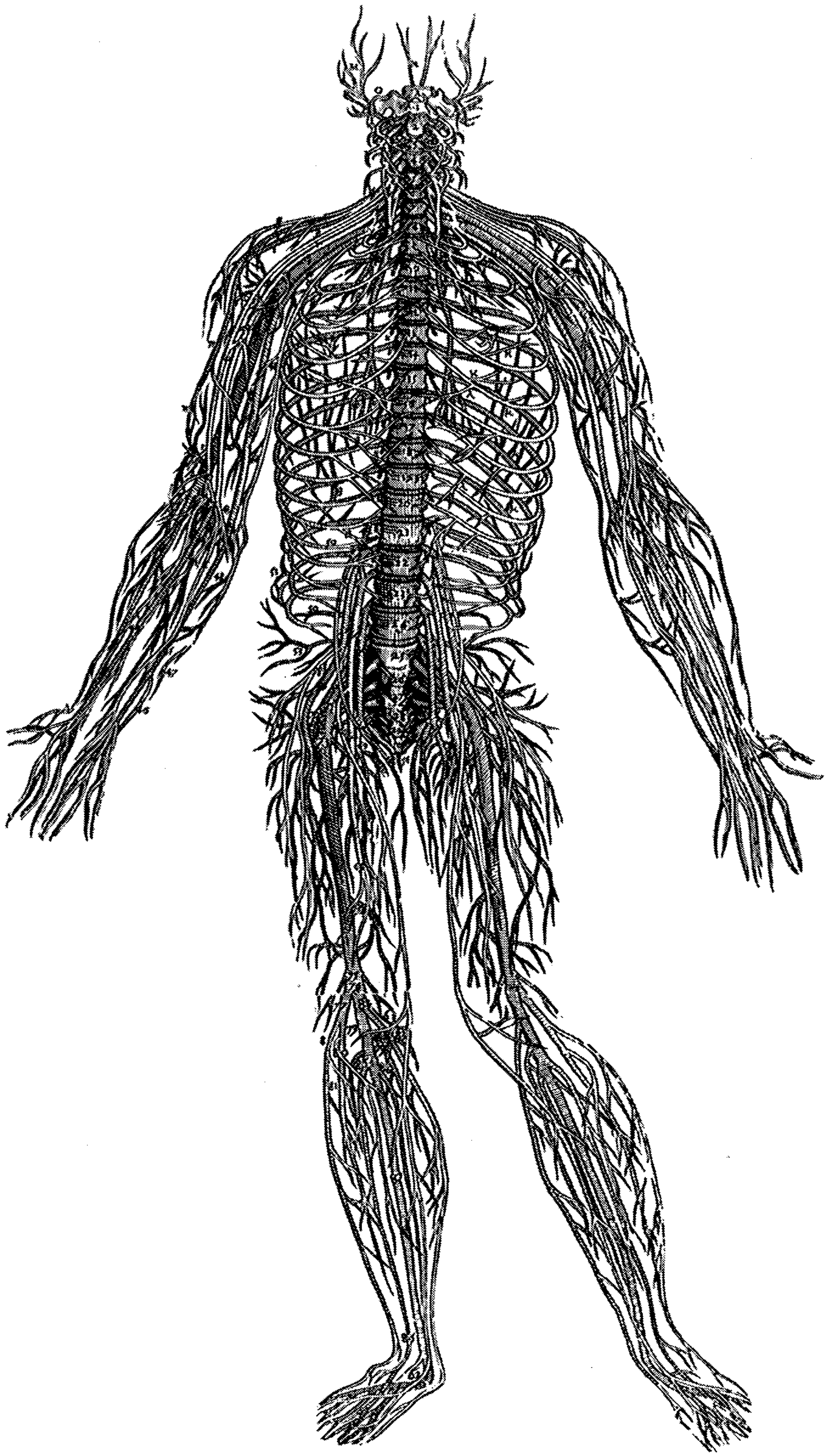
AUTOPOIESIS

A class of organization, characteristic of at least all living systems. In contradistinction with allopoietic systems, an autopoietic one is defined by productions of precisely those components which integrate it (in whichever space). Thus the phenomenology of autopoiesis is the phenomenology of autonomy insofar as the result of the system's dynamics is the system itself. [F.V.]



AUTOPOIESIS

An autopoietic system is defined as a unity by a network of productions of components which through their interactions give rise to these same productions. Thus the production network produces the components and the components, through their interactions, constitute the system as a unity in the space in which they exist and make the network possible by defining and realizing its topology. The product of operation of an autopoietic system is itself. Living systems are autopoietic systems. [K.L.W.]



Neurophysiology of Cognition

HUMBERTO MATURANA

University of Chile, Santiago

Whether stated or not, the fundamental quest of neurophysiology is to understand the processes that determine the so-called higher functions in the brain. The present work is an explicit attempt in that direction. Whatever success I have in this endeavor I owe to the enriching discussions I had with my friend Heinz Von Foerster and my student Francisco Varela.

STATEMENT OF THE PROBLEM

Propositions

Cognition is a biological phenomenon and can only be understood as such; any epistemological insight in the domain of knowledge requires this understanding.

If such an understanding is to be attained, two questions have to be considered:

1) What is cognition as a process?
2) How is it carried out?

I shall proceed accordingly.

Work on this paper was done while the author was Visiting Miller Professor at the University of Illinois.
This paper grew out of work jointly sponsored by AFOSR Grant AF 49(638)–1680 and by AFOSR Grant 7–67.

The connotations of the terms used will become apparent through their usage in the text, and the complete presentation is expected to provide the foundation for its parts.

COGNITIVE PROCESS IN GENERAL

The Observer

Anything said is said by an observer. In his discourse an observer speaks to another observer who could be himself, and whatever applies to one applies to the other as well. The observer is a human being, that is, a living system.

The observer beholds simultaneously the entity that he considers (an organism, in our case) and the universe in which it lies (the organism's environment). This allows him to interact independently with both and to have interactions that are necessarily outside the domain of interactions of the observed entity (the organism).

It is an attribute of the observer to be able to interact independently with the observed entity and with its relations; for him both are units of interactions (entities).

For the observer an entity is an entity (a unit of interactions) when he can describe it. To describe is to enumerate the actual or potential interactions and relations of the described entity. Accordingly, the observer can describe an entity only if there is at least one other entity from which he can distinguish it and with which he can observe it to interact or relate, even if this is the observer himself, and which serves as a reference for the description.

The set of all interactions into which an entity can enter is its domain of interactions. The set of all the relations (interactions through the observer) in which an entity can be observed is its domain of relations. (This latter domain lies in the cognitive domain of the observer.) An entity is an entity if it has a domain of interactions; and if it has a domain of interactions, it has a domain of relations. The observer can define an entity by specifying its domain of interactions; thus, part of an entity, a group of entities, or their relations can be made units of interactions (entities) by the observer.

The observer can define himself as an entity by specifying his own domain of interactions.

Any understanding of the cognitive process must account for the observer and his role in it.

The Living System

Living systems are units of interactions; they exist in an environment. From a purely biological point of view they cannot be understood independently of the part of the environment with which they interact, the niche; nor can the niche be defined independently of the living system that occupies it.[1]

Living systems as they exist on the earth today are characterized by exergonic metabolism, growth, and replication (and reproduction), all organized in a closed causal circular process that allows for evolutionary changes in the way the circularity is maintained, but not for loss of the circularity: exergonic metabolism is required to provide energy for the endergonic synthesis of specific polymers (proteins, nucleic acids, lipides, polysaccharides) from the corresponding monomers, that is for growth and reproduction; special replication procedures are needed to secure that the polymers synthesized be specific, that is, that they should have the monomeric sequence proper to their class; specific polymers (enzymes) are required for the exergonic metabolism and the synthesis of specific polymers (proteins, nucleic acids, lipides, polysaccharides). This circular organization determines that the components that specify it be those whose synthesis it secures. Hence, its circular nature is essential for its maintenance and its operation as a unit. That which is not in it is external to it or does not exist. This circular organization is the living organization.

It is the circularity of its organization that makes a living system a unit of interactions, and it is this circularity that it has to maintain in order to remain a living system and to retain its identity through different interactions. All the peculiar characteristics of the different kinds of organisms are superimposed on this basic circularity and are subservient to it, securing its continuance through successive interactions in an always changing environment. A living system defines through its organization the domain of all the interactions into which it can possibly enter without losing its identity; its identity is maintained as long as the basic circularity that defines the system as a unit of interactions remains unbroken. (Strictly, the identity of a unit of interactions that otherwise changes continuously is maintained only with respect to the observer for whom its character of being a unit of interactions remains unchanged.)

Due to the circular nature of its organization, a living system has a self-referring domain of interactions (self-referring system): the condition for being a unit of interactions of a living system is maintained, for its organization has functional significance only in relation to the maintenance of its circularity and defines its domain of interactions accordingly.

Living systems as interacting units cannot enter into interactions that are not prescribed by their organization. The circularity of their organization continuously brings them back to the same internal state (same with respect to the cyclic process). Each internal state requires that certain conditions (interactions with the environment) be satisfied to proceed to the next one. The circular organization implies the prediction that a necessary interaction that took place once will take place again. If this does not happen, the system disintegrates; if the predicted interaction takes place, the system maintains its identity (integrity) and enters into a new prediction. In a continuously changing environment these predictions can be successful only if the environment does not change in that which is predicted. Accordingly, the predictions implied in the organization of the living system are not predictions of particular events but of classes of interactions. Every interaction is a particular interaction, but every prediction is a prediction of a class of interactions that is defined by those features in its members which will allow the living system to retain its circular organization after the interaction and, thus, interact again. This makes living systems inferential systems, and their domain of interactions a cognitive domain.

The niche is defined by the classes of interactions into which an organism can enter. The niche is, then, to the observer, part of the environment; to the living system, all that with which it can interact without losing its identity. For the observer the environment in which the organism lies is larger than the niche, but this environment lies in the domain of interactions (cognition) of the observer, not of the organism. If the living system enters into an interaction not prescribed by its organization, it enters it not as the unit of interactions defined by this organization (but as some other unit of interactions or part thereof), and this interaction remains outside its cognitive domain. For every living system, its niche is represented in its organization as the domain of its possible interactions, and this domain constitutes its entire cognitive reality.

Evolution

Evolutionary change in living systems is the result of that aspect of their organization which secures the maintenance of their basic circularity, and which allows in each reproductive step for changes in the way it is attained. Reproduction and evolution are not essential for the living organization, but they have been essential for the evolution of the cognitive domains of the living systems on earth.

For a change to occur in the domain of interactions of a unit of interactions without it losing its identity (with respect to an observer), it must suffer an internal change. Conversely, if an internal change occurs in a unit of interactions, without it losing its identity, its domain of interactions changes. A living system suffers an internal change without loss of identity if the new prediction, brought forth by the internal change, is a prediction of interactions which do not interfere with its fundamental circular organization. A system changes only if its domain of interactions changes.

In reproduction the new unit of interactions has the same domain of interactions as the parental one only if it has the same organization. Conversely, the new unit of interactions has a different domain of interactions than the parental one only if its organization is different, and, hence, implies different predictions about the niche.

Since predictions about the niche are inferences about classes of events to be encountered (interactions), particular encounters that for the observer (and hence for other units of relations, such as internal parts of the same living system) are different, may be indistinguishable for the organism. The observer can describe them as different members of a class defined by the mode of interaction of the organism. As a consequence, interactions that are identical for the organism, but are different for parts of it, may result in different modifications of its internal states and, hence, different changes in its domain of interactions without loss of identity. These changes may bring about the production of an offspring with a domain of interactions different from the parental one. If this is the case and the new system predicts a niche that cannot be actualized, it disintegrates; otherwise it maintains its identity, and a new cycle begins.

What changes from generation to generation is the internal structure of the unit of interactions: that which through a circular organization retains its identity through numerous interactions. The manner in which the unit of interactions is compounded, whether by a single unit, or through the aggregation of numerous ones that together become a new larger unit of interactions (multicellular organisms, insect societies, etc.) is of no significance; what evolves is the unit of interactions. The evolution of the living systems is the evolution of the niches of the units of interactions, hence the evolution of the cognitive domains.

The Cognitive Process

A cognitive system is a system whose organization defines a domain of interactions in which it can act with relevance to the maintenance of the system itself, and the process of cognition is the actual (inductive) acting (or behaving) in this domain. Living systems are cognitive systems, and living, as a process, is a process of cognition. This statement is valid for all organisms, with and without a nervous system.

If a living system enters into a cognitive interaction, its internal state is changed in a manner relevant to its maintenance and it enters into a new interaction without losing its identity. In an organism without a nervous system (or its functional equivalent) its interactions are of chemical or physical nature (a molecule is absorbed and an enzymatic process is initiated; a photon is captured and a step in photosynthesis is carried out). For such an organism the relations holding between the physical events remain outside its domain of interactions. The nervous system enlarges the domain of interactions of the organism by making the internal states of the living systems also modifiable in a relevant manner by *pure relations*, and not only by physical events: the observer sees that the sensors of an animal (say, a cat) are modified by light, and that the animal (the cat) is modified by a visible entity (a bird). The sensors change through a physical interaction, the absorption of light quanta; the animal is modified through its interaction with the relations that hold between the activated sensors that absorbed light quanta at the sensory surface. The nervous system expands the cognitive domain of the living system by making it possible for it to interact with *pure relations;* it does not create cognition.

Although the nervous system expands the domain of interactions of the organism by bringing into this domain interactions with *pure relations*, the function of the nervous system is subservient to the necessary circularity of the living organization.

The nervous system, by expanding the domain of interactions of the organism, has transformed the unit of interactions and has subjected acting and interacting in the domain of *pure relations* to the process of evolution. As a consequence there are organisms that include as a subset of their possible interactions, interactions with their own internal states (as states resulting from external and internal interactions) as if these were independent entities, generating the apparent paradox of including in their cognitive domain their cognitive domain. In us, this paradox is resolved by what we call abstract thinking (a new cognitive domain).

Furthermore, the expansion of the cognitive domain into domains of pure relations by virtue of a nervous system allows for nonphysical interactions (representations) between organisms that orient each other toward other interactions. Herein lies the origin of communication. But this very process generates another apparent paradox: there are organisms that generate representations of their interactions by specifying entities with which they interact as if these entities belonged to an independent domain, while as representations they only map their own interactions. In us this paradox is resolved simultaneously in two ways.

1) By our becoming observers, we recursively generate representations of our interactions; and by interacting with their relations we remain in a domain of interactions that is always larger than that of the representations.
2) By our becoming conscious, we make descriptions of ourselves (that is representations); and by interacting with our descriptions we can describe ourselves describing ourselves—in an indefinite recursive process.

COGNITIVE PROCESS IN PARTICULAR
(The Nervous System)

Nerve Cells

The neuron is the anatomical and functional unit of the nervous system. It is the anatomical unit because it is a cell and as such it is an integrated and self-referring metabolic and genetic unit (a living system indeed). It is the functional unit because as a collector and as a distributor of influences it functions as a unit.

Anatomically and functionally the neuron is formed by a collector (dendrites and sometimes the cell body) united via a distributive element (the axon, and in some cases, also the cell body and main dendrites), capable of conducting propagated spikes to an effector area formed by the terminal branching of the axon. The functional state of the collector area depends both on its internal state (reference state) and on the state of activity of the effector areas synapsing on it. Correspondingly, the state of activity of the effector area of a neuron depends on the state of activity of its collector area and on the presynaptic and nonsynaptic interactions with other effector areas that may take place in the neuropil and in the immediate vicinity of the next collector areas. This is true even in the case of amacrines cells, in which collector and effector areas may be intermingled. The distributive element determines where the effector exerts its influence.

At any moment the state of activity of a nerve cell is a function of the spatiotemporal configuration of its input, as determined by the relative activity holding between the afferent neurons, that modulates the reference state proper of the collector area. It is known that in many neurons the recurrence of a given afferent spatiotemporal configuration results in the recurrence of the same state of activity, independently of the way in which such a spatio temporal configuration is generated.[2,3] This is the understanding that two states of activity in a given cell are the *same* (equivalent) if they belong to the same class, as defined by the pattern of activity that they generate in their effector area, and not because they are a one-to-one mapping of each other. Also, the spatiotemporal configuration of the input on a neuron that causes in it the recurrence of a given state of activity is a class of afferent influences defined by a pattern in the relations holding between the active afferents and the collector. Thus, there are neurons for which a given class of response is elicited by a given class of afferent influences.

The nerve impulses that travel along the distributive element originate at the point where this element emerges from the collector area. Each nerve impulse is the result of the state of excitation of the collector area at a given moment (as determined by the spatiotemporal configuration of the afferent excitatory and inhibitory influences acting on it and, if any, on its own internal generating mechanism) that spreads, reaching a given threshold at the point of emergence of the distributor. Excitatory and inhibitory influences, however, do not superimpose linearly; their relative participation in determining the production of nerve impulses, and hence the state of activity of the neuron, depends on their relative spatial distribution on the collector area. Inhibition works by shunting off the spreading excitatory processes. As a result, the relative contributions of a point of excitation and of a point of inhibition in the generation of a nerve impulse depend on where, on the collector, they stand with respect to each other and with respect to the point of emergence of the distributive element. Excitation and inhibition have to be seen as integral parts in the definition of the spatiotemporal configuration of afferent influences, not as independent processes. The shape of the collector area (its geometry) has to be seen as selecting the spatiotemporal configuration of afferent influences to which the cell will respond.

Although I expect in a significant number of neurons a continuous change in their transfer functions as a result of their past history, I think that for the understanding of the functional organization of the nervous system it is necessary at any given moment to consider all nerve cells as responding with definite transfer functions to classes of afferent spatiotemporal configurations in their input, and not to particular situations. This because of two reasons:

1) I consider that any interaction through the nervous system is represented by a state of activity in a collection of cells, and that this state of activity should lead to a given behavior. This behavior should be repeatable to the extent that the interaction is reproducible.
2) I consider that the nervous system always functions in the present. (The present is the time interval necessary for an interaction to take place; *past, future,* and *time,* in general, exist only for an observer.) Although nerve cells may be continuously changing their mode of operation, their past history can explain to an observer how their present mode of operation was reached, but not its present participation in the determination of behavior.

If nerve cells respond to classes of events, not to particular situations, they necessarily treat as equivalent particular afferent configurations that are otherwise unrelated.

Architecture

In any given nervous system the great majority (and perhaps the totality) of its neurons can be assigned to well defined morphological classes, each characterized by a given pattern of distribution of the collector and effector areas of its members. As a result, the members of the same class hold similar relations with each other and with other neurons: the shapes of the nerve cells (collector area, distributive conductor, and effector area) specify their connectivity. These shapes are genetically determined and have been attained through evolution; the whole architecture of the brain is genetically determined and has been attained through evolution. The following implications, I think, are significant for the understanding of the functional organization of the nervous system.

1) There is a necessary genetic variability in the shape of nerve cells, as well as a variability that results from interactions of the organism with independent events during development. The functional organization of the nervous system must be such as to tolerate this variability.
2) No two nervous systems of animals of the same species (particularly if they have many cells) are identical, and they resemble each other only to the extent that they are organized according to the same general pattern. It is the organization defining the class—and not any particular connectivity—which determines the mode of functioning of any given kind of nervous system.

The shapes of nerve cells and their packing are such that there is in general a great overlapping in the collector and effector areas of neurons of the same class. Also, the spatial distribution and the interconnections between different classes of neurons is such that any particular part of the nervous system is in general simultaneously related to many other parts; the parts interconnected, however, differ in different species, and, as a result, they have different interacting capabilities.

The organism ends at the boundary that its self-referring organization defines in the maintenance of its identity. At this boundary there are sensors through which the nervous system interacts in the domain of relations. These sensors are in general constituted by collections of sensory elements (cells) with similar—though not identical—properties (classes of properties), which in their mode of interaction with the nervous system share the characteristics of neurons in general. As a result, whenever the organism enters into an interaction within the physical domain of interactions of the sensors, generally not one, but many sensory elements are excited.

Function

The way the nervous system functions is bound to its anatomical organization. The functioning of the nervous system, however, has two aspects: one which refers to the domain of interactions defined by the nervous system (relations in general); the other which refers to the particular part of that domain used by a given species (particular classes of relations); different species interact with different sets of relations (have different niches).

The nervous system only interacts with relations. However, since the functioning of the nervous system is anatomy bound, these interactions are necessarily mediated by physical interactions. For animals to discriminate objects by vision, the receptors in the eyes must absorb light quanta to be activated; yet the objects that the animal sees are determinded by the relations of receptor-induced activity within the functional and anatomical organization of the retina. Moreover, since the domain of interactions of the organism is defined by its structure, and this structure implies the prediction of the niche, the relations with which the nervous system interacts are defined by this prediction and arise in the domain of interactions of the organism. The (anatomical and functional) organization of the living system defines a *point of view,* a bias or a posture from which the interactions take place, and that determines the possible relations accessible to the nervous system. The anatomy of the retina and the properties of the various cell types define which relations, holding between the active receptors when a given visual object is viewed, will be accessible to the nervous system.

Due to the properties of neurons, and due to the architecture of the nervous system, interactions within the nervous system necessarily give origin to activity in aggregates of cells. Also, for the same reasons, any given cell may assume the same state of activity under many different circumstances of interactions of the organism. Thus, under no circumstance is it possible to associate the activity of any particular cell to any patricular interaction of the living system. (This can be clearly shown by discussing the visual system.[4]) When any particular interaction takes place at the level of the sensors, the relations accessible to the nervous system are given at this level in a certain state of relative activity of the sensing elements and not in the state of activity of any particular one. At the same time, although operational localizations can be established in the nervous system,[5] these localizations are to be viewed in terms of areas where certain modalities of interactions converge, and not as localizations of faculties or functions, even though certain functions cannot be performed in the absence of these areas. The nervous system is localized in terms of the organism's interaction surfaces, but not in terms of the representations of the interactions it can generate.

Representation

The fundamental anatomical and functional organization of the nervous system is basically uniform: the same functions and operations (excitation, inhibition, lateral interaction, recursive inhibition, etc.) are performed in its various parts, although in different contexts and integrated in a different manner. A partial destruction of the nervous system does not alter this basic uniformity and, although the parts left untouched cannot do the same things that the whole does, they appear in their mode of operation identical to the untouched whole. To the observer, once the boundary of the sensors is passed, the nervous system, as a mode of organization, seems to begin at any arbitrary point that he may choose to consider: the answer to the question "What is an input to the nervous system?" depends entirely on the chosen point of observation. This basic uniformity of organization can best be expressed by saying: all that is accessible to the nervous system at any point are states of relative activity holding between nerve cells, and all that to which any given state of relative activity can give origin are further states of relative activity in other nerve cells by directly conforming those states of relative activity to which they respond. The effector neurons are not an exception to this since they, by causing an effector activity, cause a change in the state of relative activity of the receptor elements at the receptor surfaces, by generating an interaction. This has a fundamental consequence: unless they imply their origin (through concomitant events or their locations), there is no possible distinction between internally and externally generated states of nervous activity.

The relations with which the nervous system interacts are relations given by the physical interactions of the organism, hence depend on its anatomical organization. For the observer the organism interacts with a given entity that he can describe in his cognitive domain. Yet, what modify the nervous system of the observed organism are the changes in activity of the nerve cells associated with the sensing elements, changes that henceforth constitute an embodiment of the relations that arise through interaction. These relations are not those that the observer can describe as holding between the component properties of the entity in his cognitive domain; they are relations generated in the interaction itself and depend on both the structural organization of the organism and the properties of the entity that match the domain of interactions that this organization defines.

Whenever such a relation recurs at the sensory surface, the same state of relative activity arises among the neurons in contact with the sensing elements. Two interactions that produce the same state of relative activity are identical for the nervous system, no matter how different they may be in the cognitive domain of the observer.

Any given state of activity of the nervous system consists of states of relative activity holding between neurons. If an interaction takes place, the state of activity of the nervous system is modified by the change in relative activity of the neurons which, in close association with the sensing elements, embody the relations given in the interaction. What is then represented in the different states of activity of the nervous system are the relations given in the interactions of the organism at its sensory surfaces and not an environment independent of it (least of all a description—in terms of entities—that lies in the cognitive domain of the observer).

Every relation is represented in a state of relative activity of nerve cells, but also, every state of relative activity acts modifying the relative activity of other nerve cells. Relations, thus, through their embodiment in states of relative activity become units of interactions and generate additional relations, again embodied in states of relative activity which can in turn become units of internal interactions in the nervous system.

The classes of relations that can be represented have been defined by evolution:

1) by the evolution of the structural organization of the organism in general, and of the sensors in particular, that define the classes of relations that will be accessible to the nervous system,
2) by the evolution of the particular organization of the nervous system, corresponding to that of a given class of animals (species), that defines the mode how these relations generate a behavior relevant to the maintenance of the organism.

Of a given class of relations the particular relation encountered as a result of a present interaction is represented by a particular state of activity given in the present. This is independent of history. However, the relevance of the behavior generated by this state of activity for the maintenance of the living system is history dependent, and may depend on both the evolutionary history of the species and the past experience of the organism. In the former case, one would talk of instinctive behavior and in the second case, of learning. The description of learning in terms of past and present behavior lies in the cognitive domain of the observer. The organism always behaves in the present. The observer, however, by generating a description can treat as if in the present interactions which now do not recur. This apparent paradox he resolves generating the notion of time: past, present, and future.

There is no difference in the nature of the representation of internally and externally generated interactions. In an organism capable of interacting with the embodiment of his own interactions the nature of the representation does not change. What changes are the relations represented.

Since the nervous system interacts only with states of relative activity holding between neurons, and since a perception and a representation are states of relative activity holding between neurons, there is no objection in principle for the nervous system to interact with representations of its interactions. The distinction between the two kinds of interactions can only arise through the concomitance of events that indicate the source of the state of relative activity that embody them, or through the outcome of the new interactions which they initiated. A nervous system that is capable of treating its internally generated states as it treats its externally generated states (that is, distinguishing their origin) is capable of abstract thinking.

Description

If a given state of relative activity in the nerve cells originates a given behavior, the recurrence of the same state of relative activity should give origin to the same behavior, no matter how the recurrent state originates. The relevance of such a behavior is determined by the significance that it has for the maintenance of the living organization. The living system, due to its circular organization, is an inductive system and functions always in a predictive manner: what occurred once will occur again. Its organization (both genetic and otherwise) is conservative and repeats only that which works. For this same reason living systems are historical systems: the relevance of a given conduct or mode of behaving is determined always in the past. The goal state (in the language of the observer) that controls the development of an organism is—except for mutations—determined by the parent organism. In these circumstances, then, a behavior is relevant if it allows the organism to behave again in a similar manner, and does so by maintaining its basic circularity. With the expansion of the cognitive domain during evolution the types of behavior have changed as well as how they implement their relevance: different kinds of behavior are relevant to the maintenance of the basic circularity of the living organization through different domains of interactions.

Through a given interaction a new interaction is generated whose relevance arises from the way the system maintains its circular organization, neutralizing the disturbances introduced by the first one and giving origin to new interactions. To the observer the second interaction appears as a description of the niche of the observed organism. And, indeed, the niche of an organism is the set of all the interactions into which it can enter; hence, the cognitive domain of an organism is a description of its niche. This description, however, is a description in terms of interactions into which the organism can enter, and not in terms of representations of environmental states.

An organism can modify the behavior of another organism in two basic ways:

1) by interacting with it in a manner such that the behaviors of the two organisms are directed to the interaction but each oriented from his own perspective, i.e. courtship. A chain of interlocked behavior can thus be produced between the two organisms;
2) by orienting the behavior of the other organism to some part of his domain of interactions different from the present interaction itself, but comparable to the orientation of the orienting organism. This can take place only if the domains of interactions of the two organisms are widely coincident; in this case no interlocked chain of behavior is elicited because the subsequent conduct of the two organisms depends on the outcome of independent, though parallel, interactions.

In the first case it can be said that the two organisms interact, in the second case that they communicate. The second case is the basis for any linguistic behavior: the first organism generates (as is apparent to the observer) a communicable implicit description of the common niche by a behavior that orients the other organism within his domain of interactions, and elicits in it a conduct that the first could also have had but which is relevant independently of him. The elicited conduct, to the extent that it is an interaction into which the second organism enters, is a description of the niche. To the observer, the eliciting behavior is denotative, he can treat it as if it were pointing to a feature of the niche that can be described by the appropriate behavior. To the communicating organism the eliciting behavior is connotative, it implies the behavior that describes that particular aspect of the niche.

Linguistic behavior is orienting behavior (it orients the listener within his cognitive domain in a manner comparable to the orientation of the speaker) and its evolutionary history must be interdependent with the evolutionary increase in the complexity of the cognitive domain (the niche), through tool making, for example, in strongly social animals. (The use of tools recursively increases the domain of interactions of an organism.)

During evolution only those interactions that are relevant to the maintenance of the circular organization of the living system are incorporated in his cognitive domain. As a consequence, the laws of interaction that govern the behavior of the organism (identity, distinction, order, etc.) are the same that it can encounter in its interactions.

If an organism can generate a communicable description of its interactions and then interact with the communicable description, the process can, in principle, be carried on in a potentially infinite recursive manner, and the organism becomes an observer. It can describe its interactions and communicate its descriptions to others or to itself, and through this very same process it can describe itself describing itself. Thus, discourse, originated through communicable descriptions, generates the apparent paradox of consciousness as a domain of pure self-description (hence, self-observation).

The Observer: Epistemological and Ontological Comments

The cognitive domain is the entire domain of interactions. This cognitive domain can be enlarged if new modes of interactions can be generated. Instruments enlarge our cognitive domain.

The possibility of enlargement of the cognitive domain is unlimited. Our brain, the brain of the observer, has specialized during evolution as an instrument to discriminate relations; both internally and externally generated relations, but relations given through and by the interactions. We cannot say in absolute terms what is the *input* to our nervous system (the nervous system of the observer) because *every* state of the nervous system can simultaneously be the input as well as the receiver and, hence, every change of state modifies the nervous system as an interacting unit.

We can say that every interaction changes us because it modifies our internal state, changing the posture or perspective from which we enter into a new interaction. This necessarily creates new relations that we may detect.

The observer generates a spoken description (orienting behavior) of his cognitive domain, which includes his interactions with instruments and through instruments. Whatever description he makes, however, the description is a set of permitted states of relative activity in his nervous system embodying his interactions. These permitted states of relative activity are made possible by the anatomical and functional organization of his nervous system, which, in turn, has evolved as a system subservient to the basic circularity of the living organizations, and, hence, embody an inescapable logic: that logic that allows for a match between the organization of the living system and the interactions into which it can enter without losing its identity.

The observer can describe a system that gives origin to a system that can describe, hence, to an observer. This domain of discourse is a closed domain; it is not possible to step outside of it. Because it is a closed domain it is possible to make the following ontological statement: the logic of the description is the logic of the describing (living) system (and his cognitive domain).

This logic demands a substratum for the occurrence of the discourse. We cannot talk about this substratum in absolute terms, however, because we would have to describe it, and a description is a set of interactions into which the describer and the listener can enter, and their discourse about these interactions will be another set of possible interactions that will remain in the same domain. Thus, although this substratum is required for epistemological reasons, nothing can be said about it other than what is meant in the ontological statement made above.

For epistemological reasons we can say: there are properties; they are manifold and remain constant through interactions. The invariance of properties through interactions gives functional origin to entities or units of interactions, and, since entities are generated through the interactions that define them, they originate independent relational domains: no reductionism is impossible.

AIMS OF NEUROPHYSIOLOGY OF COGNITION

The observer can always remain in a domain of interactions encompassing that of the observed entity. The observer has a nervous system. He is able to interact with his own interactions and, hence, he is able to interact with (observe) his description of his niche. He can do this because in the general mode of organization of the nervous system there is no intrinsic difference between internally and externally generated states of nervous activity.

If the genetically determined (evolved) neurophysiological processes which permit orienting behavior (communicative behavior) do not specify particular orientations, but secure them in general, the organism can learn to orient itself to itself, that is, it can describe itself (self-description). Through describing itself in a recursive manner, the organism generates consciousness. Consciousness, then, is not a neurophysiological phenomenon, it is an epiphenomenon of orienting behavior that lies entirely in the linguistic domain. The consequences of this are twofold.

1) The linguistic domain as a domain of orienting behavior requires not only two interacting organisms, but it also requires a basic coincidence in the domain of interactions of the latter, so that they may indeed orient each other to comparable interactions in their respective cognitive domains. Also, the specifiability of content of orienting behavior allows for a purely consensual (cultural) evolution in the domain of orienting behavior, without necessarily involving in it a further evolution of the nervous system: the contents of the linguistic domain are specified by agreement between the organisms interacting in this domain. For these reasons the linguistic domain in general, and consciousness in particular, appear to be independent of the biological substratum that generates them. This independence, however, is not complete, and the linguistic and conscious domains in man are biologically limited: (*a*) by the neurological and neurophysiological organization of the brain which specifies both, the domain of possible interactions through determining the actual possibilities for the confluence of different sensory modalities, and the complexity of the patterns of orienting behavior it can distinguish; (*b*) by the necessary subservience of these domains to the maintenance of the identity of the organism (its basic circular organization).
2) If consciousness (and the linguistic domain in general) is not a neurophysiological phenomenon, but arises as an epiphenomenon from orienting behavior, the questions that one should then ask neurophysiology refer to the synthesis of meaningful behavior in general and of orienting behavior in particular. In other words, the encompassing neurophysiological questions become: How does the nervous system interact with its own internal states (states of rela-

tive nervous activity that represent internal or external interactions), and is modified by them as if they were independent entities? And, how does the nervous system generate interactions which specify states of relative nervous activity which it can treat as entities (descriptions of the niche) with which it can interact in a recursive manner? In a system with a self-referred cognitive domain like a living system, meaning arises in relation to the interactions to which a given interaction gives origin by a sequence of processes through which the organism passes without losing its identity. In these circumstances, the concrete task is to elucidate how a given effector performance is synthesized so that it brings forth an interaction which, in turn, specifies a given state of activity at the receptor surfaces (that in turn specifies a state of relative activity in the subjacent neurons, and so on). How does this state of activity interact with the internal states of the nervous system to synthesize a new effector performance? All this is occurring in a deterministic manner akin to reflex processes, but different from these in that both the states of relative activity involved and the immediate relevance of any interaction for the generation of new interactions (meaning) are submitted to change through interactions.

Although there is the fundamental core of structural organization in the nervous system that is genetically and environmentally determined (during development, maturation, and under hormonal actions), which specifies the possible modes of conduct, the overall behavior of the organism should be looked at as a synthetic process; a synthetic process in which the relative weight of the interactions of the organism in the niche and its changing internal states in the generation of new behavior is determined by the past history (behavioral and developmental) of the organism. Thus, there are two complementary aspects to the understanding of the neurophysiology of cognition:

1) one that considers the nervous system as a state-determined system performing operations that in principle are indistinguishable from reflex processes, and which actually realize the behavioral processes that take place[6]
2) one that considers how behavior acquires meaning by changing the transition rules (the relative weight of the different states of nervous activity) of the state-determined nervous system through actual behavior and survival. Consciousness, as all behavior that can be generated through language, arises in this manner.

ARTIFICIAL INTELLIGENCE

To my understanding, the aim of research in artificial intelligence is to make an artificial cognitive system. Since much of what has been said above applies to machines, the following comments are ventured.

Machines differ from living systems not in the principles used for their function, but in their reference. The organization of machines is such that they always have an alloreferential domain of interactions, that is, the relevance of their functioning is determined by how they satisfy the desires of their maker. Living systems, on the contrary, as a result of their basic circular organization, necessarily have a self-referred domain of interactions; that is, the relevance of their functioning is determined by how it permits the maintenance of their basic circularity. If we consider the ordinary machines as alloreferential machines, we can consider the living systems as self-referential machines.

The artificial cognitive systems can be of two kinds:

1) those that will describe their interactions in our terms, that is, recognize what we recognize
2) those that will make descriptions in their own terms, but which we have to map afterward into ours.

In either case, I think, there is no need for imitating what occurs in our brain. In the first case, however, it would be essential to give the machine a domain of interactions like ours, not our description of it.

POSTSCRIPTUM

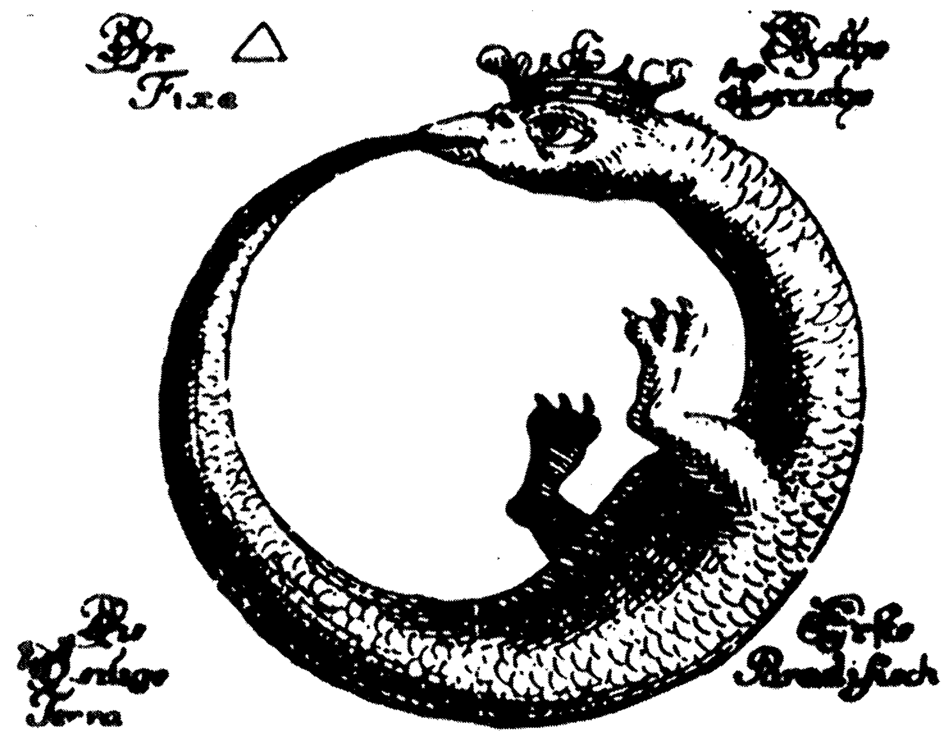
There are numerous experimental studies and observations that show the neurophysiological processes associated with complex reflexes.[7] Furthermore, much is known about networks and feedback interconnections associated with motor performances, particularly in relation to the cerebellum[8] and spinal cord. Also, much has been investigated in relation to the reticular formation and its possible role in the control of behavior.[9] The detailed list of specifiable modes of behavior that could be provided seems endless—how long must it grow? There are questions for which it is impossible to provide an answer because they are made in one relational domain and the answer must be expressed in the terms proper to another. Something like this, I think, has occurred in the study of the nervous system. The sum total of its component operation does not seem to account for the functioning of the whole, either because we have not yet discovered how certain operations are performed, or because the functioning of the whole belongs to a different phenomenological domain. The latter is obviously my view. It is stressed here, however, because of two reasons.

1) The consequences of considering the higher functions of man (and mammals in general) as being outside of the neurophysiological domain are paramount. It is not necessary to account for them in terms of neurophysiological process networks, spike potentials or whatever is of this domain; we have to account for the much simpler deterministic behavioral process which will give rise to them as a consequence of their operation in the domain of interactions between organisms. The nervous system is a state-determined system whose transition rules change and become specified through interactions that satisfy the basic circularity of the living organization.
2) There are different domains of interactions; these different domains cannot explain each other, that is, it is not possible to describe the phenomena of one domain with the notions proper to another. Any nexus between different domains is provided by the observer who can interact with the states of nervous activity generated in his brain by his concomitant interactions in several different domains, as if these conjoined states of activity constituted an independent entity (a relation). No reductionist position is tenable, and the number of such different and independent domains is, in principle, unlimited. A grasp of this situation is essential for the understanding of biological phenomena, their epiphenomena, and their evolution.

It was sometimes said during the conference that in the group there were no other biologists, hence no one could oppose my views. With this publication this, at least, is not true anymore.

REFERENCES

1. H. Von Foerster, "From Stimulus to Symbol," *Sign, Image, Symbol,* ed. G. Kepes (New York: George Braziller, Inc., 1966), pp. 42–61.
2. H. R. Maturana and S. Frenk, "Directional Movement and Horizontal Edge Detectors in the Pigeon Retina," *Science,* 142 (1963), 977–79.
3. F. Morrell, "Electrical Signs of Sensory Coding," *The Neurosciences, A Study Program,* ed. G. C. Quarton *et al.* (New York: The Rockefeller University Press, 1967), pp. 452–68.
4. H. R. Maturana, G. Uribe, and S. Frenk, "A Biological Theory of Relativistic Colour Coding in the Primate Retina," *Arch. Biologia y Med. Exp.,* Supplemento No. 1 (Santiago: University of Chile, 1968).
5. N. Geschwind, "Disconnexion Syndromes in Animals and Man," *Brain,* 88 (1965), 237–93.
6. W. R. Ashby, *Design for a Brain* (New York: John Wiley & Sons, 1952).
7. G. Horridge, *Interneurons, Their Origin, Action, Specificity, Growth, and Plasticity* (London: W. H. Freeman and Company, 1968).
8. J. C. Eccles, M. Ito, and J. Szentagothai, *The Cerebellum as a Neuronal Machine* (New York: Springer-Verlag, 1967).
9. W. L. Kilmer, *The Reticular Formation, Part II: The Biology of the Reticular Formation* (Interim Scientific Report No. 3, Division of Engineering Researching, Michigan State University [East Lansing, Michigan, 1969]).



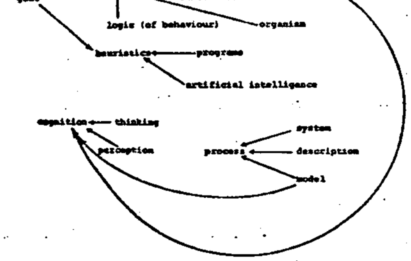
REPRODUCTION

In biology, "reproduction" makes reference to three distinct processes. Replication is the production of entities by a mechanism that is distinct from these entities. In replication there is a historical uncoupling of the entity produced from the mechanism that produces it. Copy refers to a process that is identical to a mapping operation where there is a mechanism for isomorphic transformation. Again, the mechanism for producing the entity is distinct from the entity produced. Self-division refers to a process that is peculiar to autopoietic systems. In this process, the unity produced is identical in its organization to the original unity, and it is a result of the maintenance of the original unity's autopoiesis. [K.L.W.]

EVOLUTION

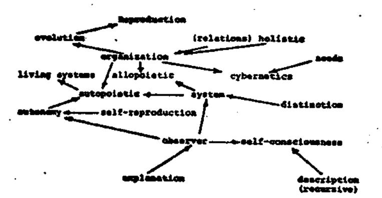
The useful use of uncertainty through a process wherein some degree of randomness modifies the logic of behavior thus producing a mutation which is then tested for survival value against the embedding environment. Those logics which survive are in turn affected by randomness, and thus mutated and themselves tested for survival. The result yields either an extinction of the original logic or some modified logic which fits an "ecological niche." In a dynamic environment some continual modification of the logic is required for survival of the species. Thus, too little randomness can result in extinction. But, excessive randomness can also produce extinction. Most random changes are dysfunctional.

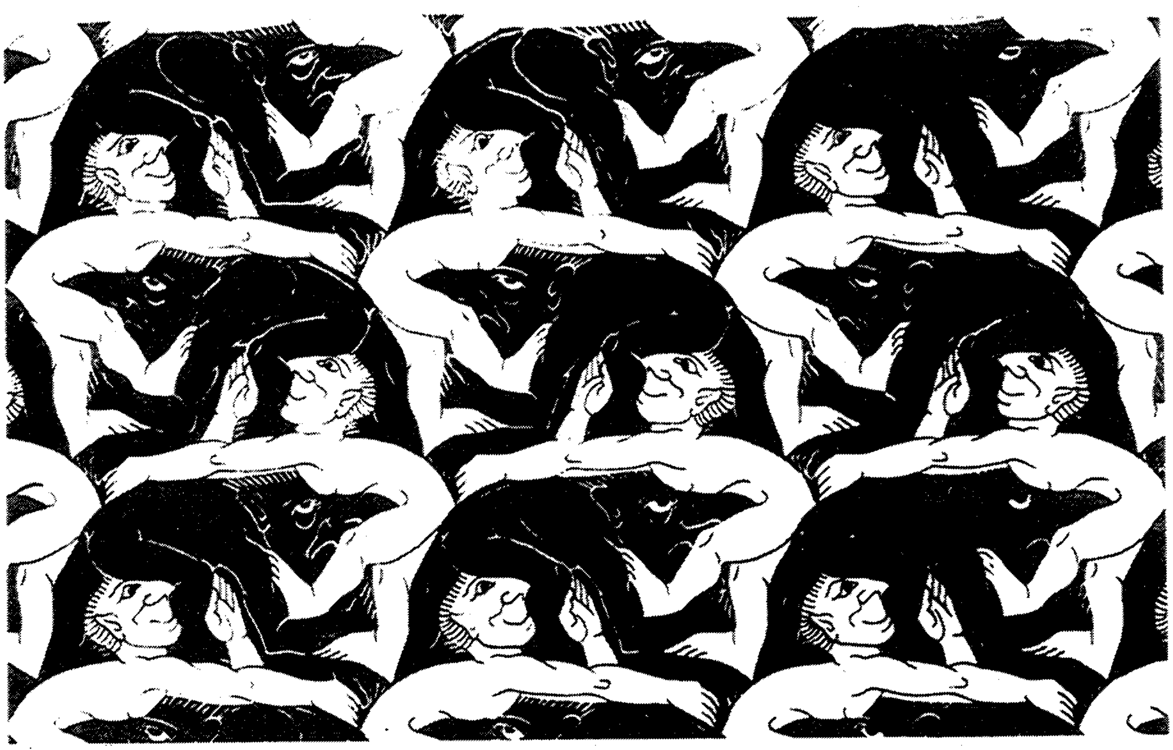
With sufficient intelligence evolution can be partially simulated and controlled. In this manner creatures become more and more specialized and therefore more significantly dependent upon the suitability of their mutation in a changing world. These same factors apply in the biological, psychological, and sociological levels. In summary this is a goal-seeking process with survival being the goal and the exploration being conducted in a grossly insufficient manner, yet herein lies the germ of imagination, ingenuity, and creativity. [L.F.]



EVOLUTION

A history of change of a type of organization as embodied in a population of distinct systems. The systems must be related by sequential reproductive steps, in which the particular structure of each one is a modification of the structure of the preceeding one, which is its historical predecessor. Sequential reproduction and change of structure in each reproductive step are the necessary and sufficient conditions for evolution. (Thus there is a history of the earth, but not an evolution of it; there is an evolution of the idea of space, not only a history of it). [F.V.]





SCIENTIFIC METHOD

Science deals with facts, and tries to generate theories about the facts; theories which in turn are supported or refuted by facts. Assuming that a sharp distinction may be postulated to exist between "facts" and "theories", "science" may proceed in a number of ways. One involves first fact-gathering by observation or experiment, followed by "getting" (somehow extracting) a hypothesis or theory from the facts. Alternatively, one may first generate a hypothesis or conjecture, and then "test" it by gathering evidence in the form of "relevant" facts, observations, or experiments.

Since an infinite number of facts could conceivably be collected, the first method encounters a problem of relevance, as well as the necessity of elucidating some method by which induction may occur, with resultant hypothesis generation. Since only a finite number of observations can be made, conclusions resulting from <u>de</u>duction will cover a broader range of cases than does the evidence. So from statements like "Cybernetician A is incomprehensible," "cybernetician B is incomprehensible," cybernetician C is incomprehensible," which cover only <u>some</u> of the cases, one arrives at "all cyberneticians are incomprehensible," which, although perhaps true, is not a compelling conclusion, given the some all jump. Yet our laws of physics make this same jump ("all bodies fall at a given rate...") indicating the possible validity and rationality of some generalization of this type. But this is another problem.

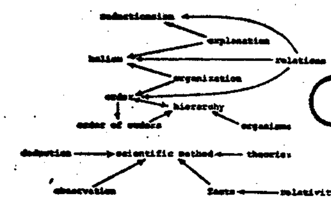
The second alternative presented also encounters problems with regard to distinguishing "relevant" facts from the wealth of possibilities. Objectivity may well be lost and bias introduced by the experimenter as he makes such choices. It may be argued that the problem itself distinguishes relevant and irrelevant facts, but the implication is then that formation of the problem may already presuppose an assumption, for a "problem" only exists if some prior expectation has been violated. Completeness too can never be a-sured, since one has no way of being certain she has amassed <u>all</u> the relevant facts. This method seems more concerned with a search for ultimate causes, while the first simply relates observed phenomena.

Both methods aside, one may wonder whether the study of scientific method is itself relevant. Does science in truth proceed in a methodological way? Any answer may depend on the level at which one makes observations. Elucidation of the structure of DNA, for example, was the culmination of much analysis with a common focus and eventual success. Progress was semi-stepwise, with Pauling's α-helix and Rosalind Franklin's X-ray crystallography necessary precursors of Watson and double Helix. Hypotheses were made, tested and discarded or incorporated as warranted by the results. As the amound of information amassed increased, eventual success became a foregone conclusion, and the analysis became almost a race, with a Nobel prize as reward. Such is the nature of the method which led science to one of its most significant discoveries. Kelkule, on the other hand, elucidated the 6 carbon ring structure vital to organic chemistry by using the "method" of going to sleep--a dream of snakes biting each other's tails suggested the structure. Perhaps the methods in the two cases are not fundamentally different; both seem worthy of consideration.

"It has been remarked over and over again how harmoniously the whole organism hangs together, and how throughout its fabric one part is related and fitted to another in strictly functional correlation. But this conception, though never denied, is sometimes apt to be forgotten in the course of that process of more and more minute analysis by which, for simplicity's sake, we seek to unravel the intracacies of a complex organism.

"As we analyze a thing into its parts or into its properties, we tend to magnify these, to exaggerate their apparent independence, and to hide from outselves (at least for a time) the essential integrity and individuality of the composite whole...

"The biologist, as well as the philosopher, learns to recognize that the whole is not merely the sum of its parts. It is this and much more than this. For it is not a bundle of parts but an organization of parts in their mutual arrangement, fitting one with another, in what Aristotle calls 'a single and indivisible principle of unity'; and this is no mere metaphysical conception; but is in biology the fundamental truth which lies at the basis of Geoffrey's (or Goethe's) law of 'compensation,' or 'balancement of growth.'" D'arcy Thompson (from On Growth and Form) [S.H.]



SCIENTIFIC METHOD

The scientific method offers an input-throughput-output algorithm for logical approach to empirical questions. This algorithm is usually described as the information flow through the following steps:

1. State or define the problem
2. Collect relevant information
3. Develop an hypothesis
4. Design and conduct an appropriate experiment (empirical test of hypothesis)
5. Analyse data or experimental results.

Step 5 is designed to lead to reformulation of the problem→ the algorithm repeats. Such an ideal information flow procedure tends to behavioristically describe the rational-empirical process of scientific inquiry, but overlooks a significant cognitive factor in empirical information processing--intuition. [D.S.]

CYBERNETICS OF CYBERNETICS (PHYSIOLOGY OF REVOLUTION)

Heinz Von Foerster

The existence of something called "Social Sciences" is indication of the refusal to let sciences be social. This holds, of course, also for the Social Sciences. The corruption of Science so as to become an activity conducted in a social vacuum began with the onset of a contagious delusion, namely, that the business of Science is to arrive at "objective statements", as if there could be such things. Clearly, subjective statements are made by subjects; and correspondingly, objective statements are made by objects--only these damn things are not going to talk. Should Science be able to talk again it better be social. Now, remember Gordon Pask ("The Meaning of Cybernetics in the Behavioral Sciences", p. 25):

> **Taciturn and Language Oriented Systems**
>
> Taciturn systems are those for which the observer asserts or discovers the goal (purpose *in*), which is thereafter equated with the purpose *for* the system in question. In contrast, language oriented systems can be asked or instructed to *adopt* goals by anyone who knows the object language and they may state and describe their own goals, using the same medium; in a very real sense these are "general purpose" systems.

It is at this point where we mature from cybernetics (where the observer enters the system only by stipulating its purpose) to cybernetics of cybernetics (where the observer enters the system by stipulating his own purpose). And also remember G. Spencer Brown (Laws of Form, p. 104):

> Let us then consider, for a moment, the world as described by the physicist. It consists of a number of fundamental particles which, if shot through their own space, appear as waves, and are thus (as in Chapter 11), of the same laminated structure as pearls or onions, and other wave forms called electromagnetic which it is convenient, by Occam's razor, to consider as travelling through space with a standard velocity. All these appear bound by certain natural laws which indicate the form of their relationship.

Now the physicist himself, who describes all this, is, in his own account, himself constructed of it. He is, in short, made of a conglomeration of the very particulars he describes, no more, no less, bound together by and obeying such general laws as he himself has managed to find and to record.

Thus we cannot escape the fact that the world we know is constructed in order (and thus in such a way as to be able) to see itself.

This is indeed amazing.

Not so much in view of what it sees, although this may appear fantastic enough, but in respect of the fact that it *can* see *at all*.

But *in order* to do so, evidently it must first cut itself up into at least one state which sees, and at least one other state which is seen. In this severed and mutilated condition, whatever it sees is *only partially* itself. We may take it that the world undoubtedly is itself (i.e. is indistinct from itself), but, in any attempt to see itself as an object, it must, equally undoubtedly, act* so as to make itself distinct from, and therefore false to, itself. In this condition it will always partially elude itself.

Thus the world, when ever it appears as a physical universe*, must always seem to us, its representatives, to be playing a kind of hide-and-seek with itself. What is revealed will be concealed, but what is concealed will again be revealed. And since we ourselves represent it, this occultation will be apparent in our life in general, and in our mathematics in particular.

In this game of hide-and-go-seek we must be prepared to continuously flip between two approaches that are created by the elementary act of drawing a distinction:

DRAW A DISTINCTION

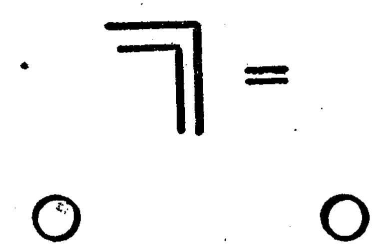
(A space is severed)

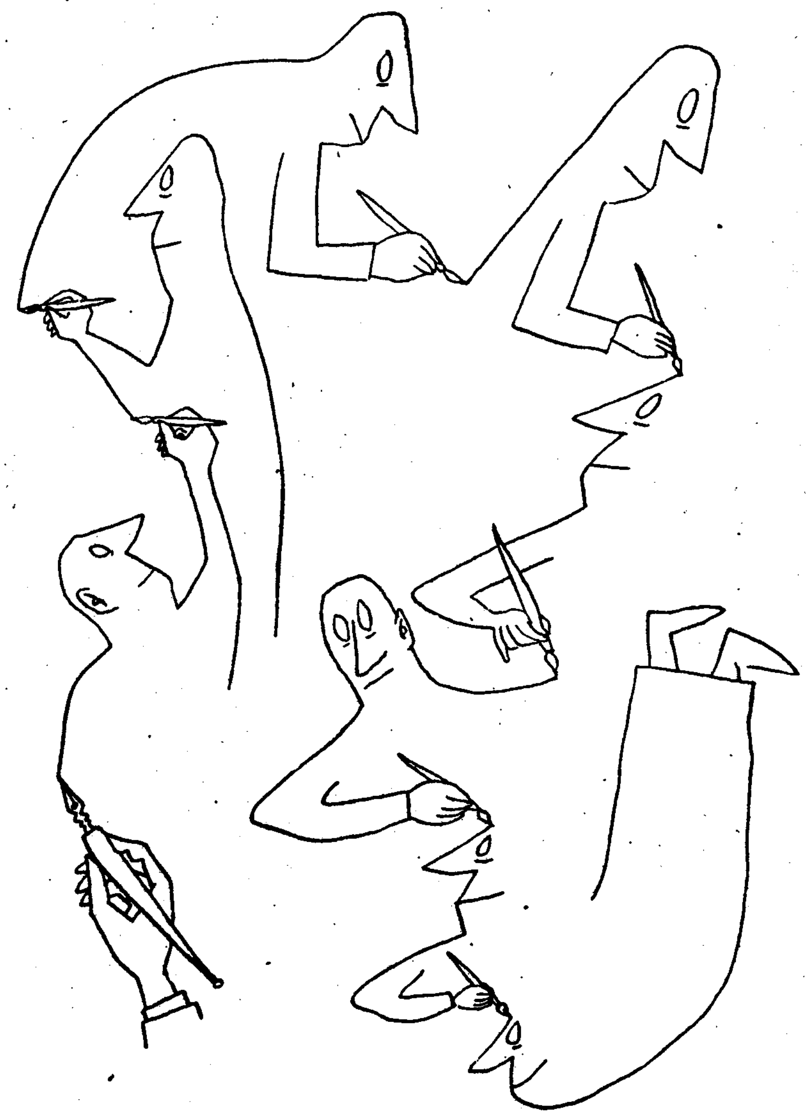
Creation	Parts are created Relations are cut	Relations are created Parts are isolated
Approach	Reductionism	Holism
Attend to	Parts	Relations
Paradigm	Set Properties of Elements	Computation Properties of Process
Causality	Linear	Circular
Structure	Open	Closed
Organization	Allopoiesis	Autopoiesis
Regulation	Heteronomy	Autonomy
Observer	Excluded	Included

SYSTEMIC MONISM

(Pask, ibid, p. 19)

Since it holds that composition and decomposition are universally possible, systemic monisim is a reductionist philosophy. But, in general, it is *holistic* rather than *atomistic* since, apart from a few trivial cases, the *whole* goal directed system is *more* than the sum of its goal directed parts.





ON MAKING A SCIENTIFIC REVOLUTION

Stuart A. Umpleby

In class I presented the view that the book we are assembling should be thought of at least in part as a handbook for making a scientific revolution, where the field to be revolutionized would be the social (behavioral) sciences. This view assumes that cybernetics has much to contribute to social science and that this promise has not yet been realized. Because of this disparity, I first interpreted the phrase "cybernetics of cybernetics" as the application of cybernetics principles to achieve acceptance of cybernetics theory by the relevant disciplines. That is, use cybernetics as the theory to guide the revolution.

However, most people in class seem to have defined the "cybernetics of cybernetics" on the individual (related discipline psychology) level rather than the social system (related discipline political science) level. Thus first order cybernetics is taken to refer to the purpose of the model; second order cybernetics refers to the purpose of the modeler. The scientific revolution point of view defines first order cybernetics as the development of basic theoretical principles; second order cybernetics as the use of those principles to win acceptance and consequently widespread application of them.

Although the cyberneticians have succeeded remarkably in developing a theory, they have yet to win its widespread acceptance. Not to assign responsibility for winning acceptance of the theory to cyberneticians would leave one without a plan of action for achieving acceptance.

The statement, "cybernetics has had little impact on social science," states an observable fact. However, the statement, "social scientists have absorbed very little of cybernetics," is predictable from what we know about selective perception, and thus should not be surprising. Consider again Thomas Kuhn's thesis that scientific progress does not occur as a smooth, autonomous process with superior ideas replacing inferior ones, but rather as the result of intense struggle very similar to a political revolution. Scientists who depart from the prevailing paradigm and who do not also study the rules and tactics of the scientific rough and tumble may not be purged or turned out of public office, but they may find it difficult to find research funds or receive academic promotions.

Using this perspective of scientific change as a process similar to politics, one could hardly expect an establishment (behavioralists) to welcome a set of ideas (cybernetics) which would result in their removal from positions of prominance. Thus when a radically new set of ideas comes along, it is in the interests of the scientific establishment (those with the greatest access to the leading journals and sources of funds, and who control academic promotions) to say that the new ideas are merely extentions of the ideas which they developed and consequently the most fundamental works are their works, not the new works. It is of course in the interests of the new group to accentuate the differences. To claim that the new point of view is fundamentally different is to require that graduate students read one's own works rather than the older establishment's works, thus increasing one's own standing at the expense of the existing scientific elite.

The basic point is that in science, as in politics, one should not rely on the opposition to implement one's program. Going one step further, to teach students a new scientific paradigm (as in a book called the Cybernetics of Cybernetics) without also teaching them how to cope with the problems of academic politics (promotion and tenure for example) in which adherence to the new point of view will inevitably involve them would be like throwing a lamb to a pack of wolves. The shortage of well established academic enclaves for cybernetics suggests to me that this may have happened in the past and should not be permitted to happen in the future.

Making a revolution, scientific or otherwise, involves two basic strategies--formulating the arguments to persuade and hold a growing group of followers and moving knowledgable, sympathetic people into positions of power. Some details of these two strategies are outlined below:

A. Develop a body of explanations, experiments, and successfully solved puzzles and then "publicize" these until they become well known.
   1. Whenever possible point out contradictions in the dominant paradigm and show how the new paradigm resolves the apparent contradictions.
   2. Ask questions (in letters to journals, at conferences) which the current paradigm cannot answer, but which the new paradigm deals with easily.
   3. Call attention to findings or common experiences which do not fit the current paradigm, but which do fit the new one.
   4. When explaining principles of the new paradigm, use examples familiar to the group one is addressing.
   5. Do not make the error of using the new paradigm to resolve current problems of the old paradigm, at least do not make this the core of one's argument. This strategy defines the new paradigm as merely a slight modification or improvement of the current paradigm. Make the core of the argument a challenge to the most fundamental assumptions of the current paradigm. A resolution of current issues of interest can then be used as a means to line up support among people interested in those issues and currently working on them.

B. Whenever possible practitioners of the new approach should be moved into positions of power and responsibility.
   1. Funding agencies.
   2. Leading publications. Setting up a new journal is less preferable to taking over an established journal. It is the difference between setting up one's own lines of communication and taking over the opposition's lines of communication for one's own purposes.
   3. Top ranked universities. Get the best graduate students possible, train them as thoroughly as possible and then get them appointments in the leading universities.
   4. Make sure the next generation is a close-knit group and stays in contact with each other. Develop a sense of a common professional destiny among those who know the new paradigm.
   5. When a new theoretical approach is very young, publish n in one or a very few journals with a large number of subscribers. As word of the new theory spreads, people will try to find the original papers. If they are hard to find, the revolution may be swallowed up in the flood of scientific papers. If this situation does develop, edited compilations of the scattered, early journal articles are essential.
   6. Establish new departments. One way to get control of existing university funds is to win department chairmanships for the developers of the new paradigm. An alternative (not mutually exclusive approach) is to set up a new department.

If the decision is made to include a section on winning academic acceptance for cybernetics it should present, as thoroughly as possible the basic arguments favoring cybernetics as opposed to behavioral science. It should also include a history of cybernetics (research groups, periodicals, major conferences, funding agencies, new departments, number of PhD's, etc.) up to the present time.

INTUITION

Intuition lay in the subliminal realm of information throughput in scientific inquiry. N. R. Hanson, in Patterns of Discovery, discusses intuition with regard to the first, or definitional-discovery step of the "scientific method": one solves problems algorithmically, but discovers them intuitively. Poincarè made the same observation: "It is by logic that we prove, but by intuition that we discover." The scientific method is logical and non-intuitive on its surface. Insight and intuition can be infused with this manifestly rational process, however, by application of heuristics. [D.S.]

ALGEBRA

A calculus in which only such general properties of an arithmetic are represented as are true irrespective of the values of the constants operated on. E.g. $a^2-b^2=(a-b)(a+b)$ irrespective of the numerical values of a, b. There are basically two kinds of algebras, those which completely represent the properties of their respective arithmetics, and those (subject to Godel's and Church's theorems of decision) which do not. E.g. in common arithmetic which is subject to Godel's theorem, there are accidental factorizations of numbers in excess of what can be determined algebraically. But the commonly used algebras of logic and set theory (non-numerical), and algebras to a modulus (numerical), neither of which are subject to Godel's theorem, do completely represent their arithmetics which thus have no accidental property. [G.S-B.]

ARITHMETIC

A calculus in which the constants operated on all have specific values, as in 2x3=6. The term includes the general properties (apart from the algebra) of a calculus, e.g. that there is no greatest prime number, and the art of discovering and proving the truth of such propositions (called theorems) is generally considered paramount in the science of mathematics. This art is distinct from that of computation, and no theorem can be proved by computer. (Euclid's geometric 'theorems' and Whitehead's and Russell's logic 'theorems' are strictly speaking not theorems but algebraic consequences, and thus demonstrable by computation alone). See algebra.

[G.S-B.]

Enrico Fibonacci

Come hear and sea what eye have found.

Eye am a person two.

Reality is eye and eye alone.
The world is a figment too.

If ever yew are in great pane.
Do just us on yore neighbor.
I don't wanna dew Anny work ether.

Say this group is chaos.
This groop is Kay uoss.
How come?
Why buy car or horse.
Hose the leeder hear any way?

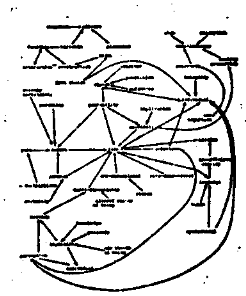
Eye don't under stand.
Thats thought full.
Yore crazy.
I's are four seeing.
Know Know you don't stand under.
Thats Greek fore noing.
Eye Kant by that.
The hole is greater than the some of it's parts.

KW

NUMBER

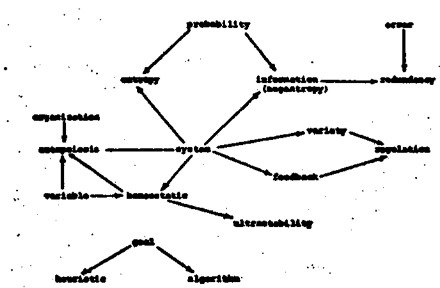
Number is the repetition of some cycle. In mathematical terms, if the circuit of unit radius is 2π, then $e^{2\pi i} = 1$ where i is $\sqrt{-1}$, and i itself is given by $e^{\pi i/2} = i$, which seems like bootstrapping until we transform this last equation to $i^i = e^{-\pi/2}$, which is approximately 0.208, thus defining i as that number which, when raised to itself as a power, yields very nearly a fifth. Number, being repetition, is thus inexorably connected with counting, and counting with tallying. Tallying is keeping pace with the recurrences of some effectively defined event by the making of a distinct mark to denote each such occurrence and re-occurrence. Thus, numbers could be tallied long before they were counted or named. Number names probably arose from the parctical need to refer in a convenient, understandable way to discrepancies between two tallies. The words two (Italian due, Sanskrit dva) and divide, and thou (Latin and Italian tu) are all from the same root. A divided mark is two marks, and to say "thou" or "you" implies the "me" of the speaker, i.e. two persons. Thus the entrance of "you" (tu) means two. Tu is two, as Gertrude Stein would say.

Just as there are positive (forward) and negative (backward) numbers (the ordinary ones) there are positively and negatively sideways and vertical numbers of various species called hypernumbers. See "Hypernumbers and Their Spaces" in Journal for the Study of Consciousness, vol. 5 (1972-3), p. 250. [C.M.]



ALGORITHM

A full comprehensive set of instructions for reaching a known goal. To illustrate: 'first left, second right, turn right at the Red Lion, my house is the third on the left'. [S.B.]



KoLaKowsKi

In Praise of Inconsistency

I speak of consistency in one sense only, limited to the correspondence between behavior and thought, to the inner harmony between general principles and their application. Therefore I consider a consistent man to be simply one who, possessing a certain number of general, absolute concepts, strives earnestly in all he does, and in all his opinions about what should be done, to remain in the fullest possible accord with those concepts. A consistent man is one who considers killing evil and refuses to enter military service; one who is convinced of the superiority of monogamy over other forms of family life and so does not deceive his wife. Consistent, too, is the policeman who believes regulations must be observed and therefore gives out summonses to jaywalkers and the like.

There exists in the history of culture a rare race of highly talented authors of the extreme Right, whose works provide invaluable material for reflection upon consistency. In France this is the breed of Bossuet, De Maistre, and Maurras. Men of considerable intellectual courage, they have not been afraid to carry their assumptions to their logical conclusions and loudly voice their judgments on every matter in which their principles were engaged.

Joseph de Maistre knows exactly what is the best order of the world, as ordained by God. He knows, too, what is most precious in this order and what is subordinate. Next, he demonstrates his amazing consistency by applying these general assumptions to all concrete questions: The world is so fashioned that evil must exist; given the existence of evil, there must be punishment. This being so, someone must inflict punishment and that "someone" is an indispensable element in the social order, worthy of respect for that reason. De Maistre then writes in praise of the hangman: "All greatness, all power, the hierarchy as a whole rest upon the hangman: he is the terror and the mainstay of human society. Remove this misconstrued factor from the world and instantly order will yield to chaos, thrones will shake, and society perish. God, who created authority, also created punishment." It follows that the hangman, because his profession is shrouded in dread, "is an anomalous being and to include him in the family of man requires a special dispensation, a *fiat* of creative power. He was created just as the world was."

Similarly: Spiritual crimes are worse than those of the flesh because the good of the spirit carries more weight. Spiritual crimes are also more odious because they offend against God's majesty, which is greater than that of terrestrial sovereigns; and this leads De Maistre to praise the Spanish Inquisition. In like fashion, Galileo was himself responsible for his trial because he could not refrain from writing despite his promise, because he defended the compatibility between Copernicus and the Bible, because he wrote in the vernacular and not in Latin. De Maistre concludes by lauding the tribunal that found Galileo guilty.

We may salute this fine example of consistency, of strict application of principle. On the other hand, we must note that humanity has survived only thanks to inconsistency.

What is required of a soldier going off to war? Uniquely that he be consistent in his righteous duty to defend his country. (I say "defend" because, as we know, all that seems to exist in wartime is "defense," and always "righteous," at that.) Battles fought by consistent soldiers can end only when the last man on one of the sides gives up the ghost. What is required of a citizen? Consistent loyalty to the state or government. Therefore a consistent citizen will always be proud to cooperate with the secret police, knowing it to be necessary to the existence of the state, to its glory and growth. To prove this is so is the easiest thing in the world, and every citizen who hesitates to write systematically to the secret police informing on his neighbors is surely inconsistent. Let us assume that we consider a certain matter to be the most important in the world; for example, a universal obligation to wear a top hat. Why, then, should we object to imposing our idea by means of war, aggression, provocation, blackmail, assassination, intimidation, terror, murder, or torture?

The race of those who vacillate and are soft, the inconsistent people, precisely those who happily eat steak for dinner but are totally incapable of slaughtering a chicken; those who do not wish to contravene the laws of the land yet do not denounce others to the secret police; those who go to war but in a hopeless situation surrender as prisoners rather than die in a last-ditch fight; those who prize frankness but cannot bring themselves to tell a famous painter that his work is terrible, nervously uttering words of praise which they do not mean—in short, the race of inconsistent people—continues to be one of the greatest sources of hope that possibly the human species will somehow manage to survive. For this is the race of which part believes in God and the superiority of eternal salvation over temporal well-being, yet does not demand that heretics be converted at the stake; while the other part, not believing in God, espouses revolutionary changes in social conditions yet rejects methods purporting to bring about these changes which openly contradict a certain moral tradition in which these people were raised.

In other words, total consistency is tantamount in practice to fanaticism, while inconsistency is the source of tolerance. Why should anyone inflexibly convinced of the exclusive truth of his concepts regarding any and all questions be willing to tolerate opposing ideas? What good can he expect of a situation in which everyone is free to express opinions that to his mind are patently false and therefore harmful to society? By what right should he abstain from using any means whatsoever to attain the goal he regards as correct?

We could say at this point that tolerance is extorted, that the only things tolerated are those which, for lack of ammunition, cannot be destroyed. And as a rule the only people are those who are so strong that their opponents cannot eliminate them with impunity. This observation is certainly well documented by history, but it does not explain everything. If a power relationship were the sole basis of tolerance, and if in addition fanatical consistency ruled the minds of the antagonists, the two groups would be permanently involved in trying to eradicate each other. Since this does not occur, or at least not always, it is only as the blessed result of inconsistency, an inconsistency which does not necessarily spring from conscious acceptance of the principle of tolerance, but merely manifests itself as if that principle were accepted to some degree.

Inconsistency is simply a secret awareness of the contradictions of this world. By contradictions I mean the various values that are, notoriously throughout history, introduced into society by mutually antagonistic forces. If convictions of the absolute and exclusive superiority of a given value to which all else is subordinate were to spread and be practiced widely, they would of necessity transform the world into an ever-larger battleground—which indeed does occur from time to time. The lack of consistency checks this tendency.

Inconsistency as an individual attitude is merely a consciously sustained reserve of uncertainty, a permanent feeling of possible personal error, or if not that, then of the possibility that one's antagonist is right. We have been speaking all this time about the relationship between thinking and the bases of practical action. Now, all thought that can in any way manifest itself as a causative factor in practical conduct is the affirmation of a value. In turn—and this is one of the most important principles we wish to formulate—*the world of values is not logically dualistic*, as opposed to the world of theoretical thought. In other words, there are values that exclude each other without ceasing to be values (although there are no mutually exclusive truths that still remain truths). Daily life shows at every step what a truism this statement is.

Inconsistency, in the sense we use it here, is simply a *refusal once and for all to choose beforehand between* any *values whatever which mutually exclude each other*. A clear awareness of the eternal and incurable antinomy in the world of values is nothing but conscious inconsistency, though inconsistency is more often practiced than proclaimed. Inconsistency is a constant effort to cheat life, which incessantly tries to place us before alternative doors, each of which is an entrance but through neither of which we can return. Once we have entered we are compelled to fight to the end, to the last bullet, for life or death, with him who entered through the other door. Thus we try to dodge, to maneuver, to use all the tricks and traps, all the suspect manipulations and stratagems, subterfuges and evasions, the chicanery, half-truths, hints, and circumspection—anything to keep from being pushed through either of the doors that opens upon a single direction.

These attempts to deceive life, to conciliate implacable antagonists—these efforts to evade the fatal "either-or" between contradictory values—all this is not the result of a temporary derangement in people's lives that will be removed with the advent of the new era. It is the result of human nature, whose antinomies are always with us. Accepting them as part of man's universal lot, we can elude these antinomies through inconsistency, in order not to reject permanently something we value just because something else we esteem is eternally contradictory to it. So we try to postpone final decisions until the end of life overtakes us, the sole situation in which there is no longer any possibility of choice.

At this point someone may ask: Is this any different from the common-sense wisdom Aristotle set forth in his Nicomachean Ethic? Indeed, his idea is based on the premise that there exist virtues and antivirtues, as well as corresponding vices and antivices. Thrift is a virtue and parsimony a vice; but generosity is also a virtue, while extravagance is wrong. We can—so taught the father of Europe's intellectual tradition—reconcile contradictory virtues without falling prey to the opposite vices. Let us be at once frugal and open-handed, but neither miserly nor prodigal. Let us hold that middle ground between recklessness and cowardice which harmoniously combines valor and prudence. As between a feverish lust for fame and timorous humility, let us seek to maintain a position which unites healthy ambition with modesty. Equally removed from brutal vengefulness and abject submissiveness, let us be both firm and gentle. For the truly generous man is not wasteful but thrifty; just as the courageous man is not reckless but cautious. And so forth.

Have we, in praising inconsistency, come to the point of merely repeating the age-old dictum of the golden mean? Let me confess at once that my idea does not pretend to the slightest consanguinity with this middle-of-the-road theory; it is, in fact, the exact opposite. Aristotle's ethic! Aristotle's ethic was clearly earthly, but his earth was flat. He expounded a novel concept that unified the Hellenistic world. He conquered the world because he was the embodiment of the spirit of universal conciliation at a time when unity was needed most. In metaphysics, politics, and moral doctrine he personified this unity. But Aristotle's genius is alien to us because we live in a world of extremes.

If we look more closely at this Aristotelian ethic we see that its main current is a longing for synthesis and a belief that between any two extremes one can find a mean that will preserve the best of each and reject what is harmful. It is assumed that a reasonable mind can harmonize what to the immoderate one appear to be contradictions. In other words, Aristotle believes that the contradictions in human attitudes that erupt into social antagonisms are not inherent in the world but are caused by lack of reasonableness. Thus antinomies are created by man, who in one way or another misuses the good in the world.

My praise of inconsistency, however, springs from a completely different source. It posits that contradictions in values do not stem from their abuse and therefore are not merely appearances that can be overcome by intelligent moderation. These contradictions inhere in the world of values and cannot be reconciled in any synthesis. Reasonable inconsistency does not seek to forge a synthesis between extremes, knowing it does not exist, since values as such exclude each other integrally. The real world of values is inconsistent; that is to say, it is made up of antagonistic elements. To grant them full recognition simultaneously is impossible, yet each demands total acceptance. This is not a matter of logical contradictions, because values are not theoretical theses. It is a contradiction which lies at the heart of human behavior.

Inconsistency is thus a certain attitude which, having realized that this is the situation, knows the extremes to be irreconcilable yet refuses to reject either because it recognizes each as valid.

Naturally, I do not mean that no concrete conflict of two values clashing in a given situation can ever be reconciled. My thesis is not concerned with any pair of contradictions within a defined context, but with the condition of contradiction as such. In other words, I believe it is possible to synthesize or surmount actual contradictions, but at the same time I am convinced—in accordance with the experience of history as a whole—that one contradiction disappears only to give way to another; that therefore, no universal synthesis is possible. In the world we live in, contradictions cannot be reconciled; once resolved, they pertain no longer to this world but to a dead one, regarding which we need no longer take a practical stand. Contradictions pursue us as long as we act in a world of values, which simply means for as long as we live.

Let us take an example from everyday political life. We believe that nations have the right to decide their own fate. This belief affirms a certain value. We also believe that certain important social institutions preserved in the life of our nation are detrimental to its development. Yet we see unmistakably that our people prize these very institutions highly and obviously have no intention of listening to our arguments on the subject. For instance, we are not only immune to the benefits of religious consolation but deeply convinced, besides, that the continuing influence of religious institutions upon public life is damaging. Nonetheless, this influence not only manifests itself, but undoubtedly does so in accordance with the will of the people. How should we behave? We do not want to renounce either of the opposing principles involved. We do not want to destroy by force and against the people's will an institution manifestly supported by the majority. Nor do we wish to abandon the fight to abolish this institution.

Some might call this a trite situation, one that in no way excludes a synthesis. That synthesis will come about in the course of history. We expect that as people grow sufficiently enlightened the present conflict will cease to exist. Meanwhile, though, we must—with total consistency—strive to educate people so that in the future they will be disposed to accept our concepts and choose of their own free will to extirpate institutions that, in our opinion, hamper the nation's growth.

Unfortunately, this sage advice does not dispel our doubts. Certainly nothing prevents us from believing that in the course of history, measured by the yardstick of generations, the will of the people will undergo such a change, and that their level of maturity will rise to the point where this particular conflict will disappear. Nevertheless, this hope is of fleeting value in practical daily behavior. Let us suppose that in this country I have to vote on whether the teaching of religion should be introduced in public schools, when heretofore there was none; or, on the contrary, whether to discontinue existing religious education. Then I have no other choice: I vote either in agreement with the will of the people, or else in accordance with what I consider their good. I must vote—this obligation has been forced upon me by circumstances beyond my control. I cannot remain consistent toward both my principles at once, but I do not want to forsake either. This example is neither contrived nor exceptional. How many deputies in any number of parliaments vote one way while secretly hoping the opposition will win?

The problem of the antinomy inherent in the principle of tolerance is eternal and eternally unresolved: how to preach and practice tolerance toward ideas and movements which are intolerant. We act against our basic tenet if we silence these ideas and movements by force; we also act against our principle if we tolerate them, for we thus enable them to triumph and destroy the principle of tolerance in social practice. And it is cold comfort under the circumstances to hope that this contradiction will be solved in the process of historical development, either because, having slaughtered all the enemies of tolerance, we shall be able to apply it boundlessly; or else because these movements will in the course of time discard their intolerance. In practical everyday actions and in our daily participation in society, such perspectives help us minimally in making decisions.

These examples are not fictitious. Our lives are bound up in conflicting loyalties that we must choose between in concrete situations. We must break one bond in favor of another, while still not questioning the first. Loyalty to the individual, to one's own outlook on the world, to human communities in which we find ourselves either accidentally or of free choice, loyalty to nations, parties, governments, friends, to ourselves and those close to us, to our own nature and our convictions, to the present and the future, to concrete things and universalities—there are as many insurmountable contradictions as there are loyalties. An authentic synthesis resolving chronic conflicts rarely occurs; most often the supposed synthesis is superficial and fraudulent. We deceive ourselves with it in order to appear consistent, for one of the values instilled in us since childhood is consistency. Our proposition, aimed at making us realize that in these conditions consistency is an ideological fiction, is thus also intended to remove at least one kind of conflict: that which results from a belief in consistency as a value. So, proclaiming the contradictory nature of the world, we strive to attentuate it at least at one point, for, as we see, conflicts multiply because they are misunderstood. In other words, praise of inconsistency is at the same time the rejection of a specific value, that of the consistent life. The contradiction between the value of a consistent life and that of a basically reasonable one belongs to the species of conflict which may perhaps be removed unilaterally—not by synthesis, but by the repudiation of one of the sides to the dispute.

And immediately the question arises: Can we really proclaim the principle of inconsistency in a perfect formulation, which means, in essence, consistently? Is there no sphere of human events for which we can postulate total consistency, thus falling in turn into conflict with the above-mentioned repudiation? We must answer this question affirmatively. Such a sphere does exist. We call such events elementary situations.

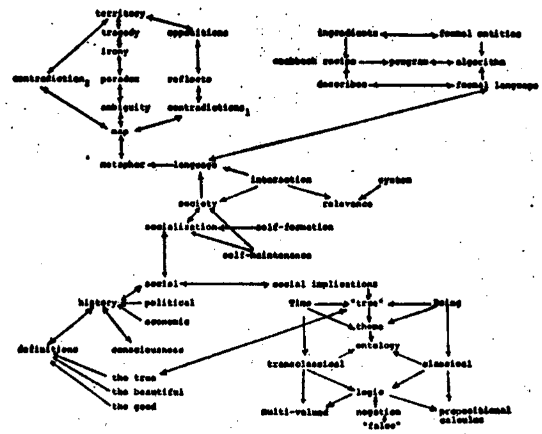
Elementary situations are those in which tactics perish; that is, those human situations in which our moral attitude is unchanged regardless of the way these situations arrive at their culmination. If a man is dying of hunger and I can feed him, then there is no confluence of circumstances in which it would be right to say "It is, nevertheless, tactically better to let him die"; or, if I cannot help him, to say "Tactically it is better to hush up the fact that he died of hunger." Open aggression, genocide, torture, mistreatment of the defenseless—all these are elementary situations. In such situations the values of inconsistency cease to play a role, and here we suddenly confront a dual-valued world. In this way our praise of inconsistency is also inconsistent. Inconsistency has certain limits within which it is valid: the limits wherein reality is contradictory. But reality is contradictory only up to a certain point. (We are speaking at all times about the reality of values and not about the reality which is the subject of theoretical speculation.) For let us also carefully bear in mind that to be consistent in inconsistency means to contradict by an act (the application of a certain consistency) something the affirmation of which (the affirmation of inconsistency) is the substance of that act; in short, to fall into an impossible situation, into an antinomy.

Let us therefore also be inconsistent in our inconsistency, and apply the principle of inconsistency to itself. But, someone may reproach us, only then do we practice strict inconsistency, only then do we attain total consistency in the practice of inconsistency—for if we were always inconsistent, but our very inconsistency were completely consistent, then by that very fact we would not always be inconsistent. When, however, we limit our inconsistency, that is, when we are not always inconsistent, only then do we become absolutely inconsistent. In other words, we have arrived at the most classical antinomy of terms: consistent inconsistency is not consistent inconsistency (for it excludes inconsistency itself from the principle of inconsistency); inconsistent inconsistency, on the other hand, is actually consistent inconsistency. To this extent, therefore, we propose to preserve the principle of consistency as a value, by practicing the principle of inconsistency inconsistently. To this extent we mold our praise of inconsistency to a perfect form, protesting against the practice of inconsistency in its perfect form.

So much for praise of inconsistency. The rest cannot be verbalized. The rest must be done.

PROGRAM

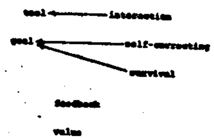
A cookbook recipe is a <u>program</u> (or "algorithm"). A good recipe, namely one that is complete in that it has full specification of the ingredients; that describes the appearance of the dish in each stage of its preparation; that gives the sequence of steps in the preparation of the dish in an order that is logical, convenient, and efficient; and that produces in the end infallibly a culinary delight, is the ideal and archetype for all programs. ---Computer programs differ from cookbook recipes only in that the ingredients are formal entities and the recipe itself is given in a formal language. In all other respects, a good computer program is the same as a good cookbook recipe -- since for the computer gourmet, an elegant result is a culinary delight.

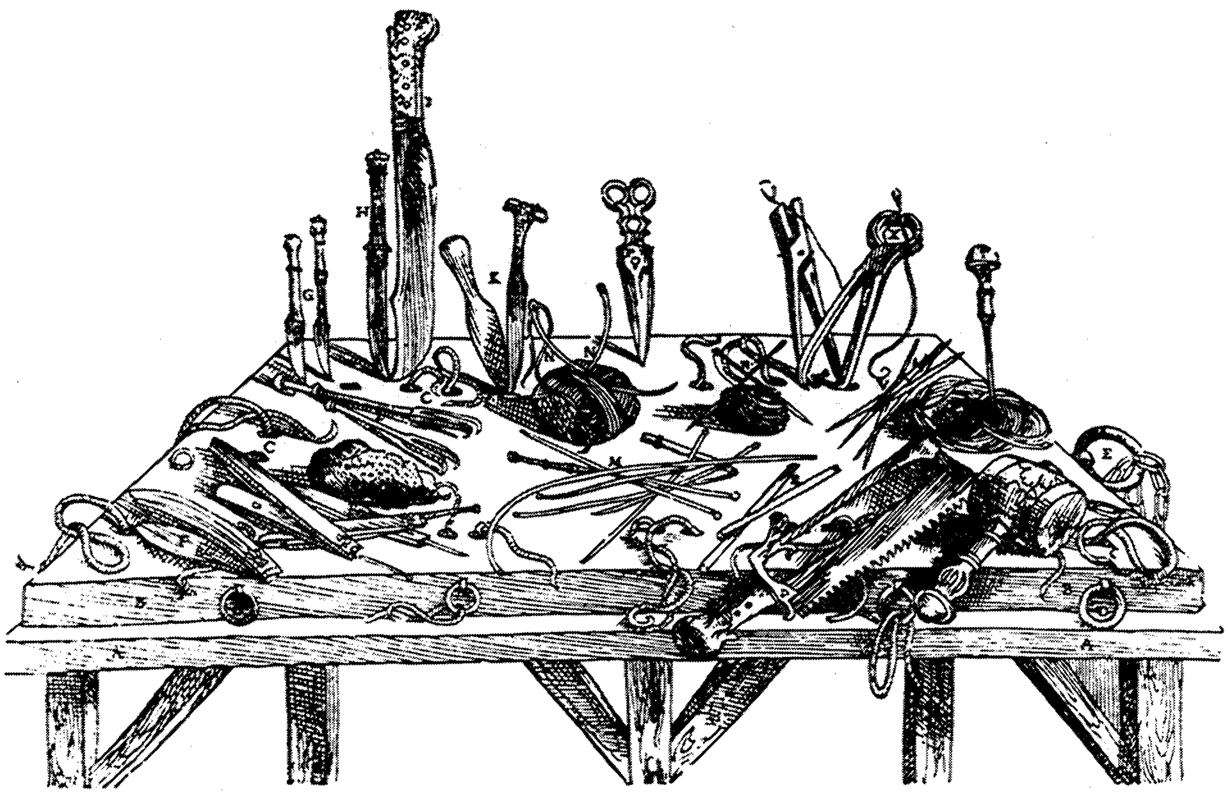


[R.H.H.]

TOOL

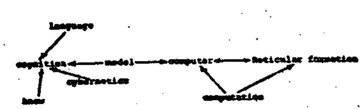
A tool consists of a use at one end and a grasp at the other. Tools, tasks, and users co-adapt and co-evolve in rich interaction. [S.Br.]





COMPUTATION

Computation refers to the execution of any procedure (algorithm) for transforming one set of data (called "inputs") into another (called "outputs"). A computing machine implements this transformation by means of a program which represents the algorithm in a well-defined machine language, operating on any of a variety of storage devices (e.g., registers, memory, disks, tapes, card readers, printers) containing the input, output and any intermediate data in the form of device-dependent codes. The brain, on the other hand, implements this transformation by neurons. Warren McCulloch identified "domains of computation" in the brain which differ from each other in the algorithms they implement, in the types of their neurons and in the interconnections of their neurons. (See Computer) [L.S., W.D.]



Reflections on Notation and Cognition

Derek Schultz

In the preface to Wittgenstein's Tractatus Logico-Philosophicus, Bertrand Russell, in discussing notational systems and logic, remarked, "...a perfect notation would be a substitute for thought."[1] This passing comment raises important questions to cognitive psychology: What would constitute a "perfect notation" such that it might be a substitute for thought? Further, by what mechanism would notation supplant thought, and what type of thought might it supplant? As basic as these questions seem, Psychological Abstracts references no articles under "notation", and only thirty-seven under related topics: mathematics, symbols, signs, logic. To understand notation may be to understand some aspect of cognition. Surely, such a venture is worthwhile.

Naturally, the first question is of definition: what is a notation? Generally, a notation seems to be a specialized symbolic system: specialized in that it is employed for communicative or problem solving purposes in only a narrow range of situations; symbolic in that the graphic elements tend to represent things other than themselves; and finally, a system, in that the graphic symbols are interrelated and dependent upon one another. To better understand notational systems in genre, the diversity of such systems may be examined. Table I lists a variety of notational systems with brief descriptions of those that may not be familiar.

The great diversity and pervasiveness of notational systems exemplified in Table I belies their psychological significance. To approach understanding of the cognitive factors and mechanisms relevant to notational systems, classification of these diverse notations may be useful. One such classification is presented in Table II. Two classificatory criteria are employed: The first is the implicit primary goal to which the notational system is applied: either communication of information, i.e., communicative; or as an aid in problem-solving, i.e., heuristic; or both. Certainly, to be heuristic a notation must be communicative, but the criterion is of primary goal. A second criterion is similarly employed: What is the process or paradigm by which the notation functions (in a strictly general sense)? The notation may be iconic--a diagrammatic representation of the real world, within a single frame of time. That is, an iconic notation diagrams a physical situation within a temporal snapshot. Other notations are more symbolic, employing elements representative of objects or ideas in a non-iconic way, and atemporally. Finally, a

Table I: Notational Diversity

1. Arithmetic
2. Babbage Notation - a movement notation for mechanisms
3. Benesh Movement Notation - a dance notation
4. Blueprints
5. Braille
6. Calligraphs - oriental graphemes
7. Chemical Equations - e.g., $2H_2 + O_2 \rightarrow 2H_2O$.
8. Chemical Structures - molecular diagrams, e.g. organic chemistry
9. Chess Notation
10. Circuit Diagrams
11. Eshkol Notation - a dance notation
12. Flow Diagrams - especially in computer programming
13. Football Diagrams - of plays
14. Geological Maps
15. Geometry Diagrams
16. Gestures
17. Graphs - of any type
18. Hieroglyphs
19. Insurance Maps
20. Kenographs - a three-valued logic
21. Labanotation - a widely used dance notation
22. Language - words or symbols
23. Laws of Form - a symbolic logic with one operator
24. Lewinian Topological Psychology - diagrams psychological spaces
25. Logographs - trademarks
26. Mathematics
27. Meteorological Maps
28. Morse code
29. Musical Notation
30. Notation for Iconics - a specialized notation for visual patterns
31. Notation for visual patterns
32. Physics Equations e.g. $F - Ma$
33. Pictographs - early writing systems
34. Predicate Calculus - advanced symbolic logic
35. Propositional Calculus - e.g. $P \rightarrow Q$
36. Road Maps
37. Semaphore Signs
38. Shorthand
39. Signs - of psychological significance, e.g., stereotypes
40. Source Code - computer programming language
41. Topographic Maps
42. Venn Diagrams

notation may be synthetic--transfering information from one mode to another along a time continuum. For example, musical notation translates information from audition to vision along a sequence of time frames--bars or measures. It should be understood that these rough classificatory criteria are conceived as mutually exclusive, but are not pragmatically so. For instance, Braille can be classified as either symbolic or synthetic notation with good argument. Other such ambiguities exist within this classification schema, and this classification is just one among those possible for this notational diversity. This particular schema was presented for psychological relevance: it categories on both process and goal, and is thus highly cybernetic. Another psychologically relevant classification might be generated by scaling each of the 42 notations on twenty or twenty-five criteria, and factor-analyzing the results. Such an analysis may imply psychologically relevant factors in notations. A similar approach might be to scale each notation on a highly standardized factor-analytic system: the semantic differential. [2]

At a different level, other criteria may be relevant to categorizing notations. Within the group of notations primarily oriented toward problem-solving, earlier labelled "heuristic", a distinction may be made between those whereby the solution is ground out in a stepwise or algorithmic fashion--Mechanical notations--and those whereby the solution may be seen or be made perceptually obvious--Perspicuous notations. The former is exemplified by mathematics and propositional calculus, the latter by graphs and Venn diagrams. Perspicuous notations, in this schema, are more truely heuristic: an algorithm is not available by these notations.

Whatever classification scheme is used, however, the criteria will imply notational variables--factors upon which notations differ. An alternative approach to examining notations is via their shared universals, or notational constants. In a recently published book, Huggins and Entwistle [3] presented a list of invariance factors which they deemed relevant ot iconic communication--communication by pictures and visual images. The factors they present seem likewise highly relevant to the similar situation of notational systems. Their factors can be presented as hypothesized notational constants:

1. Layout of images on screen (i.e., in space)
2. Identification and portrayal of relations
3. Avoidance of irrelevant information
4. Use or misuse of color
5. Use of motion or spatial perspective

Notational Classification III:

Process/Paradigm	Goal (Primary): Communicative	Heuristic	Either Goal
Iconic (1 Time Frame)	Road Maps Topographic Maps Geological Maps Meteorological Maps Insurance Maps Blueprints Circuit Diagrams Chemical Structures		Geometry Diagrams Graphs Topological Psych. Diagrams (Lewin) Venn Diagrams
Symbolic (Atemporal)	Pictographs Hieroglyphs Logographs Signs Language Calligraphs Braille Shorthand Source Code Gestures	Arithmetic Mathematics Propositional Calculus Predicate Calculus Laws of Form Kenograms	Physical Equations Chemical Equations Notation for Vis. Patterns Symbols
Synthetic (Time Continua)	Labanotation Eshkol Notation Benesh Mvmnt Notation Musical Notation Notation for Iconics Babbage Notation Chess Notation		Football Diagrams Flow Diagrams

6. Invention or design of glyphs

This is not to say, for example, that all notational systems layout images spatially in the same way. Rather, the layout of notational symbols or images is presumed constant in accordance with some psychological variable. Thus, the Huggins and Entwistle factors imply analysis of notations at a higher level than mere classification: these factors address themselves to psychological constructs or variables. A rough summarization can be made of such cognitive and psychological factors:

1. Perceptual Factors:
   - A. Discrimination Variables
   - B. Attention Variables
   - C. Variables hinging on both A and B.
2. Cognitive Factors:
   - A. Symbol-independence variables
   - B. Symbol-relational variables

Such factors may be more easily grasped in a different framework. Consider a notational system, symbolic logic perhaps. Certain perceptual factors seem highly relevant: the symbols must be highly discriminable, i.e., easy to separate visually. Moreover, and related to discriminability, the symbols must be attentionally separable: the attention on one symbol or group of symbols should not be interfered with by any other symbols. This latter objective may be achieved by symbol classification: one set of symbols with particular relevance may be in Greek letters, another set in capital Arabic letters, yet another set in small Arabic letters. The letter set itself then carries information. The same objective may be achieved positionally, not just graphemically: a superscript is a power sign, an exponent; a subscript is a set sign, and so forth. Such compositional rules have long been employed in mathematics generally, and specifically in statistics. The potency of such compositional factors was precisely the point Menger addressed in "Gulliver in the Land Without One, Two, Three." [4] Such compositional schema are reflected perceptually or spatially, but address cognitive factors. Thus, discrimination and attention factors in notation may be considered primarily perceptual, but compositional rules manifested spatially might more appropriately be perceptual-cognitives.

A higher order of cognitive factors may exist, however. Weston, Von Foerster [5], Pask [6], and Maruyama [7] have all discussed "cognitive paradigms"--ways in which people think differently about certain types of problems, and the paradigms of thinking employed tend to categorize thinkers. In "Artificial Intelligence", Weston and Von Foerster reported that what is perceived as a problem by seniors in biology may be perceived as a solution by seniors in electrical engineering. Since many notations are oriented toward problem-solving, questions of "What is a problem?" and "What is a solution?" are certainly relevant here. It seems that cognitive paradigms define "problem" and "solution" by perceived context. It is not unreasonable, then, to presume that notational systems chosen as heuristics by groups whose cognitive paradigms differ will themselves differ. Inferences to cognition may be possible upon examination of such notational differences.

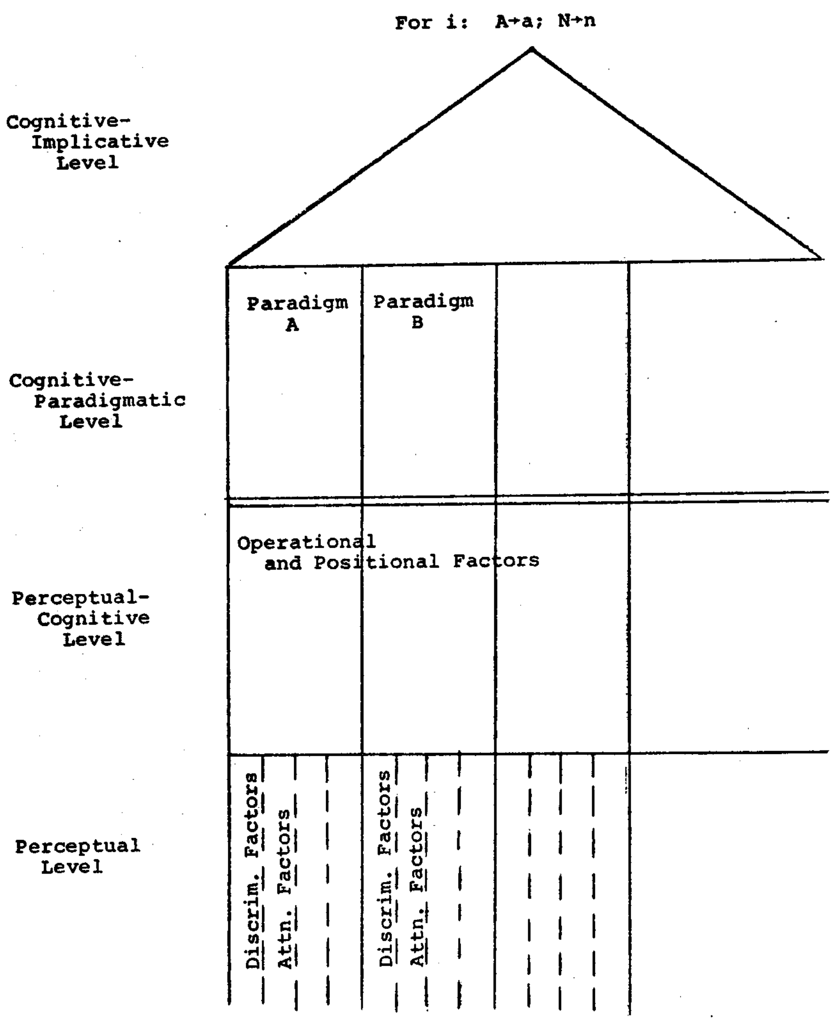
In a similar framework, Gordon Pask has noted the difference between "lumpers" and "stringers": the former seek information in large amorphous, heterogeneous chunks, the latter in a systematic stepwise piecemeal fashion. Maruyama extends a similar idea both in schema and in scope. The actual number of cognitive paradigms is moot, but their apparent existence is significant. We think; we think about notations; we think with notations. We think in systematically different ways, and notations seem to differ systematically. It may be possible, then, to infer cognitive paradigms from a systematic examination of notations. Thus, an upper cognitive level, that of cognitive paradigm, seems essential to examining notations.

But higher still, despite the individual differences that are concommitant to nearly all psychological traits, a universal factor may be presumed that will unite the diversity of notations and cognitive paradigms. Such a universal may be stated in the form of an implication: "If person i employs cognitive paradigm A, he will employ notational paradigm a. If person i employs cognitive paradigm B, his notational vehicle will be of type b." and so on. This highest cognitive level, that of cognitive to notational paradigmatic implication, could be considered the notational constant. The context of this level is largest; it may lay at the contextual limit in notation.

The relations just discussed are presented diagrammatically in Figure 1.

A schema for notation in a psychological framework has been hypothesized. It is vague because it is general. It is now essential to step beyond the potential theoretical framework by which notation is seen to be relevant to levels of psychological theory, and to consider possible variables by which notation may facilitate or even supplant thinking, with an eye toward research. That is, if notation somehow "thinks for us", the question is "How?" This question has

Fig. 1



rarely, if ever, been raised in orthodox experimental cognitive psychology, or if it has been, only in the most specific contexts (e.g., arithmetic). Thus, at this stage, intuition and introspection seem to offer up psychological variables potentially relevant to notations, and, from these possibilities, tests must be devised.

One such intuitively generated variable is imagery, especially visual imagery. Specifically, it may be the case that a notational system may supplant visual imagery that may be concommitant with problem-solving. Since the turn of the century, discussion of "imageless thought" and imagery in problem-solving behavior has abounded in psychological literature. Reviews of and allusions to this question are found in Arnheim [8], Newell and Simon [9], Paivio [10], Segal [11], Sheehan [12], and Pollio [13]. If visual imagery facilitates problem-solving in certain situations and if particular notations may supplant visual imagery, then it may be the case that those notations "think for us" by "imaging for us."

A related potential variable is memory, once again especially in the visual modality. Paivio [14] has presented a "visual peg" hypothesis: visual imagery aids memory by providing a perceptual "peg" upon which a chunk of information may be hung and subsequently easily retrieved. This process seems to be that which Luria's famous mnemonist [15] employed. In any event, if memory itself is the relevant variable in problem-solving, and imagery merely a means of memory enhancement, then notational systems might be examined via memory tasks.

A third approach to the psychological mechanisms by which notational systems seem to facilitate thought is by information-transmission analysis. A good notation may package more information than does a natural language presentation into a chunk of perception. Thus, the memory-load may be eased, or cognitive examination of the information presented may be facilitated. Exemplary of this possibility are graphs: a perfect correlation is not easily perceived from data alone, but is quickly grasped when those data are graphed.

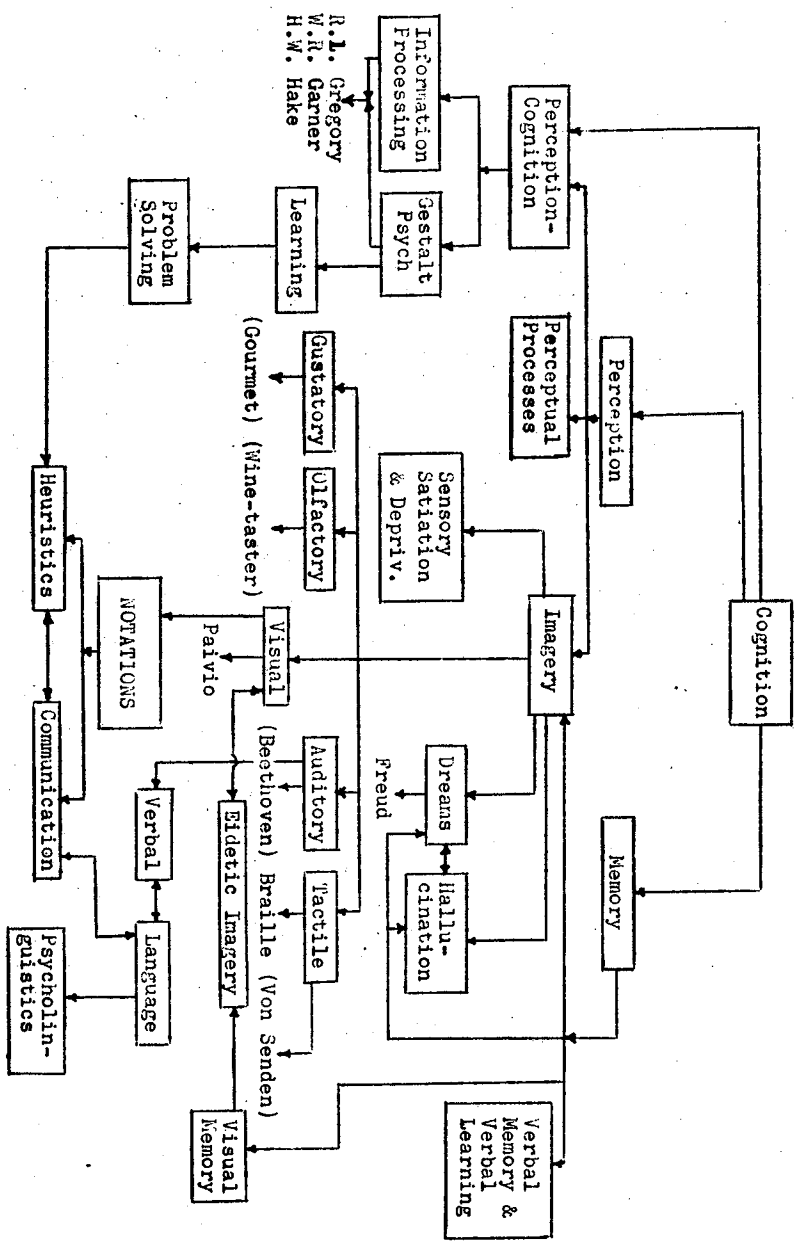
Certainly, notational systems may function in very different ways on certain psychological variables: a Venn diagram may aid imagery, but propositional calculus may not. The users of notations may vary on cognitive paradigms; the notations themselves seem to differ on relevant psychological variables; and the relevant psychological variables are undoubtedly more numerous and complex than this cursory presentation outlines. Figure 2 diagrams, in a schematic fashion only, areas of conventional cognitive psychology, possibly relevant to the examination of notational systems, and how these areas may interrelate.

The preceding analysis of notation in a cognitive perspective is designed to raise questions, not to answer them. Notational systems seem to offer a great potential for insight into human cognition...a potential that seems up to now overlooked. Fortunately or not, only psychological research will reveal the fruitfulness of the orientation herein presented.

Footnotes

1. Russell, Bertrand, Preface to Wittgenstein's Tractacus Logico-Philosophicus, Routledge and Kegan Paul, Ltd; London, 1922, p.xviii.
2. Osgood, C.E., Personal Communication, 1974.
3. Huggins, W.H. and Entwistle, D.R., 1974, Iconic Communications: An annotated bilbiography. Baltimore: Johns Hopkins, 171 pp.
4. Menger, K., 1959. "Gulliver in the Land Without One, Two, Three." The Mathematical Gazette, 43, No. 346, pp. 241-250.
5. Weston, P.E. and Von Foerster, H., 1974, Artificial Intelligence in Annual Review of Physical Chemistry, Vol. 24, Palo Alto: Annual Reviews, Inc.
6. Pask, G., Anti-Hodmanship.
7. Maruyama, M., 1973, Paradigmatology and its Application to Cross-Professional and Cross-Cultural Communication. Presented at the EDRA Conference, April, 1973.
8. Arnheim, R., 1969, Visual Thinking, Berkeley; U. of Calif. Press.
9. Newell, A., and Simon, H., 1972, Human Problem-Solving, Englewood Cliffs: Prentice-Hall.
10. Paivio, A., 1971. Imagery and Verbal Processes. Chicago: Holt, Rinehart, and Winston.
11. Segal, S.V. (ed.), 1971. Imagery: Current Cognitive Approaches. New York Academic Press.
12. Sheehan, P.W., (ed.), 1972. The Function and Nature of Imagery. New York Academic Press.
13. Pollio, H.R., 1974. The Psychology of Symbolic Activity. Reading, Mass: Addison-Wesley.
14. Paivio, A. op. cit.
15. Luria, A.R., 1965. The Mind of the Mnemonist. New York: Basic Books.

Figure 2



Derek Schultz

Cognitive Systems: A Personal View

The entailment structure on the following pages ia a schematic/diagramatic of my personal intellectual interests, and the interrelations I perceive among them. At the outset, it should be understood that all factors in the entailment structure are intellectual interests: this cognitive system does not extend beyond the intellectual domain.

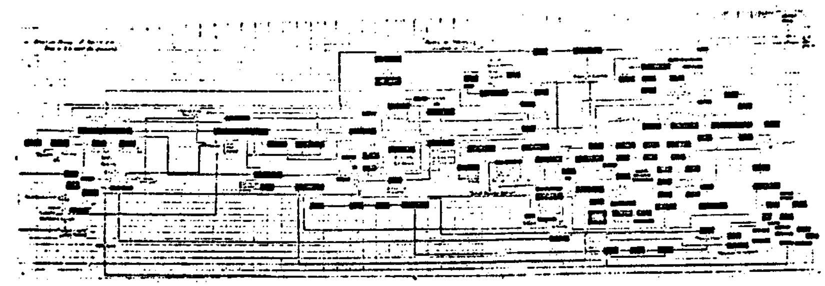
In this structure, factors encased in boxes are disciplines or topical areas, whereas those in quotes are books, papers, or general concepts. Any person associated in my mind with these factors is indicated without either boxes or quotes. The arrows roughly correspond to "leads me to" or "reminds me of",...that is FACTOR 1 "reminds me of" or "intellectually connects with" FACTOR 2.

Precisely, there are no dimensions in this structure upon which the factors are scaled: that is, the boxes are placed for convenience and aesthetic congeniality. Though spacing or distance on this structure has no particular significance, it is generally the case that the most important factors are centered between top and bottom, and a logical progression exists from left, ("Theistic Positions") to right ("Cybernetics; Systems Analysis"), on this horizontal axis.

It would, of course, be desirable to model this structure in three dimensions (say spherically) rather than merely in two... for one thing, "Cybernetics" would then be spatially proximal to "Theistic Positions" - an interesting relation!

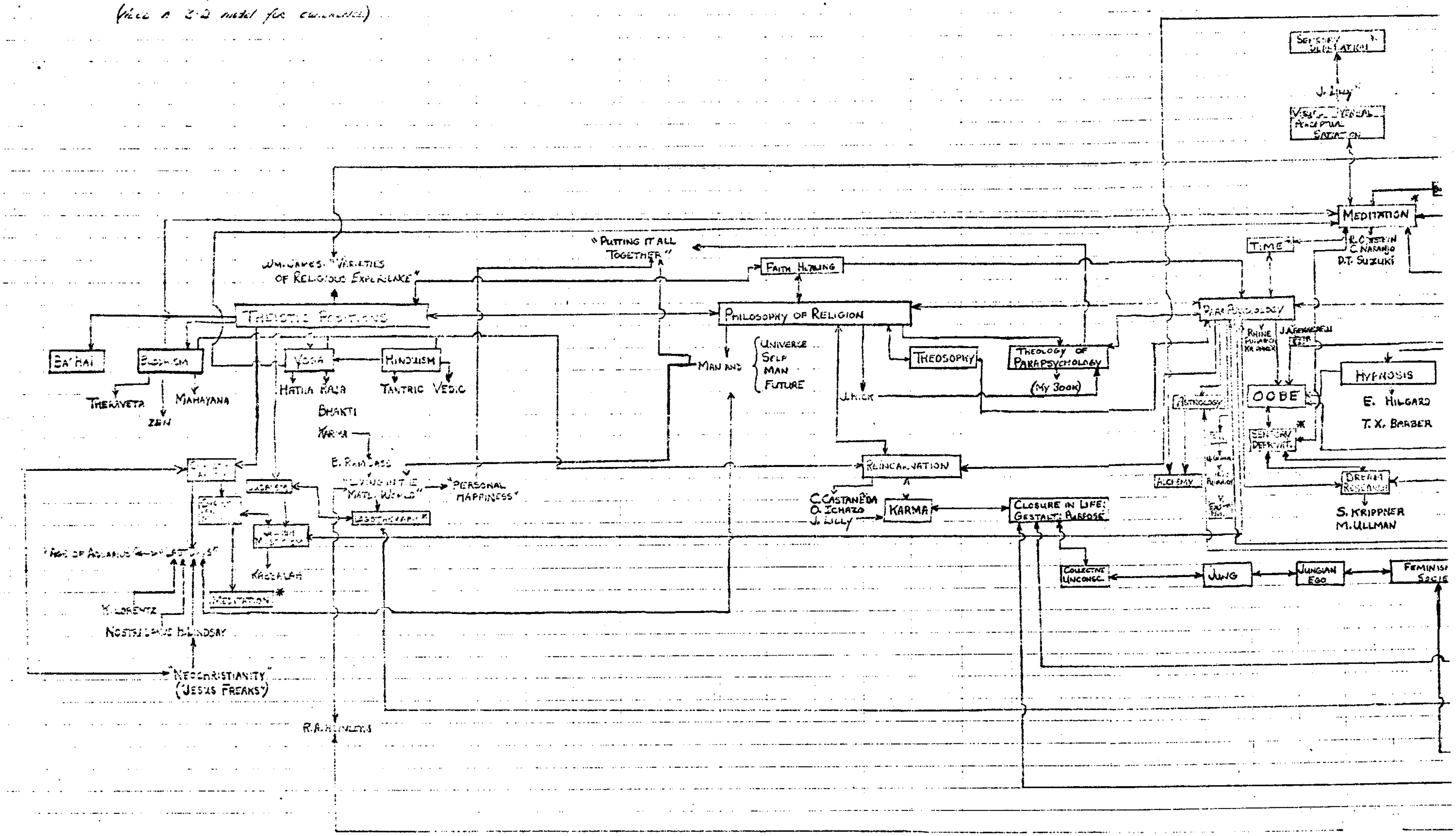
Before actually explaining details of presentation, I would make two caveats: 1) I apologize for the handwritten presentation, and, at places, visual confusion - this structure began as a personal cognitive topology; 2) I would appreciate feedback regarding this presentation, and what is entailed therein. Until about 1977, I'll be at:

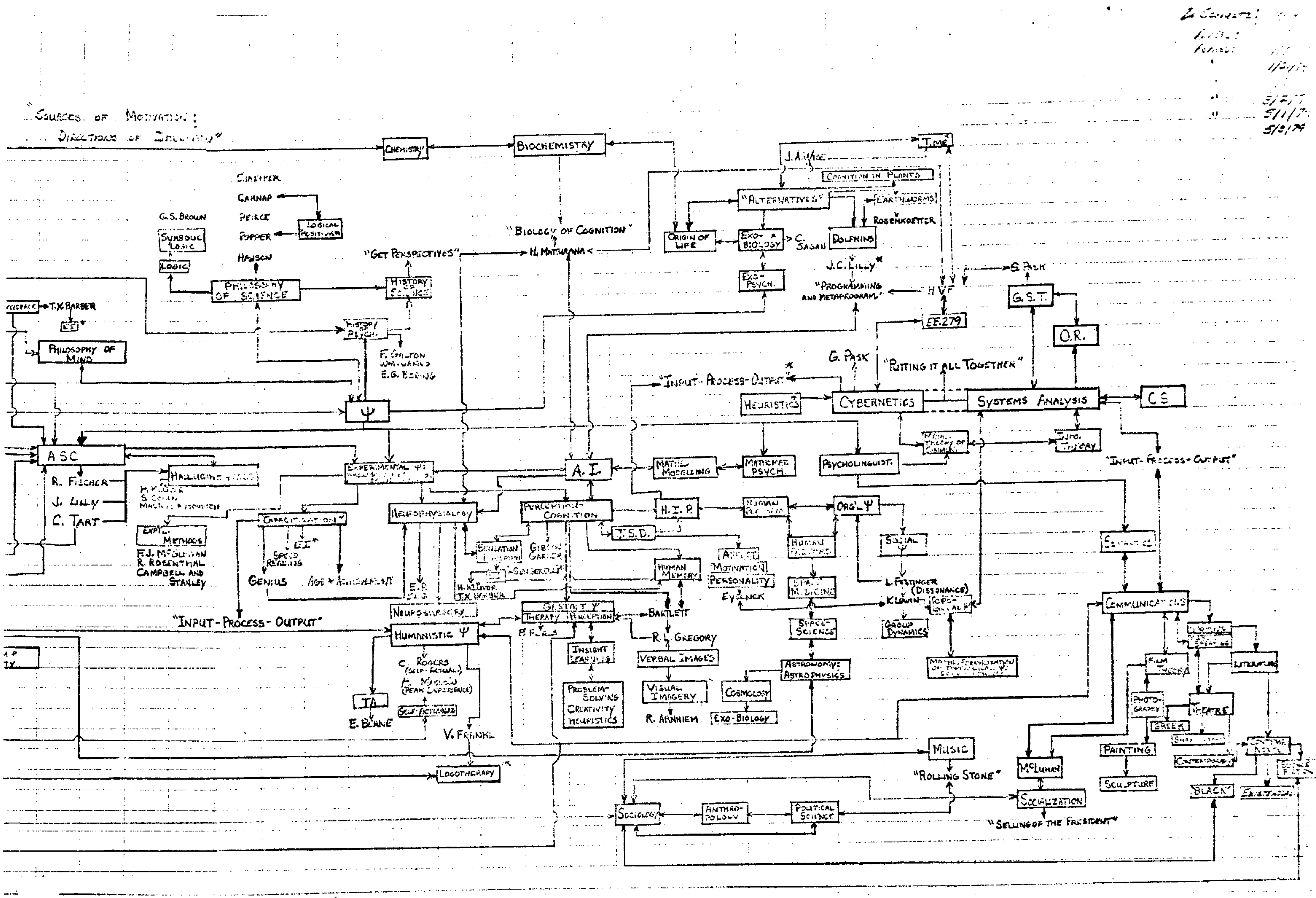
Department of Psychology
University of Illinois
Champaign, Illinois 61820



In presentation, we have taken the long diagram represented in Fig. 1, and broken it into two pages, with overlap in the adjacent margins (i.e., margins of pages x and x+1 are on opposite sides of the same sheet of paging). Though this procedure tends to destroy, or at least vitiate, the schematicism of the structure, the aforementioned logical progression on the horizontal axis lends unity. Finally, many abbreviations are employed in the structure, for which I present the following key:

AI - artificial intelligence
ASC - altered states of consciousness
CS - computer science
EEG - electroencephalography
EI - eidetic imagery
EP - evoked potentials
HIP - human information processing
HVF - Heinz Von Foerster; University of Illinois, Biological Computer Laboratory
OOBE - out-of-the-body experiences
OR - operations research
Org'l ψ - organizational psychology
PK - psychokinesis
ψ - psychology
TA - Transactional analysis
TSD - theory of signal detection
* - "this factor is represented in at least one place in the entailment structure."


MEDITATION
R. ORNSTEIN
C. NARANJO
D.T. SUZUKI
J. Lilly
TIME
HYPNOSIS
E. HILGARD
T. X. BARBER
OOBE
S. KRIPPNER
M. ULLMAN
"PUTTING IT ALL TOGETHER"
FAITH HEALING
PHILOSOPHY OF RELIGION
MAN AND
UNIVERSE
SELF
MAN
FUTURE
THEOSOPHY
THEOLOGY OF PARAPSYCHOLOGY
(MY BOOK)
J. HICK
REINCARNATION
C. CASTANEDA
O. ICHAZO
J. LILLY
KARMA
CLOSURE IN LIFE: GESTALT: PURPOSE
COLLECTIVE UNCONSC.
JUNG
JUNGIAN EGO
OF RELIGIOUS EXPERIENCE"
BA'HAI
THERAVETA
ZEN
MAHAYANA
HATHA RAJA
BHAKTI
KARMA
HINDUISM
TANTRIC VEDIC
E. RAM DASS
"PERSONAL HAPPINESS"
KABBALAH
K. LORENTZ
"NEOCHRISTIANITY"
("JESUS FREAKS")
R.A. HEINLEIN

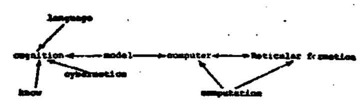

"Sources of Motivation:
Directions of Inquiry"
Chemistry
Biochemistry
"Biology of Cognition"
H. Maturana
Carnap
Peirce
Popper
Hanson
Logical Positivism
G.S. Brown
Symbolic Logic
Logic
Philosophy of Science
"Get Perspectives"
History of Science
History of Psych.
F. Galton
E.G. Boring
T.X. Barber
Philosophy of Mind
Ψ
ASC
R. Fischer
J. Lilly
C. Tart
Experimental Ψ
Exptl. Methods
F.J. McGuigan
R. Rosenthal
Campbell and Stanley
EI
Speed Reading
Genius
"Input-Process-Output"
Neurophysiology
Neurosurgery
Humanistic Ψ
C. Rogers
(Self-Actual.)
(Peak Experience)
Self-Actualiz.
TA
E. Berne
V. Frankl
Logotherapy
A.I.
Perception-Cognition
T.S.D.
Gestalt Ψ
Therapy
Perception
F. Perls
Insight Learning
Problem-Solving
Creativity
Heuristics
Human Memory
Bartlett
R.L. Gregory
Verbal Images
Visual Imagery
R. Arnheim
Personality
Eysenck
Motivation
Cosmology
Exo-Biology
Astronomy
Astro Physics
Space Science
Math'l Modelling
Mathemat. Psych.
H.I.P.
Org'l Ψ
Social
L. Festinger
(Dissonance)
K. Lewin
Group Dynamics
Psycholinguist.
Heuristics
Cybernetics
Systems Analysis
C.S.
G. Pask
"Putting it all Together"
"Input-Process-Output"
EE.279
G.S.T.
O.R.
S. Pask
H.V.F.
"Programming and Metaprogram"
J.C. Lilly
J.A. Wise
Time
Cognition in Plants
"Alternatives"
Earthworms
Rosenkoetter
Origin of Life
Exo-Biology
C. Sagan
Dolphins
Exo-Psych.
Info. Theory
"Input-Process-Output"
Communications
Literature
Film Theory
Photo-graphy
Theatre
Greek
Painting
Sculpture
McLuhan
Socialization
"Selling of the President"
Music
"Rolling Stone"
Black
Existential
Sociology
Anthro-pology
Political Science

COMPUTER

Warren McCulloch employed this term to describe both "computational domains" in the central nervous systems of animals and manufactured processing systems. The principal computational domains or, briefly computers that he identified in vertebrate animals are:

cerebral cortex which McCulloch called "the great computer"
retina
core of the reticular formation
basal ganglia
cerebellum

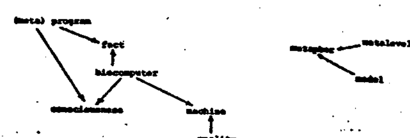
In our laboratory we have devised operating models of stereopsis, including a primitive retina and the core of the reticular formation. Our object is both to increase our understanding of these computers and to develop practical devices like them.

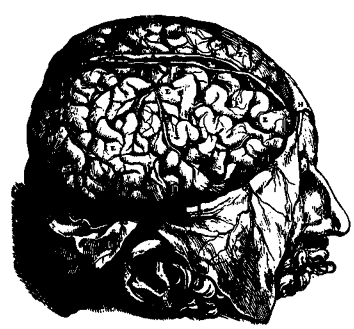


(See Reticular Formation) [L.S.]

MACHINE

That which is configured from materials(solid, liquid, gas, etc) by a biocomputer through its quantum observer-operator outside itself in its external reality which has at least some extension/creation of decision/choice capability of the given biocomputor/s. [J.L.]

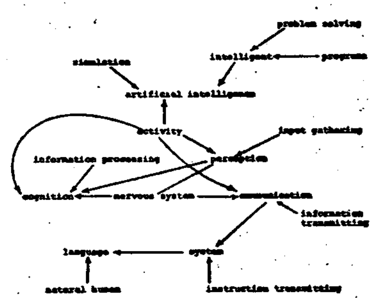


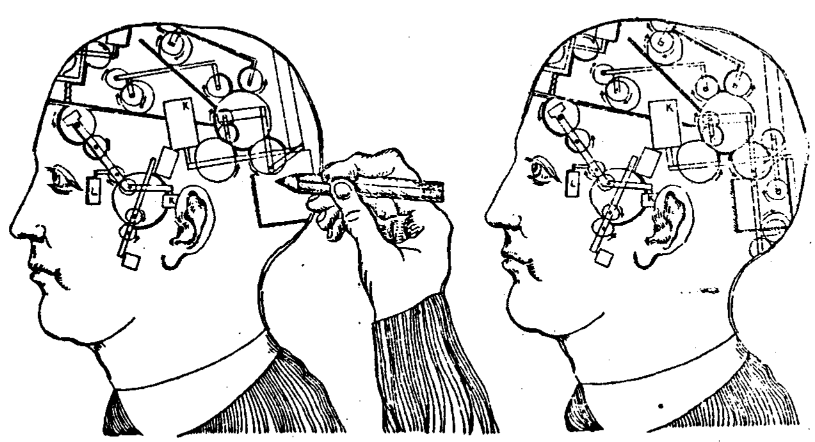


ANY SYSTEM WITH A PURPOSE FOR IT (ANY SYSTEM FOR WHICH A CYBERNETIC THEORY CAN BE CONSTRUCTED) ALSO HAS A PURPOSE IN IT, I.E. A GOAL; ALL SYSTEMS ARE GOAL DIRECTED SYSTEMS.

ARTIFICIAL INTELLIGENCE

The simulation by computer programs of activity considered intelligent. The difficulty is in deciding what is intelligent in such a way that not all computer programs are automatically included (after all, an arithmetic operation can also be considered intelligent). The common solution of this difficulty is to define intelligence as problem-sovling and give a list of problem-solving tasks that have been attempted in artificial intelligence research, such as game-playing (chess, checkers), theorem-proving, etc. [P.G.]





Amosov

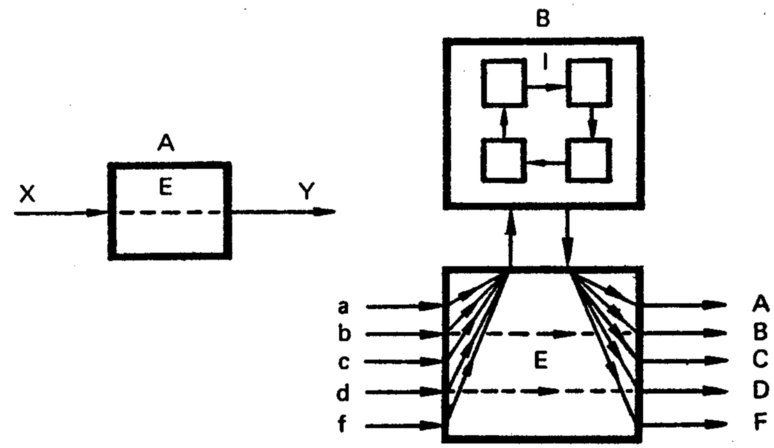
Simulation of Thinking Processes

NICOLAI M. AMOSOV
Institute of Cybernetics
Academy of Science of USSR
Kiev

Let me begin with some general statements. Any system is composed of a number of different elements, united by connections and functioning as a whole. Complex systems differ from simple ones in that they not only transform energy, but process information as well (Fig. 1).

The information which is processed by a system we consider as a certain knowledge about an object, usually presented in the form of a model. The model reflects—with some simplifications and distortions—the structure and functions of the object. Model and information are inseparable. The model is often described by a human language, creating a "model of a model" within which any sequence of actions of the system is designated by the word "program."

Simulation or modeling of the object called human thinking is necessary for the sake of studying human cognition, for creation of an artificial intelligence, and for development of sociological or economic models of human interactions. The most general model of a human being is that represented by an automaton with three types of programs: "for himself"—the instinct of self-preservation; "for stock" (family)—instinct of reproduction; and "for species" or "for society"—the programs of social behavior.



A - simple system:
E - energy, X - input, Y - output

B - complex system:
a - f - inputs, A - F - outputs

E - transformations of energy

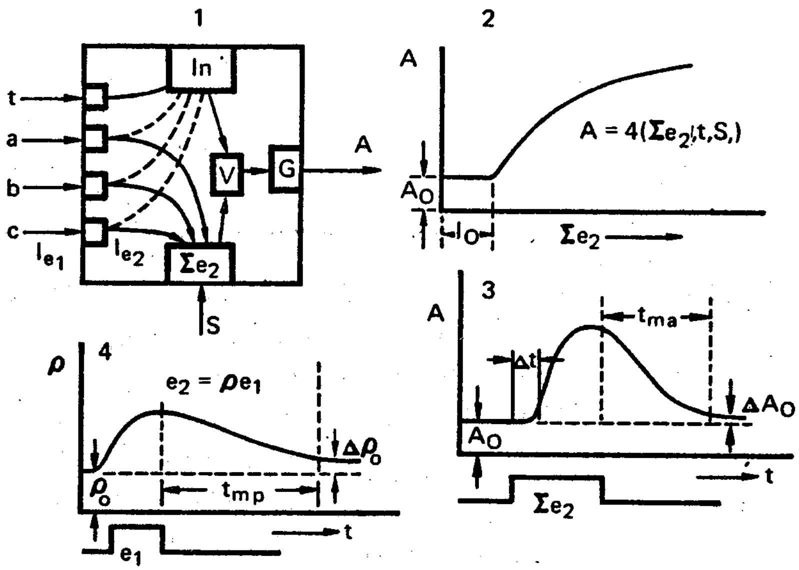
I - selections and processing of information

Figure 1. Systems.

Each of these three programs functions in the human on several layers: a rigid congenital part, laid in the subcortex and endocrine system; another, implanted by learning, in the lower area of the cortex; and a third one, generated during the process of creative work, in the highest areas of the cortex.

In our model we shall represent the cortex by a network of neurons, united by an infinite number of connections which are slightly changeable as the model undergoes some form of activity. As there is not yet much possibility of building up a neurophysiological model of the brain explaining the mental processes of man, we propose one founded upon programs of processing information in a certain hypothetical system. This system does not claim to have complete similarity with the brain, but it can be realized by a model.

The conventional diagram of an informational model is shown in Fig. 2. This figure shows also the static and dynamic charac-



a, b, c – inputs, t – time, e_1 – outer energy of input
e_2 – inner energy of input, conductance of input
A – output, G – generator of activity
Σe_2 – energetical part
In – informational part
V – "veto," S – additional input, reinforcing or inhibitory
2 – static chatacteristics of model of dependence of output energy upon input energy, 1_O – threshold, A_O – model's own activity
3 – dynamic characteristics of model
ΔA_O – increase of own activity, t_{ma} – "active memory"
t – time of delay
4 – dynamic characteristics of connections
ρ_O – level of long-term memory, $\Delta \rho_O$ – increase of memory after functioning
t_{mp} – short-term memory of connection

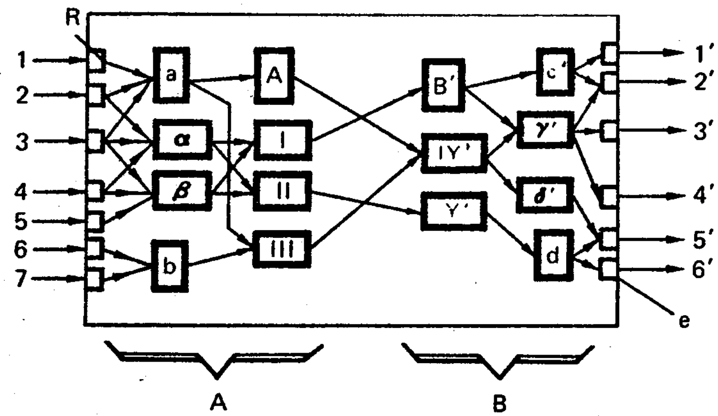
Figure 2. Block diagram of informational model.

teristics of the "generator of activity" of the model and the increase in the facility of connections following their use, that represent the main properties of the neuronal network. One should observe:

1. the formation of the conditional connection between two or more models excited simultaneously.
2. three kinds of memory:
   a. short-term active memory (t_{ma}) which is realized by a prolonged excitation of the model after the stimulus has ceased its action.
   b. short-term passive memory (t_{mp}) which is realized by the increase of the efficiency of connections during a certain period of time after some energy has passed through them and caused excitation, a property which gives the possibility to "recollect" all the chain of activity during some time after excitation.
   c. long-term memory which is realized by a stable increase of facility of connections after manifold repetitions (Δ p) and decrease of facility when connections are not used.
3. increase of activity of a trained model is expressed by lowering the threshold of excitability (1_o) and an increase in the level of spontaneous activity (A_o).
4. the parameters of activity can be changed (decreased or increased) by the central regulatory system (System of Reinforcement and Inhibition—SRI) or by the adjunction of activities.

Figure 3 shows the simplified diagram of a neuron-like network of a complex system with a large number of "inputs" and "outputs," and with two layers of semantic models, united by connections. The activity of the system is realized by the changing activity of the informational models and the circulation of the energy of the signals along the connections. As a result of this activity the outer influences upon the system, their meaning and qualities, are "recognized" and the programs of "response actions" start.

Unfortunately, such a network does not work. The reason is that the same combination of receptors (when there are many of them) can excite a lot of models simultaneously, and they start many contradictory, antagonistic movements. A mechanism is needed to inhibit most of these movements, leaving only the one which is most expedient at a given moment.

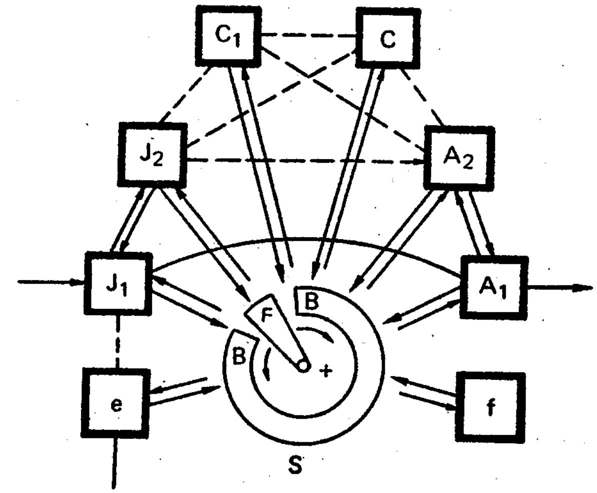


A - "ascending," B - "descending" flows of information
1 ... 7 - "inputs," R - their receptors
a, b - spatial models }
α, β - temporal models } the 1st layer
A - spatial models }
I, II, III - temporal models } the 2nd layer
B, IY, Y - spatial and temporal models of the 2nd layer
c', d, γ', δ' - spatial and temporal models of the 1st layer
e - effectors (muscles), outputs 1 - 6

Figure 3. Block diagram of information-processing network of complex system.

The problem of dominance is solved in our hypothesis by means of the System of Reinforcement and Inhibition (SRI) (Fig. 4). Its function is to select and reinforce (intensify) only one model or activity at any given moment. All the others are automatically inhibited. In a short time the connection of this model with the SRI becomes "tired," and reinforcement turns to the model having the next highest activity.

The functioning of SRI permits us to explain such terms as consciousness and subconsciousness, thinking and thought. Consciousness is a program providing at any given moment the domination of activity of one model over the others. Any model, including an image of an outer world, feeling, action, percep-



S - SRI F - reinforcement of a model
B - inhibition of the others

Different cortical models

J_1, J_2 - layered (hierarchical) models of meaning of the surrounding world
A_1, A_2 - hierarchical models of actions directed outwards
C_1, C - models of programs of consciousness
e, f - models of feelings and emotions

Figure 4. Block diagram of SRI.

tion, desire, or so on, may be in the sphere of consciousness. The model reinforced at a given moment is a *thought*. As SRI cannot stay too long upon one model, the reinforcement regularly turns to some other model, then to a third one. All of this chain of models is connected by the short-term memory.

The processing of information, i.e., the movement of the excitation from one model to another, takes place not only in the sphere of consciousness, along the chain of excited models, but it also takes place in the subconscious, between models that are partly (not completely) inhibited. The role of the subconscious is very large; the amount of information processed subconsciously is much more than the quantity of information processed in the consciousness, although the level of activity of the subconscious models is lower. The primary "recognition" of images and preparations of models for "capturing" consciousness in the next moment takes place in the subconscious, as does the starting of learned programs of actions and the influencing of feelings.

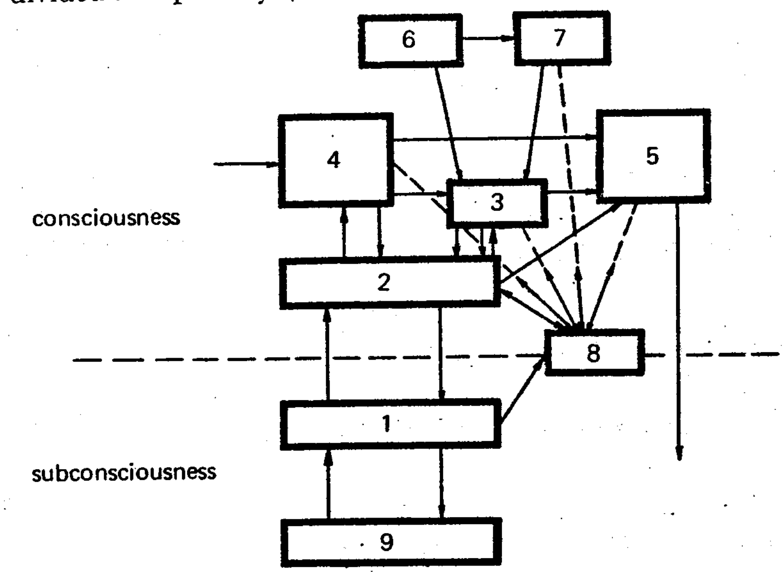
It is possible to point out six types of programs that determine human behavior. We shall go on to examine each in detail. They are:

1. Program of bodily feeling—long perception and processing of information from the body.
2. Perception of outer influences and recognition of models of meaning and qualities of the surrounding world.
3. Program of action—imparting of energy and information to the surrounding world.
4. Program of speech.
5. Consciousness.
6. Creativity and labor.

Programs of Feeling

Feelings are cortical reflections of the excitation of subcortical centers, representing instincts and animal programs "for himself" and "for stock." But in the process of learning and creative work the primary models of feelings expand. The activity of some models is suppressed, whereas the activity of others is reinforced. At the same time, the models of the outer world acquire conditional connections with feelings and thus they always take part in processing information by the program "for society." Figure 5 shows the rough block-diagram of interrelations of subcortical centers. Figure 6 shows a simplified block-diagram of models of feelings in the cortex, connected with models of the surrounding world and SRI. The connections give an element of subjectivity to the processing of information.

It is difficult to classify feelings, but it seems to be convenient to deduce the classification from the basic programs of behavior, and to distinguish sensations, feelings, and emotions, according to their intensity. Besides, the feelings can be conventionally divided into primary (or simple) and secondary (or complex).



1 - subcortical centers/models/of feelings
2 - cortical models of "simple" feelings
3 - cortical models of complex feelings
4 - models of surrounding world
5 - models of movements
6 - time
7 - models of consciousness
8 - SRI
9 - body

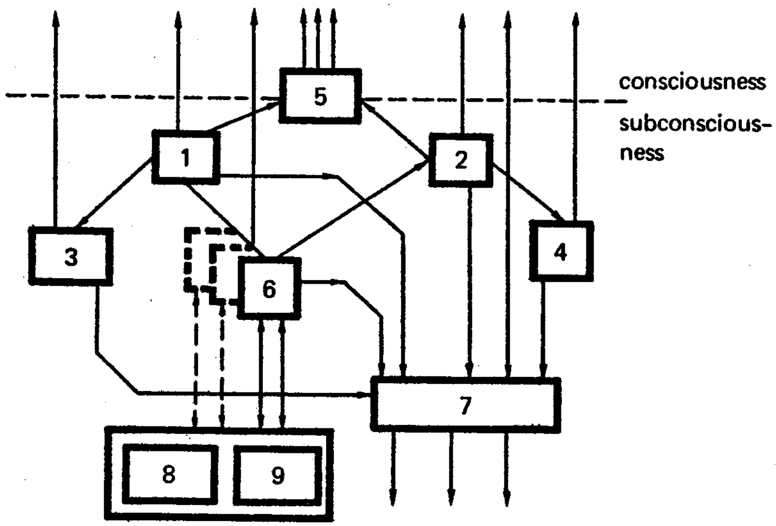
Figure 5. Interrelations of cortical and subcortical emotional centers.

Simple feelings reflect directly in the cortex the excitation of the subcortical centers connected with the body. Complex feelings are represented by models that have formed as a result of interaction of several feelings in response to outer educative influences, that have no direct connections with biological programs of human beings; for example, patriotism.

As has been already mentioned, the basic "animal" programs "for himself," or instinct of self-preservation, consist of two instincts—"defense" and "nutrition"; programs "for stock" (family), or instinct of reproduction, consist of two instincts—sexual and paternal.

Program of Perception and Processing of Outer Information

Figure 7 shows the basic elements realizing this program: the "main receptor" (R), perceiving the outer influences; the system of attenuation (T-MT-RT) of R with muscles and receptors;



1 - center of "pleasure"

2 - center of "pain"

3 - positive emotions

4 - negative emotions

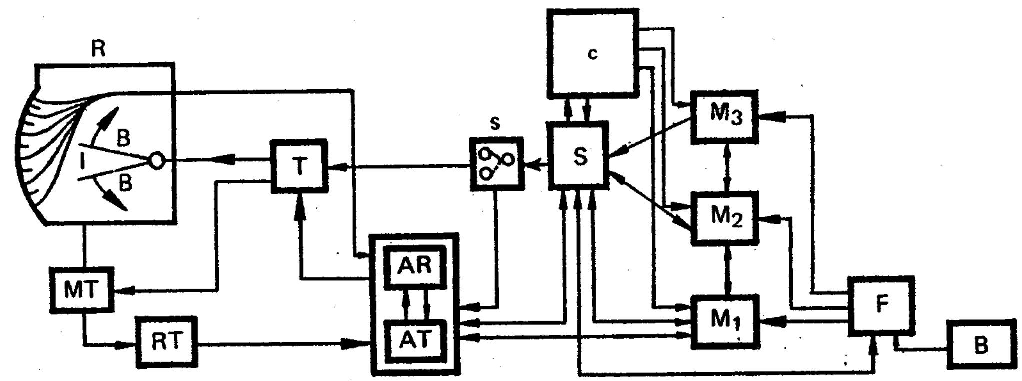
5 - SRI

6 - one of simple feelings

7 - centers of actions - unconditional reflexes, facial expressions, postures

8 - visceral organs } body
9 - endocrine glands

Figure 6. Subcortical centers of feelings.



R	- "main" receptor, MT - muscular attenuation
I	- reinforcement of a part of receptive neurons
B	- inhibition of the others neurons
RT	- receptor of attenuation, T - attenuation
AR	- analyzer of the main receptor, AT - analyzer of attenuation.
S - SRI, s	- relay of SRI from receptor to analyzer
c	- models of consciousness, F - feelings, B - body
M_1, M_2, M_3	- layered (hierarchical) models of surrounding world

Figure 7. Models of perception of outer effects, selection and processing information.

cortical analyzers of the "main receptor" (AR) and receptors of attenuation (AT), that primarily reflect images of the outer world; SRI(S) with its relay of attention-reinforcement (from receptors to analyzers); and the system of layered (hierarchial) models (M_1, M_2, M_3) of meaning and qualities, that process information from analyzers and are connected with feelings and models of consciousness.

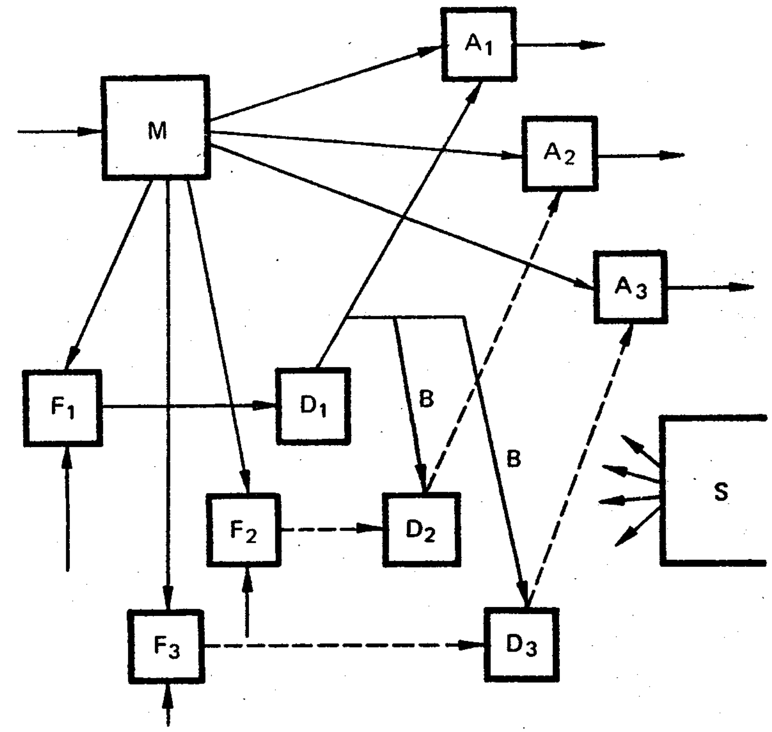
The program consists of the following subprograms:

1. Attenuation gives an orientation to the body with respect to the outer world and determines the distance to separate objects. Information about this is stored in the models of the analyzer of attenuation. They are connected with the models of images, perceived by the "main receptor" and reflected in the "main analyzer."
2. SRI can separate receptor from analyzer, so the excited models of outer objects are viewed upon by the "inner sight," and the excitation passes from them along connections to models of higher layers where they are recognized.
3. Temporal sequence of perceived and recognized images is kept for some time by the short-term memory within the connections. If the sequence is frequently repeated, it is memorized by the long-term memory, and new models are built.
4. A model stored in the memory of the analyzer can be recollected, i.e., excited from other models, when SRI separates the analyzer from receptors.

Programs of Action

These movements are the main means for a human being to affect the outer world. In such a way both energy and information are transferred outward.

In a human being the control and regulation of muscles is a function of the cortex and is realized on the basis of conditioned reflexes, i.e., of connections, formed during the individual's life. The general block-diagram of choice of actions is shown in Fig. 8. The model of stimulus M has connections with several models of action $A_1 \ldots A_3$. Sometimes these models are of an antagonistic character. Choice of a certain model is determined by the stimuli of feelings (F) and desires (D). The program reinforced by bodily feelings is started. All connections are conditional, and subordinated to the laws of memory. The model of action has a very complex structure and is formed very gradually from early childhood. Perhaps it comes from "reading out" information from visual and auditory images.



M - model of outer influence (effect)

F_1, F_2, F_3, - feelings

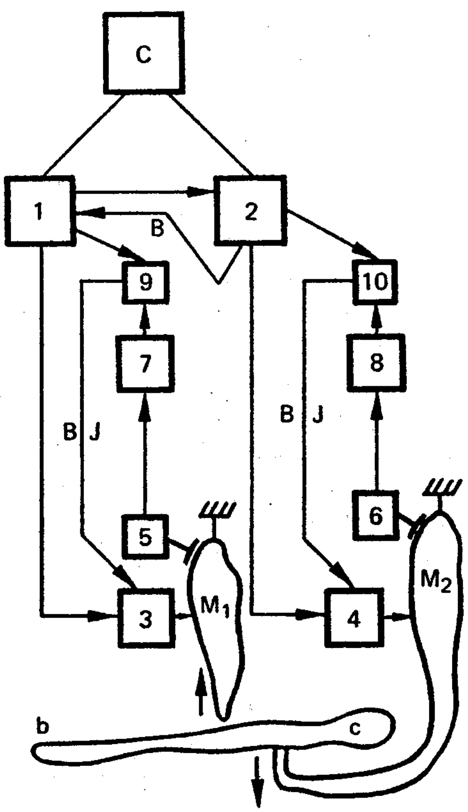
D_1, D_2, D_3, - wishes (volitions)

A_1, A_2, A_3, - models of different variants of actions

S - SRI

Figure 8. Choice of program of action.

A block-diagram of a model of movement is shown in Fig. 9. Its main model is laid in the analyzer; it is a system of neurons (1-2), connected by the usable connections with "delay." Each element of the model is connected with motorneurons (3-4), controlling the contractions of a certain muscle. The feedback from the muscle tells the degree of its contraction (5-6). Feedbacks are connected with a monitoring system, in which the



C - generalized model of the higher layer

1 - 2 - the "main model" in analyzer

3 - 4 - motoneurons of muscles

5 - 6 - receptors of muscles

7 - 8 - short-term memory

9 - 10 - models of comparison

B - J - inhibition or reinforcement

M_1 - M_2 - muscles-antagonists

b - c - bone

Figure 9. Block diagram of the model of movements.

short-term model of movement is formed. Then it is compared with a long-term model in the analyzer (9-10) and, if there is a discrepancy, reinforcement or inhibition is sent to the motorneurons.

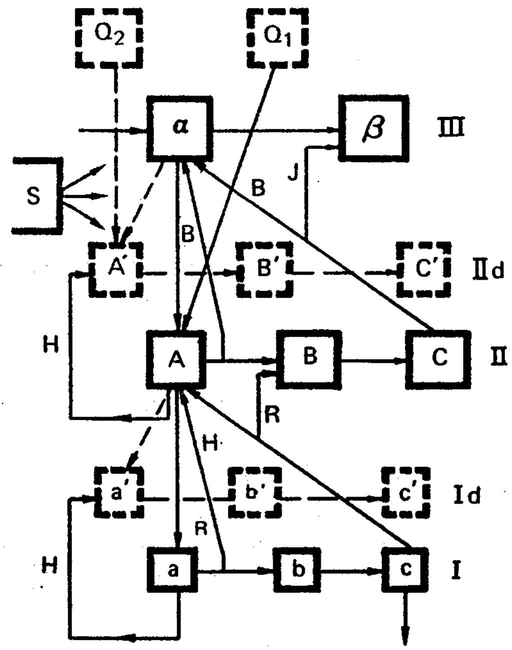
The simplest model is shown in the block-diagram. Real models are much more complex, because there are many muscles, several layers of models, and the action itself is controlled by many parameters. For each parameter there are models, monitoring systems, that record discrepancies and give correcting signals to motorneurons. In such a way any shift or displacement is detected by muscular receptors and by visual ones independently. Its short-term memory is formed in analyzers with additional qualities—speed and acceleration. The strength of contraction and fatigue are detected separately. The result of any effect upon a certain object is detected by means of a visual analyzer, and feelings are connected with the fulfillment of the action and its final result.

The model of a complex motion, as well as models of the outer world, have a many-layered structure. All layers are built similarly. There are some models reflecting different qualities; each model has feedback and monitoring systems. But this is not all. One upper layer model corresponds to several lower layer models. Only one of them functions at a time, a choice determined from "above" by models of qualities, and from below by the variant that is "ready" to fulfill the program. There are connections between the models of variants; when one of them is excited, all the rest are inhibited (Fig. 10).

Program of Speech

Speech, or the Pavlovian second signal system, is an informational language that enables us: firstly, to encode complex, not completely determined phenomena; secondly, to transfer information from one person to another; and, thirdly, to record information and to store it separately from the brain, by means of physical models.

Besides this, speech is important for learning and for building complex models in the cortical system. Models of speech work as a parallel system of concrete images. Information is processed in these systems simultaneously, while consciousness is alternatively involved in both; when one system is in the sphere of consciousness, the other is in the sphere of the subconscious



I, II, III, - models of layered (hierarchical) program of actions

I d, II d, - reserve variant of layers S - SRI

α - β - stages (models) of program of higher layer

A, B, C - models of the middle layer, main variant

a, b, c - models of the lower layer

A', B', C' - models of the middle layer, reserve variant

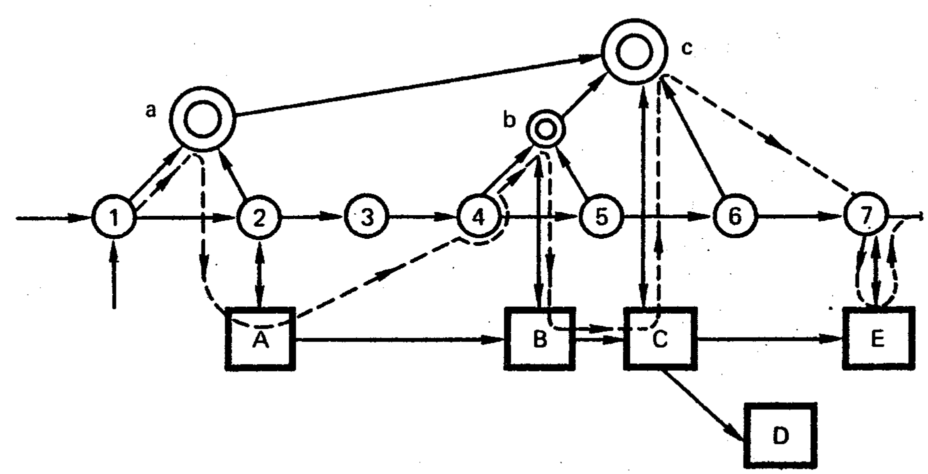
a', b', c' - models of lower layer, reserve variant

q Q_1, Q_2 - models of qualities

H - inhibitory connections, R - reinforcing connections

Figure 10. Hierarchical (layered) models of actions.

(Fig. 11, where the dotted line shows switching of consciousness).



1...7 - words, a,b,c, - models of phrases (spuch)

A ... E - models of images, ——→ - connections

- - - - → - consecutive order of switching of the consciousness

Figure 11. Two systems of models.

Structure of speech reflects the character of the information—layered complex models integrate spatial, temporal, and qualitative relations. Besides this, speech reflects the basic programs of a human being, as an individual and as an element of society ("for himself," "for stock," "for society"). All this is expressed by different means in different languages.

Any speech can be represented graphically as a set of models, things, feelings, shifts in space, or changes in time (which can be represented in other systems of coordinates as changes in space). It is reflected in the brain by the order and strength of excitation (activity) of different models. A block-diagram of perception, comprehension, and memorizing of speech is shown on Fig. 12.

The diagram shows that at first all the heard words (1-15) are memorized in the form of activating their models, while their sequence is memorized by activating connections. The connections with feelings and images α, β, γ of different significance are established immediately; and thus the activity of some words is increased. In the same time comprehension of meaning of separate phrases is going on. The models, A, B and "own" words corresponding to them are excited. The connections with feelings are established, and the activity of the significant models of meaning and their words increases.

The short-term connections between "own" words and memory are established. At the same time, models of concrete feelings and images α, β, γ are activated and connected with memory by short-term connections. Models of meaning have their own words, and these words (not heard ones) are used to express memorized speech. Less significant models are forgotten and as a result only the most important models of meaning remain. These models are expressed by key words and connected with strong feelings.

Consciousness plays an important part in all these processes. Being involved in interactions of active models, it reinforces them and inhibits less active ones. During intervals in the perception of speech, consciousness (SRI) comes back to more significant models, reinforces them, and thus promotes memorizing.

Somebody else's speech, as well as meaning, "superimpose" on one's own matrices and are modified by these matrices in whatever way is most suitable. This is memorized and becomes one's own.

Pronunciation of speech, as any other action, is composed of contents and stimuli. Each question can be answered in different, even opposite, ways. A complex of models, excited by a question, is connected with several models of response actions, and the choice of one of them is determined by feelings and desires as shown by the block-diagram, Fig. 13.

At first models of meaning M_1, M_2, M_3 and their key words

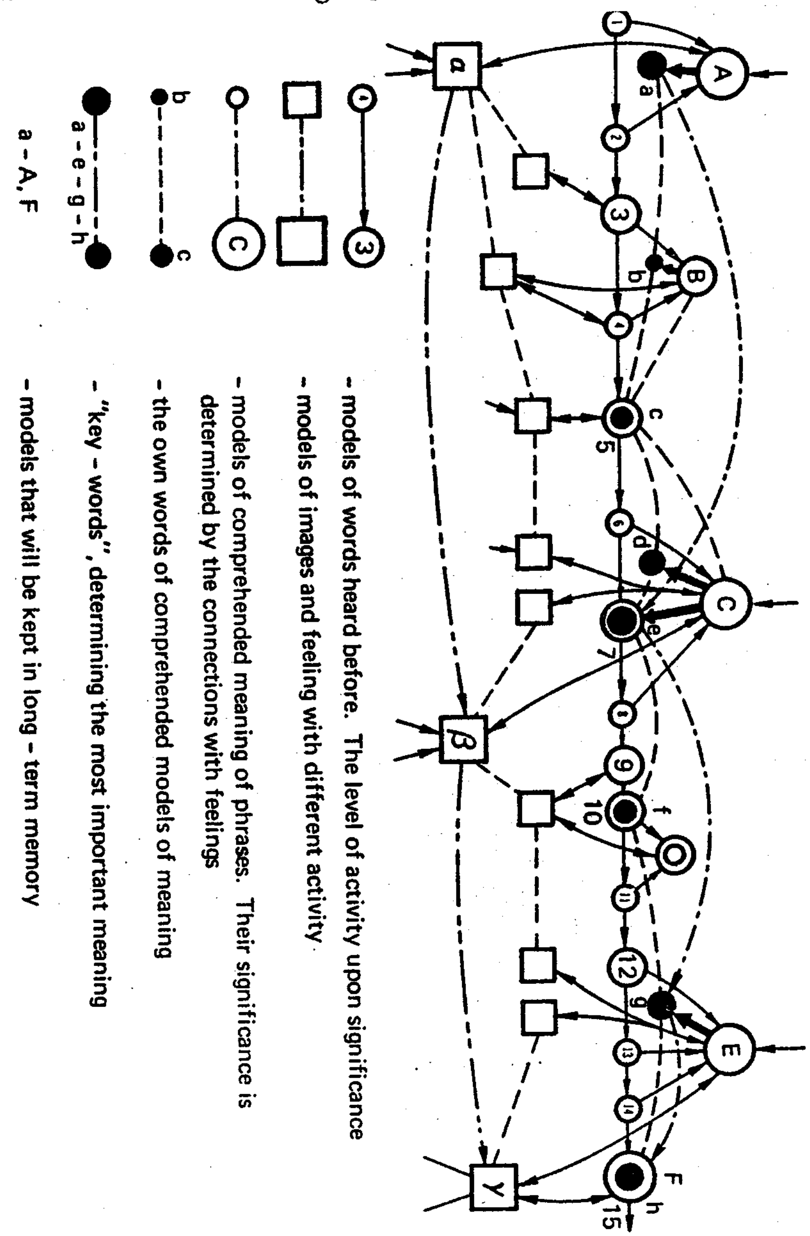


Figure 12.

Figure 12. Perception, comprehension, and memorizing of speech.

1, . . . , 5 are excited; being connected between themselves they represent a plan of speech. Models of qualities Q_1, Q_2, Q_3 and

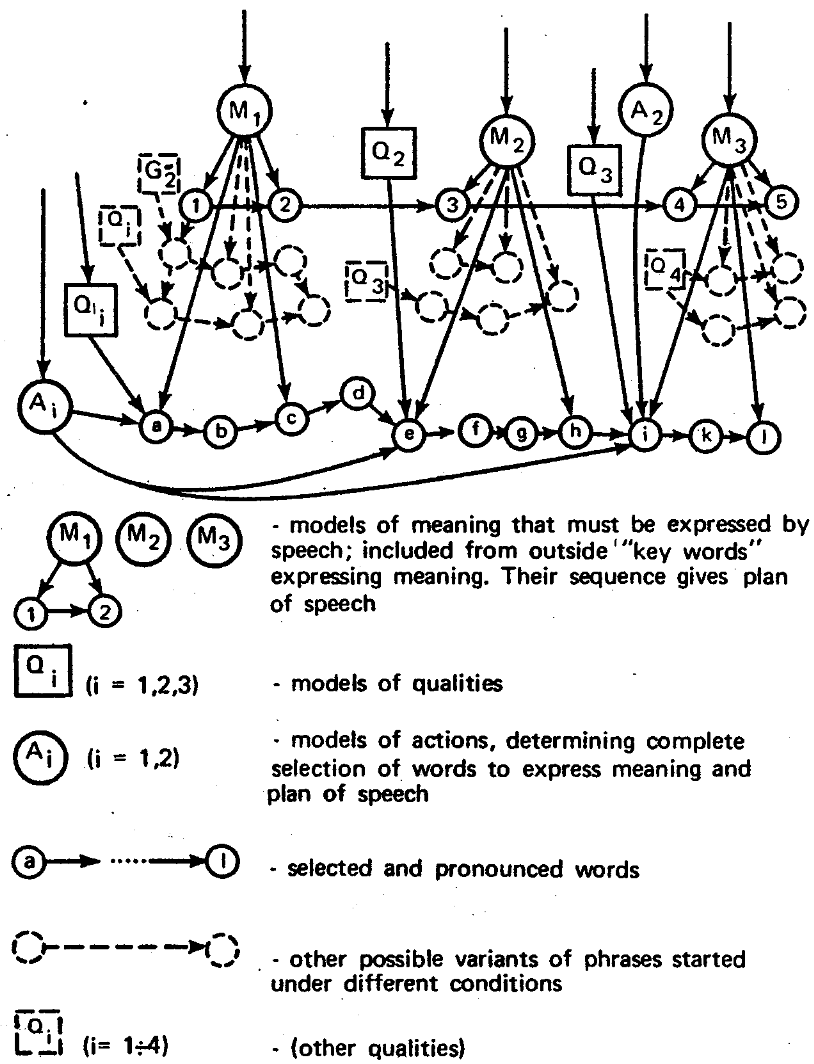


Figure 13. Formation and pronunciation of speech.

the stimulus to action A are also excited. Interaction of all these models determines the consecutive order of "reading out" meaning of pronounced words, a, . . . , e, which must be regarded as only one of the possible variants to express a certain meaning that satisfies given qualities. Different qualities q_1, q_2, q_3 require different words.

Programs of Consciousness

Consciousness can be defined as simulation of spatial, temporal, causal-sequential interrelations of objects and subjects, and informational interrelations, when models themselves are simulated. We can point out several levels of consciousness, according to the complexity of the programs.

1. Attention. Selection and reinforcement of one cortical model and inhibition of the rest.
2. Formation of models "Ego" and "Non-Ego," models of time, distinguishing between real and unreal. Simulation of one's own action.
3. "Simulation of models"—of one's own and someone else's thoughts.

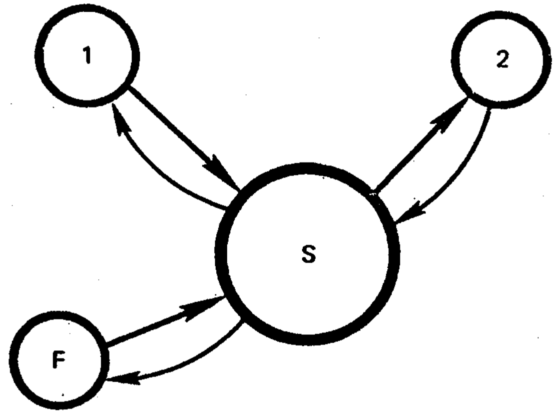
The program of will stands by itself, as a capacity to concentrate attention arbitrarily, to reinforce selected models, giving them high activity and dominance over others.

All these programs can be represented in the form of structural models. Let me describe them briefly.

1. *Attention.* The program of attention is a foundation of all the others; it is connected with SRI, which is able to reinforce activity of one model and to inhibit all the rest. Feelings are the main source of activity of SRI, but SRI also includes positive feedback—reinforcing a model gives back stimulating impulses to SRI. The training of will is based on this fact. As a result of repeated training, selected models increase their activity, and reinforce activity of SRI through the connections. In such a way an arbitrary concentration of attention is provided which can be directed to any model (Fig. 14).

2. *Models of Time.* Conditioned connections with delay are the basis for the modeling of time; they reflect true temporal relations between two models during a period of establishing connections.

Evidently, time has no dimension in memory; there is no "clock" in the brain. The true value of time is established only by means of language, by means of designation of minutes,



1 - model of the word "overcome"

2 - model of the word "pain"

F - feeling

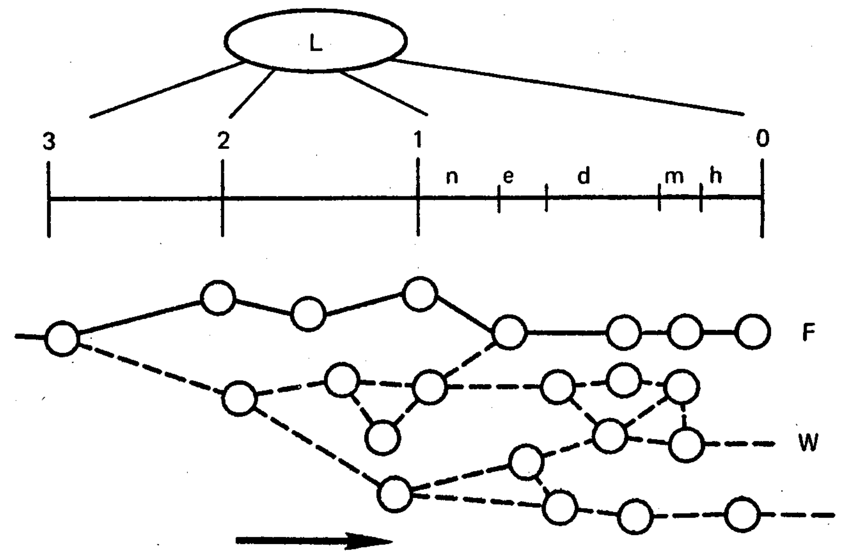
S - SRI

Figure 14. Arbitrary switching of attention.

hours, days, years, and other events, connected to form a united network—change of time, of day, or of seasons. Our model must be capable of accelerating or slowing down time, while still preserving its direction.

Figure 15 shows the block-diagram of "the past." It can be depicted by a system of models, united to form some lines, connected with each other and tied up to a calendar, a clock, and alternation of day and night. These models can be realized in visual images, sounds, or words and phrases; all these have connections with a model of the word "past." To recollect the past one must "go ahead" from a certain date or event that has been indicated from outside, or has been recollected by chance.

The present is only one flowing moment, when a human perceives the world and his attention is concentrated upon the data of his receptors. But perceived information is reflected in the analyzers and the layered models; and their activities last much longer. Thus the limits of the present are expanded and involve a part of the past (Fig. 16).



L - word "past"

1,2,3, - days

n,m,d,e, - time of day

○——○ - models of events

F - "line of family"

W - "line of job"

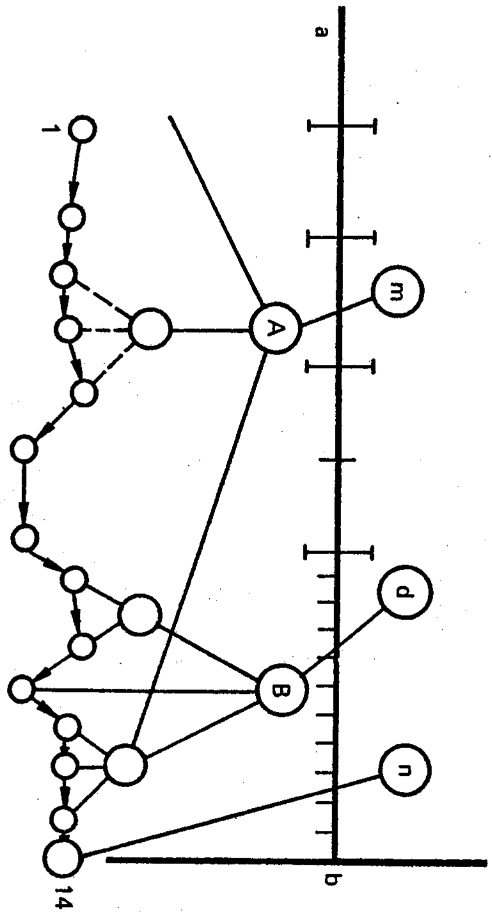
R - "line of rest"

L - model of the word "past"

Figure 15. Model of the past.

The future is a probabilistic extrapolation of the present. The program of anticipation can be regarded as an excitation of models, tied up to models of time and repeating the past and the present. This program is organized by means of language, and is expressed by models of words, although some concrete images of the past are tied up to them (Fig. 17).

Chains of events of different types, with different cyclic recurrence, and different qualities took place in the past. These events are extended to the future and different cycles are superimposed. As a rule it is impossible to find all connections between the past events; therefore, the anticipation of the future is probabilistic.



m - phrase "in this month"

d - word "today"

n - word "now"

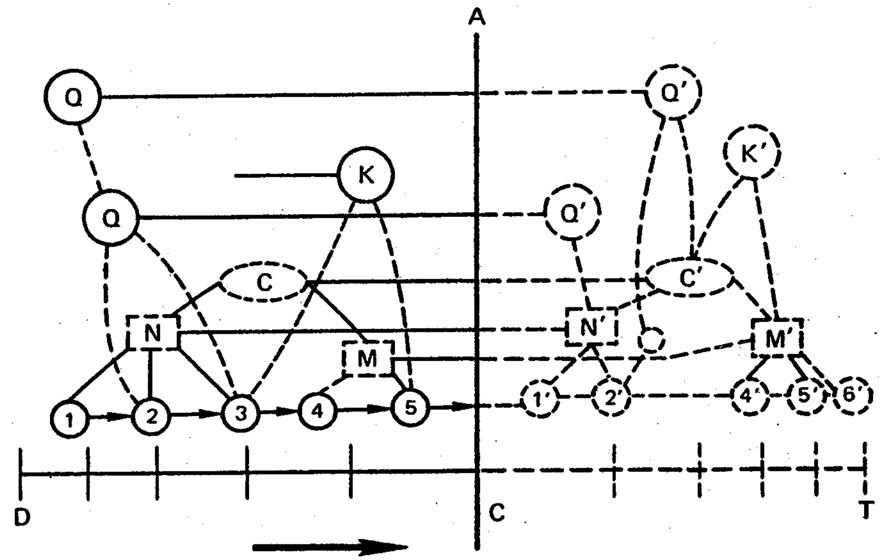
1 - 14 - models of events and their connections in memory

A - B - generalizing models

a - b - scale of time

Figure 16. Model of the present.

"Switching" from the present to the past or future is carried out arbitrarily by linkage of corresponding actions with models. The models of "words," "to recollect," "to anticipate," and "suppose" are selected; receptors are switched off by SRI. These models are initiated from the outside or get to the sphere of consciousness by chance, having been prepared by complex processes in the subconscious, where the general program of examining any phenomenon is established. This program con-



A-C - the present

D-C - the past, scale of time

C-T - the future

1,...,5 - chain of events in the past

1',...,6' - supposed events in the future

N,M,C - different cycles of events in the past

N',M',C' - supposed cycles in the future

Q - qualities in the past

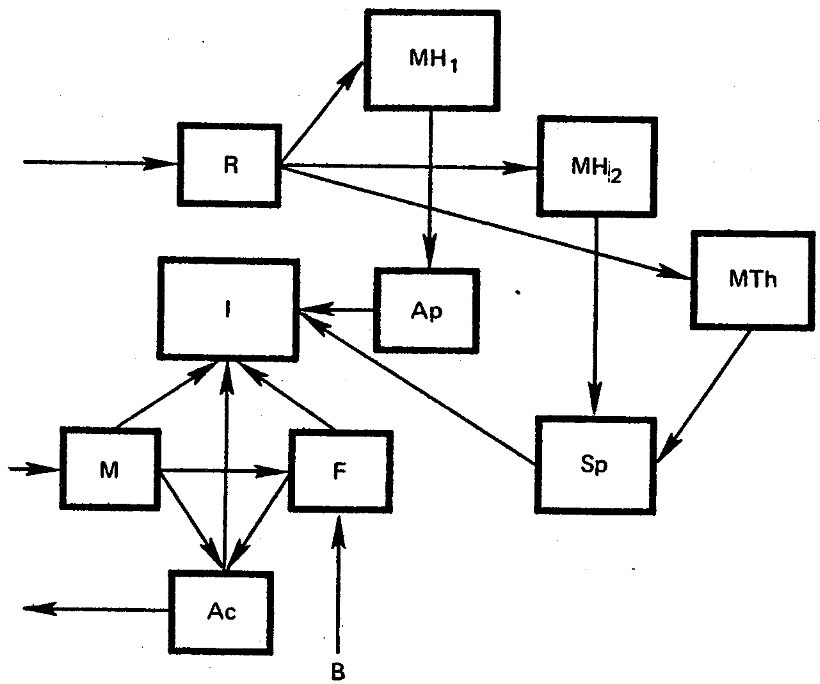
Q' - the same qualities in the future

K' - new qualities in the future

Figure 17. The future.

sists of perception of the present, examination of the past, and anticipation of the future.

The distinction between reality and unreality is realized by building up emotional and verbal models connected with proper names. An estimation of people and things is worked out, based upon feelings of sympathy and aversion and derivations from pleasure and pain that have formed as a result of personal experience and education. A block-diagram is shown in Fig. 18.



I - model of "Ego"

M - model of effect

Ac - models of response action

F - models of feelings

B - body

R - receptors

MH_1, MH_2 - models of human beings

MTh - models of things

Ap - feeling of aversion (antipathy)

Sp - feeling of sympathy

Figure 18. Model of "ego" and "not-ego."

3. *Models of Other Persons.* The notion of someone else's feelings is formed by means of perception of his outward appearance and actions in response to stimuli. They are compared with one's model of one's self, and one's own feelings are spread upon another person, with some correction, depending upon sympathy or aversion for that person.

Any human with a high level of consciousness has many models of third persons in his cortex. The actions of such third persons cannot be exhausted by the scheme—stimulus-feeling-response. This is only the first motive which is corrected after modeling of the relation of third persons both to stimulus and to response. Action is a result of integration of programs of all "present" third persons, with corrections, depending upon sympathy and aversion to them. Figure 19 shows how the plan satisfying the third person is chosen, as a result of comparison of three variants taking into consideration one's own and other persons' interests.

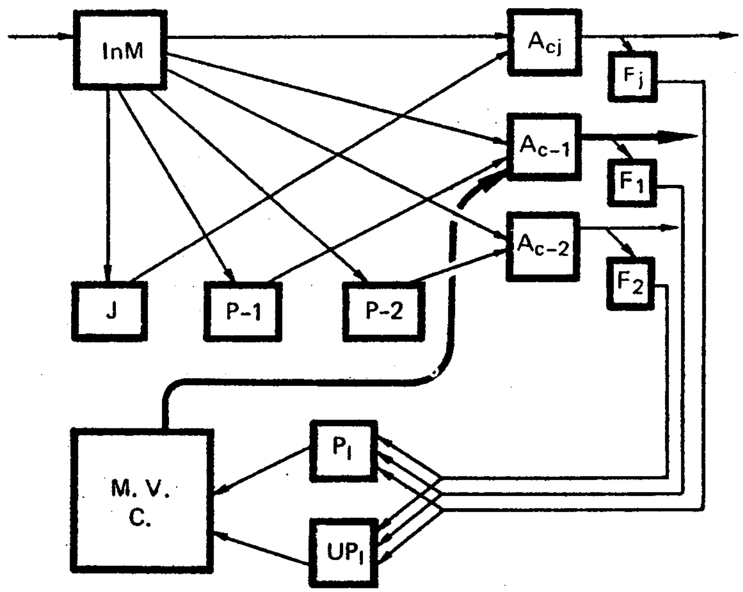
Programs of Creativity and Labor

The essence of creative work is in creating new models in the cerebral cortex. The nature of labor is to incorporate these cortical models into physical things. Every new model is represented by a combination of a number of familiar elements in space and time.

The program of creativity consists of the following stages:

1. Clearing out the task-set of models of meaning and qualities on the upper layers.
2. Search of necessary elements and their combinations to satisfy higher models. Creation of variants of models in memory.
3. Expression of the variants by physical means in the form of things.
4. Repeated perception and examination of the physical model according to criteria of meaning and qualities. Finding out new, unforeseen properties.
5. Correction of the model or even rejection of it. Beginning a new work.

Any realization of model-labor is connected with programs of movements and is accompanied by antagonistic feelings—fatigue, disappointment, or joy (if the work is successful). The



InM – models of stimulus, J – complex of feelings and wishes of a person

P_1, P_2 – models of feelings and wishes of other persons

A_{cj}, A_{c-1}, A_{c-2} – variants of possible response actions

F_j, F_1, F_2 – presumed feelings, connected with fulfillment of action

P_l, UP_l – components of pleasure – pain, connected with different variants of actions

M. V. C. – memory, estimation of variants, choice of the most pleasant, taking into consideration himself and other persons

Figure 19. Block diagram of choices of a variant of action, taking into consideration its action upon third persons.

completion of work requires will, i.e., capacity to reinforce the models that are necessary and to inhibit obstacles.

Programs of creativity and labor can be exposed in the form of block-diagrams, but they are too complex to be described in my report.

Conclusion

We have presented a hypothesis about mechanisms of human thinking and behavior; it is based upon general principles of simulation and processing information. It can be expressed in the form of a complex semantic network composed of elementary informational models, with changing degrees of activity, and united by changing connections. An artificial intelligence with all the attributes of personality can be created on this foundation. Unfortunately, great difficulties prevent us from fulfilling this task, but these difficulties are of a technical nature.

A model of purposeful behavior has been built to test the suggestions described above. It is an automaton consisting of two principal blocks (Figs. 20 and 21): one block to construct a preliminary plan of behavior on the conventional locality; the other block designed to simulate the processing of information by the human brain while realizing certain movements.

The first block is an algorithmic description of building up a plan to achieve a certain goal, given by the programmer.

The second block is a network composed of elements. It is a structural model of processing information while fulfilling the chosen plan of behavior. The consecutive order of elements selected by SRI (System of Reinforcement and Inhibition) is interpreted as the reasoning of the automaton. Semantic interpretation of elements of the network enables one to observe the motivation of choice of certain actions by the automaton and estimate its reasoning. Changing of the levels of the activity of elements not selected by SRI is compared with subconscious processing of information.

The automation is incorporated in a set of programs for the computer.

The general picture of its surroundings where the experimental investigation was carried on is shown in Fig. 22. The dotted line designates the chosen plan of movement; the arrows show the route "walked" by the automaton while fulfilling the plan. Objects of the surrounding world were interpreted as "forest," "beast," and "food."

Changing the parameters of the network, it is possible to reproduce human behavior with different types of characteristics.

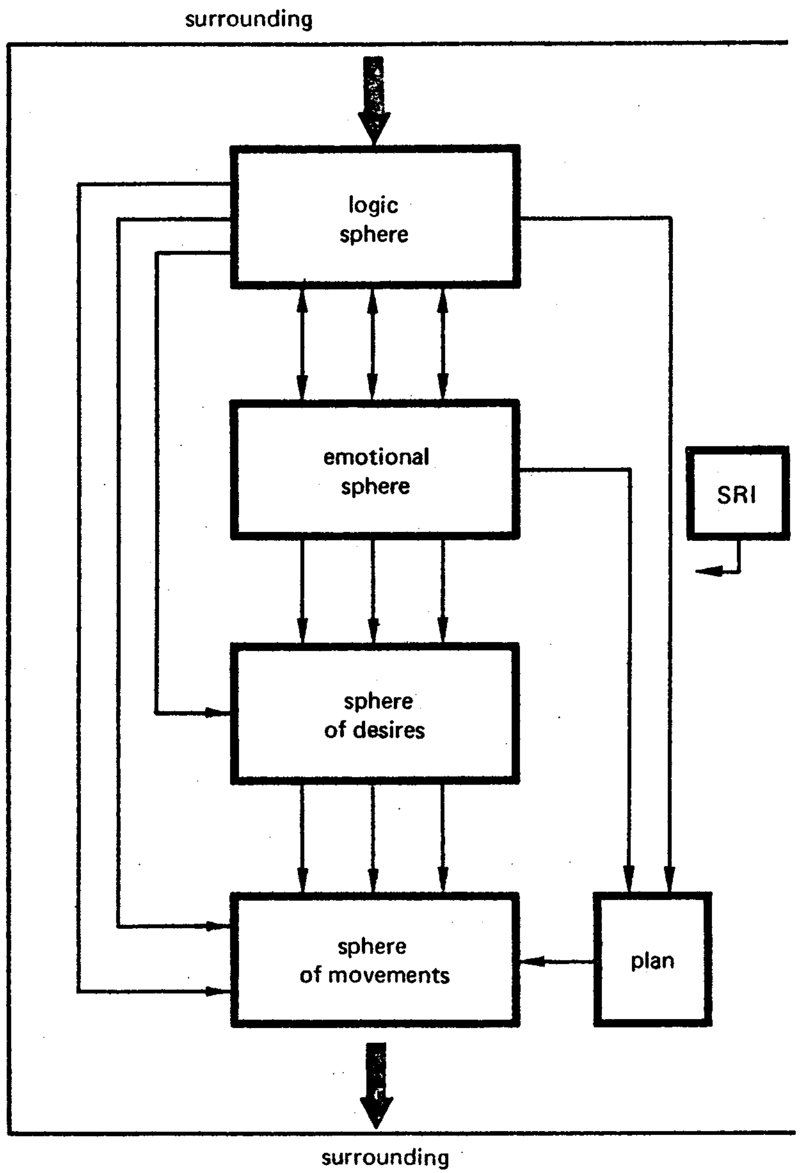


Figure 20. Block diagram of automation.

Changing parameters of SRI, one can observe changes in the mechanisms of attention, and correlations of "consciousness" and "subconsciousness."

Programs of the automaton approximately simulate the following presumed programs of the psyche:

1. Hierarchical processing of information and hierarchical structure of program of movement.
2. Program of anticipation—building up a plan.
3. Participation of feelings in building up a plan.
4. Program of elementary "consciousness" and "subconsciousness."

The accomplished work has enabled us to come to the conclusion that an automaton of the proposed type can serve to study the basic mechanisms of nervous activity, both normal and abnormal.

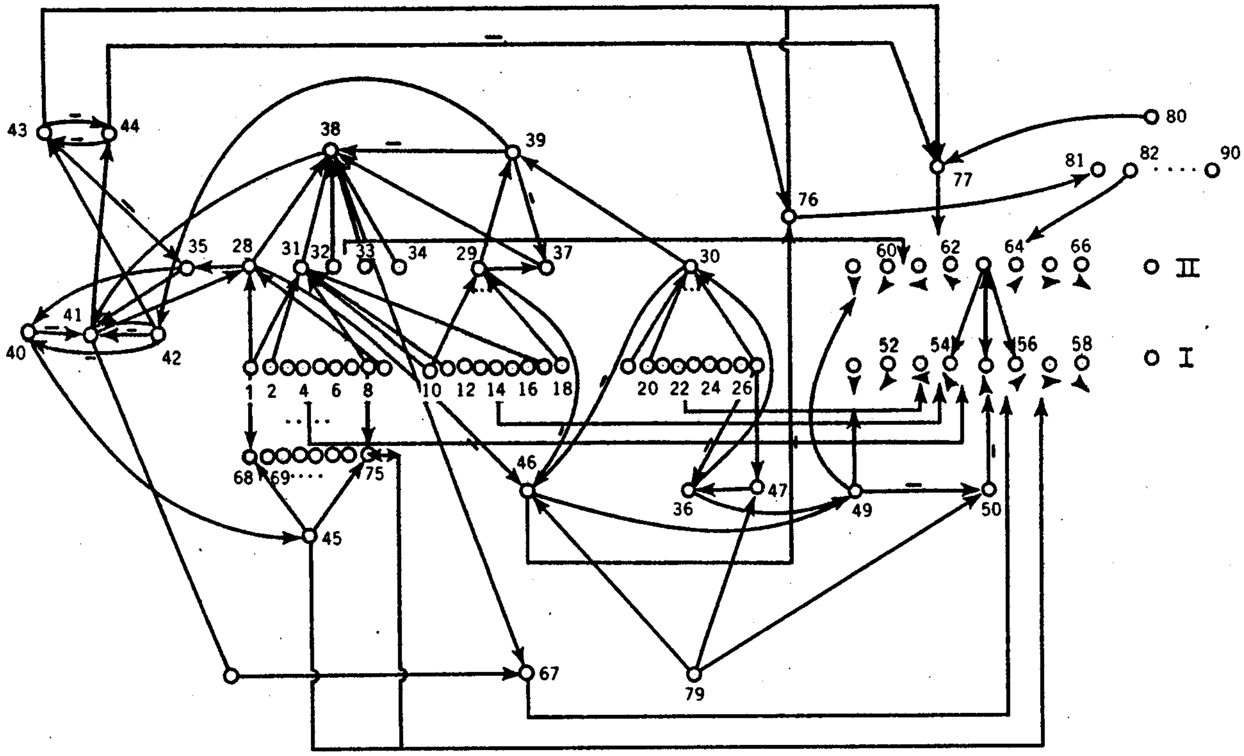


Figure 21. Semantic network of automation.

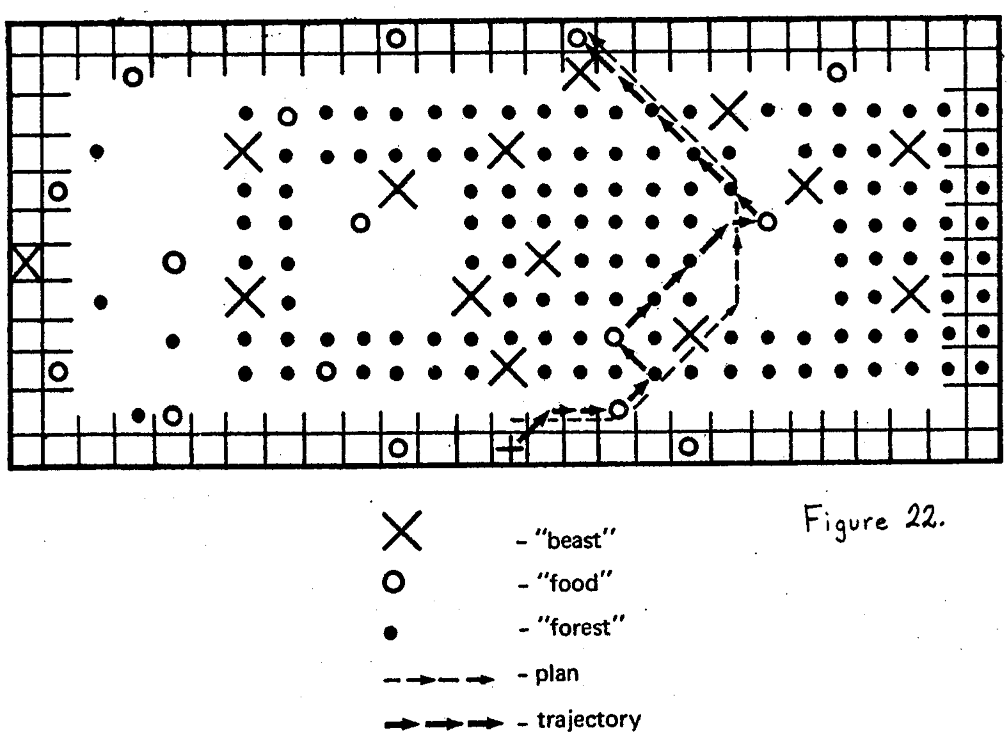
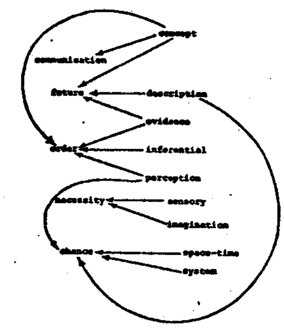


Figure 22.

FUTURE



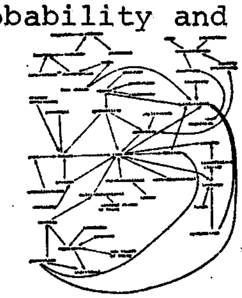
Designation of regressive description.
The word read after is is future.
Evidence of an absent undiscovered conceptual basis.
Dysfunctional assignment.
Constantly refuted nominal position.
Easily discarded or retained.
Inviolate.
Nominal preposition unrelated to assumed preparation.
Verb without precedent.
The end of declension.
Recognition of unrecognized presence.
Essential perspective only in relativistic describing activity.

[E.S.]

FUTURE

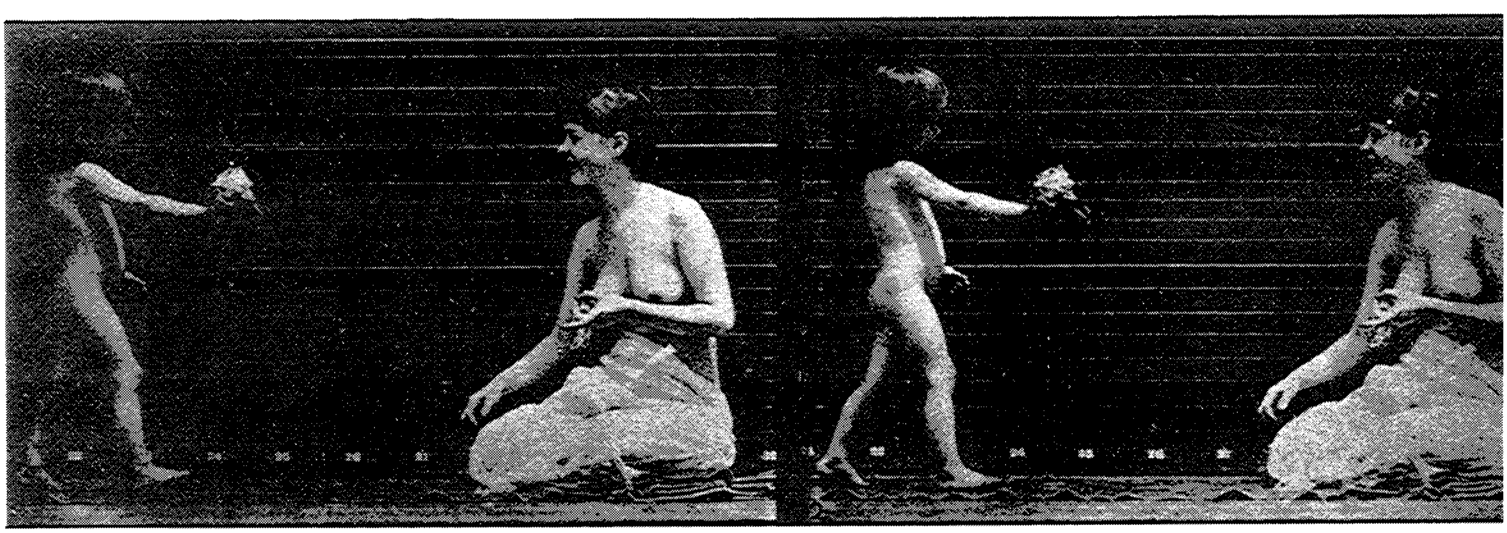
Future is a notion indissolubly bound with present and past and hence with TIME. The root of the word future lies in the Latin fuo from the Greek phyo in turn from the primitive root phy linked with the Sanskrit bhu, Old Irish biu, English be (come). All these roots refer to becoming, bringing forth, putting forth, i.e. development of potential. So future means the possibility of actualizing potentials. The past refers to already produced actualizations, and the present, to the momentary state of such actualization. Thus the notion of future is intimately related with the notion of CHANCE in the context of possibility.

But future has an even more interesting and accessibly evocatory meaning; namely, that by the exercise of DESIRE in terms of appropriate ORDERings of what we have by virtue of past and present, we can often to a critical extent, choose what kind of future we will have. Thus desire may increase probability and thus act as a self-fulfilling or future-guidance system. Dynamic resonance meditation (Journal for the Study of Consciousness, vol. 5, no. 2, p. 144) makes use of this future-oriented capability of directed or focused desire. [C.M.]



TIME

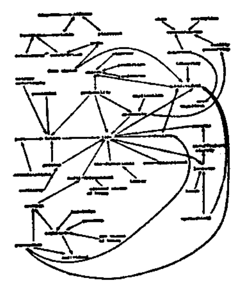
A problem that becomes relevant by the cybernetics heuristic is the definition of "time." We know of time, but not time itself. It certainly reflects the cognitive manifestations of memory. Without memory, time is imperceptible. Further, William James noted in Principles of Psychology that "Awareness of change is thus the condition on which our perception of time's hour depends." To know time, we must remember at least some change. The Lorentz-Fitzgerald contraction equations indicate that time stops for an observer travelling at c, the velocity of light. There can be no perceptible change at c, for volume has been reduced to zero and the mass of that non-volume to infinity. We know the manifestations of time, yet not time itself. A somewhat amorphous discourse leads me to the following concluding suggestion: Cybernetics, as an heuristic, entails the necessary scope to approach understanding of time itself. By cybernetic heuristics the silver on the mirror of time may be removed and the transparency of temporality may be perceived. Time is evanescent: not an entity, but the reflection of entities. Cybernetic heuristics feedback the mirror images of time and ourselves. The epistemological bastion of temporality is reperceived as the rigidity of human cognition.... [D.S.]



TIME

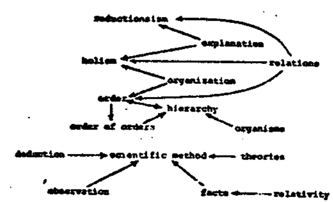
TIME is the ongoingness of potentials into regions of increasing probability, culminating in that configuration of certainties we call "the present" or "in the future," i.e. a still not certain or present state. This is one interpretation of entropy increase in a closed region.

The FUTURE is branching probability, whereas looked at from the present toward the past, historical time is always a linear succession of moments, each moment or time point holding in itself all the spaces (of whatever dimension) of the cosmos. In this sense space is the zeroth dimension of time. That higher time dimensions exist is shown by the time-distortion phenomena in altered states of consciousness (cf. the writer's paper "Trance states, precognition, and the nature of time" in Journal for the Study of Consciousness, p. 77, vol. 5, no. 1, 1972). In fact, there is a hypernumber transformation best expressed by $t \to e^t$, where $e^t = t^0 + t^1 + t^2/2 + t^3/6 + t^4/24 + t^5/120 + \ldots$, t being time and t^0 space, as already mentioned; $t = t^1$ is thus ordinary or linear time, t^2 is the two-dimensional time of most noetically altered states, noetics (as we have used the word since 1967) being the science of the nature and changes of perceptual and cognitive states. Time is also deeply related to LANGUAGE. The nature of time accounts for some PARADOXES. [C.M.]



ORDER

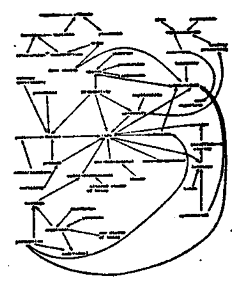
Pervasive throughout the living and physical worlds, and possibly the link through which they may be related in a fundamental way, order basically consists of a set of "similar differences" which, although they may be distinguished, also mutually relate to one another. Two types of differences may be distinguished, "constitutive differences" which determine the essence of the order, and "distinctive differences" which form some pattern and in another sense the formation of part of an overall hierarchy of orders. So what at one level of consideration is a constitutive difference may at another be a distinctive difference, and "the organization of orders into hierarchies may proceed without limit." (see D. Bohn "Some Remarks on the Notion of Order" in Towards a Theoretical Biology, vol 2, ed. C. H. Waddington.) [S.H.]



ORDER

Order is ultimately placement or fixed location. The fact that there may be regular patterns or sequences in the array of locations is not as basic as the fact that the locations are fixed and each is assigned a fixed category of occupants. Indeed any such placement-array is a pattern. We showed (in our 1974 paper "How to make a stupid machine clever by ruthlessly opportunistic programming") that the concept of a named variable is simply that of a location for a thus-specified set of occupants (the values of the variable). Hence a place-value system of numbers is homomorphic with the address-data system used in computer programming. So, variable = address = location code.

Proceeding to a high hierarchy of order, an instruction is a set of rules for manipulating data and their addresses. Then, by the above concept-equations, a function (a set of rules for variables and their values) is the homomorph of an instruction. As we also pointed out in 1962 (p. 156, Aspects of the Theory of Artificial Intelligence, ed. C. Muses, Plenum Press, New York), construction itself is a form of instruction, allowing certain things to be in certain ways only. Indeed, the very concept of structure is one of modular order. [C.M.]



ORDER

The spectrum from (inferential).

Results of subsequent determination.

Precipitate of measuring activity.

Consensual request.

Framed objective qualification.

Presupposition of problem orientation and generator of problem structure.

Evidence of the seeming duplication of phenomena.

Spatial consequent determined through arbitrary designation.

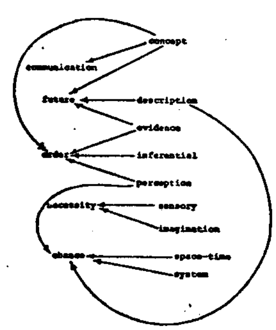
Emotional (glandular) determination.

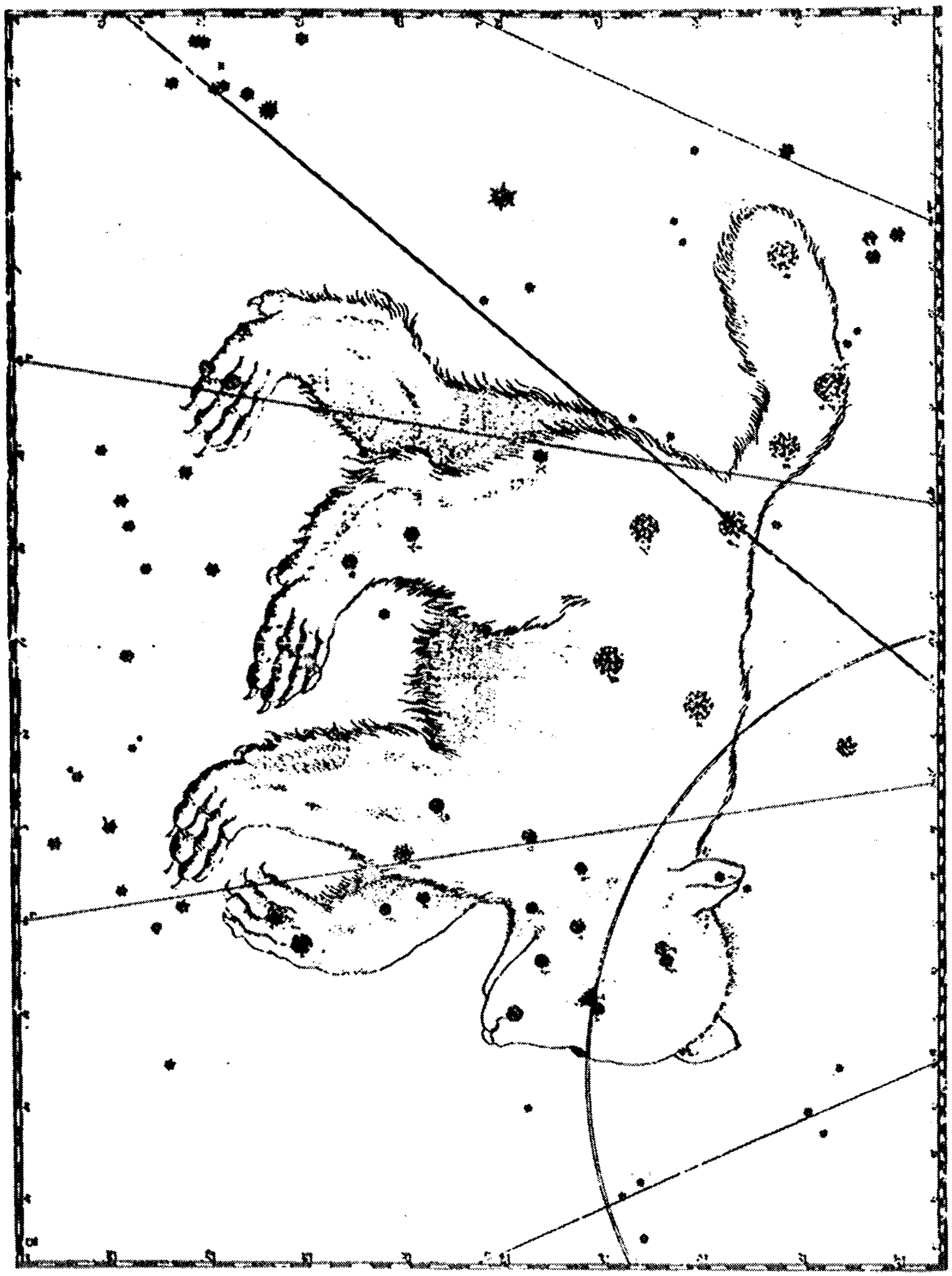
Entropic comparative formulation.

Conceptual trial.

Imposed arbitrary perceptual differentiator.

[E.S.]





Recognition of Order and Evolutionary Systems

Lars Löfgren
LUND INSTITUTE OF TECHNOLOGY
LUND, SWEDEN

The regularities of an ordered object (pattern) permits the object to be effectively described by a comparatively short description. Let ξ be a pattern, i.e., an arbitrary string of symbols from a finite alphabet. Let x be an initial tape expression that causes a universal Turing machine U to compute ξ. For each ξ and U there are many such x, but only one which is shorter than all the others. This shortest U-form of ξ is symbolized $s(\xi, U)$.

The following thesis concerning our intuitive conception of order is suggested and motivated: there is a U such that of any two equally long patterns ξ and ζ, ξ possesses more order than ζ if and only if $s(\xi, U) < s(\zeta, U)$.

The following thesis concerning our intuitive conception of randomness is suggested and motivated: there is a U such that of any two equally long patterns ξ and ζ, ξ possesses more randomness than ζ if and only if $s(\xi, U) > s(\zeta, U)$.

Various properties of the shortest U-forms are derived. It is for example found that, for no U, the set of shortest U-forms of the computable numbers is recursively enumerable. On the basis of these results and the above hypotheses, conclusions are drawn about the effective recognition of order and about the effective generation of random numbers. It is for example found that random numbers cannot in general be generated algorithmically.

The interdependence of relative order and reproduction is discussed. An object (pattern) A must possess order relative to a surrounding B in order to be reproductive (in an effectively explicable sense). This implies regularities (order) in B which in turn can be obtained if also B is reproductive. Depending upon whether B has sufficient order relative to A or to another surrounding C, we distinguish between symbiosis and, respectively, an hierarchy of reproductive systems. Evolution is discussed briefly and a study of invariances of evolution characteristics over animal formation and theory formation is suggested as a central topic in modern cybernetics.

I. Order and Simplicity

Looking at a surrounding we usually say that it possesses order if we can give a simple description of it.

For example if we have a deck of n cards where each card is marked with an integer out of $\{1, 2, \ldots, n\}$, such that no two cards have the same number, then we can describe each serial arrangement of the cards with a number such as

$$2\ 5\ 7\ 6\ 4\ 3\ 1\ 9\ 8.$$

In this example $n = 9$. Another arrangement of these cards is

$$1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9.$$

This last arrangement can be said to possess more order than the first because we can simply describe it with an expression like "each successor card is marked with $i + 1$ where i is the mark of its predecessor card." It is true that this expression may be considered more complicated than the number arrangement itself. But the description does not contain n, and hence there must be an n such that the description is arbitrarily shorter than the number arrangement itself. This may not be true, however, for some extension of the first arrangement type to some large n. In this way it is natural to say of two arrangements A and B that A possesses more order than B, if the shortest effective description of A is shorter than the shortest effective description of B.

This would lead nowhere, however, if the class of arrangements could not be taken as a representative for some kind of universe of surroundings. Hence it is natural to consider the set of all integers (words) defined on some finite alphabet β, to represent the universe of arrangements. Any such integer ξ (word) can be produced by a universal Turing machine U starting its computation from a suitable initial tape expression x (input word). In this sense we can say that any arrangement ξ can be effectively generated by a U-description x. There may be many x-words (descriptions) that produce (describe) the same ξ, but there is only one which is shorter than all the others, with shortness defined as follows. x is said to be shorter than y if the length (number of symbols) of x is shorter than the length of y, or if of equal length, x comes before y in the lexicographically order defined by the alphabet α for the input words ($\beta \subset \alpha$).

We are thus led to define the notion of *U-order* as follows. An arrangement ξ possesses more U-order than an equally long arrangement η, if and only if the shortest U-description of ξ is shorter than the shortest U-description of η.

Any Turing machine computes either a recursive or a partial recursive function. According to a generally accepted hypothesis (Church's thesis), any algorithm (i.e. an effective procedure for the production of an arrangement from its description) corresponds to a recursive or partial recursive function. This is our main support for the introduction of the notion of U-order, because a universal Turing machine can compute any function that any other Turing machine can.

To go back to the description "each successor card is marked with $i + 1$ where i is the mark of its predecessor card," this is an example of a recursive description. It is fairly short because it utilizes a property for a subsequence of cards, here pairs of cards, which is valid wherever the subsequence occurs. We can then go back (Latin: *recurrere*) to this property at every step of the reconstruction of the pattern from its description.

The problem of recognition of U-order leads us to the problem of detecting regularities in a given symbol sequence ξ. In this connection we want to take up the notion of a random number. From Popper [4] we quote the following discussion of von Mises' axiom of randomness. "The axiom of randomness or, as it is sometimes called, 'the principle of the excluded gambling system,' is designed to give mathematical expressions to the chance-like character of the sequence. Clearly a gambler would be able to improve his chances by the use of a gambling system if sequences of penny tosses showed regularities such as, say, a fairly regular appearance of tails after every run of three heads. Now the axiom of randomness postulates of all collectives that there does not exist a gambling system that can be successfully applied to them." The discussion suggests that we may say of a sequence ξ that it has a high degree of randomness if it shows very little regularities. Hence it is plausible that a nearly random sequence ξ shows a very small U-order, i.e. its shortest description will be about as complicated as ξ itself.

Let us also take up the question of "order in nature." We are inclined to say about nature that it is simple when we have succeeded in giving a "beautiful" theory for some of its processes. Here "beautiful" often means short in the sense that we have discovered some abstract concept of invariance, that permits a short recursive description. Nevertheless the computations that have to be made from this short description, in order to reconstruct the physical event, may be very long and very complicated. But this complication does not influence us at all in our judgement of the simplicity of the event. Again this supports the plausibility of saying that an event ξ is simple (or ordered) when its shortest U-form x is short. Notice that we then do not pay any attention to the length of the computation of ξ from x. It is only the length of x that matters.

Perhaps it is the simplicity of nature that permits our existence and our thinking abilities (rather than the converse). For what we do when we think is to create models of the surroundings that we notice, which indeed include other humans. Most of these models must then be shorter than the things they model, for otherwise we could not have them inside of us.

II. Properties of the Shortest Form Function

In this section we study in some detail the function that generates the shortest U-description of a pattern ξ. Properties of this shortest form function will

enable us, in the next section, to obtain limitations for the recognition of U-order. Concerning some of the terminology used in the present section, the reader is referred to Davis [2].

Let us consider Turing machines defined on a tape alphabet α. The (partial) recursive function computed by a Turing machine Z with Gödel number z, will be denoted $\psi_Z(x)$ or, in some cases, with $[z](x)$. Here x symbolizes an arbitrary "word" on the tape alphabet α. Hence if Z is started off with the initial tape expression x (in standard position with respect to the reading head), and if x belongs to the domain of ψ_Z, the machine will stop after some finite time. The terminal tape expression ξ then produced is the function value $\xi = \psi_Z(x)$. However, if x does not belong to the domain of ψ_Z, the machine will never stop.

The notation $\psi_Z(x)$ may be used even if we in fact consider a function of two or more variables, say u and v. The initial tape expression x is then a suitable encoding of the ordered pair $x = \langle u, v\rangle$, for example the expression u followed by a blank followed by the expression v.

Some Turing machines are, as we have mentioned previously, universal. The universality of a machine is manifested by its ability, via a suitable encoding, to perform any computation which could be performed by any given Turing machine. However, as Davis [3] points out, the condition must surely be added that the encoding itself be, in some suitable sense, simple. For there would not be much point in claiming universality for a Turing machine in which the encoding would require, in essence, another universal machine to carry it out. This led Davis [3] to the following definition.

Definition 1. A Turing machine U is *universal* if and only if the domain D of ψ_U is *complete*, i.e. D is recursively enumerable and for every recursively enumerable set R, there is a recursive function $\sigma(x)$ such that $R = \{x : \sigma(x) \in D\}$.

In what follows a machine denoted U, with or without index, will always be assumed universal in the sense of the above definition.

Definition 2. x is said to be a *description* of ξ *with respect to* U, or x is a *U-form* of ξ, if and only if U computes ξ from the standard initial tape expression x.

For each U and ξ there are many U-forms of ξ, but only one shortest form defined as follows.

Definition 3. x is said to be the *shortest U-form* of ξ if and only if the length of the U-form x is shorter than the length of any other U-form of ξ or, when of equal length, x comes first in the lexicographically order defined by the elements of the tape alphabet α. For each U and ξ, $s(\xi, U)$ symbolizes the shortest U-form of ξ, and $s(\xi, U)$ is called the *shortest form function*.

In order to prove theorems about the shortest form function $s(\xi, U)$, we will need the following lemmas.

Lemma 1. Let U be an arbitrary universal Turing machine. Let $g(x)$ be an arbitrary (partial) recursive function, i.e. there is a Turing machine Z with Gödel number z such that $\psi_Z(x) = g(x)$. Then there exists a *recursive* encoding function $\phi(z, x)$ such that $\psi_U[\phi(z, x)] = g(x)$.

This lemma is proved in Davis [3] on the basis of Definition 1. It could be noted that the recursiveness of the encoding function $\phi(z, x)$ ensures that it can be carried out without the use of another universal machine. For if ϕ, on the contrary, should require a universal machine, ϕ would have to be partial recursive and not recursive; compare Davis [3].

Let $T(z, x, y)$ stand for the predicate "z is the Gödel number of a Turing machine, which from the initial tape expression x produces a sequence of tape expressions, whose Gödel number is y." Further, let $U(y)$ be the terminal tape expression ξ of the sequence of expressions whose Gödel number is y. Then $U[\min_y T(z, x, y)]$ is a partial recursive function,* i.e. there is a Turing machine U_0 that computes it. On the basis of Definition 1, Davis [3] proves the following lemma.

Lemma 2. Any Turing machine U_0, which computes the partial recursive function $\psi_{U_0}(z, x) = U[\min_y T(z, x, y)]$, is universal.

The encoding $\phi_0(z, x)$ that makes U_0 imitate an arbitrary Turing machine Z with Gödel number z, is indeed quite simple. For obviously U_0 will compute $\psi_Z(x)$ $(= [z](x))$ from the initial tape expression $\langle z, x\rangle$, i.e. $\phi_0(z, x) = \langle z, x\rangle$. Thus $\langle z, x\rangle$ is a U_0-form of $[z](x)$.

Let $g(z, x)$ be an arbitrary (partial) recursive function. If z is given a fixed value c, the singulary function $g(c, x)$ will of course in general be distinct from $[c](x)$. However the following recursion theorem of Kleene, conveniently found in Davis [2], ensures the existence of a particular z-value for which the two functions are alike.

Lemma 3. (Kleene's recursion theorem). If $g(z, x)$ is a (partial) recursive function, then there exists a number e such that: $[e](x) = g(e, x)$.

Since $\langle e, x\rangle$ is a U_0-form of $[e](x)$, we immediately obtain the following lemma.

Lemma 4. If $g(x)$ is a (partial) recursive function, then there exists an argument τ such that $\psi_{U_0}(\tau) = g(\tau)$.

As we shall see, Lemma 4 remains true with U substituted for U_0.

* The value of the minimal function, $\min_y P(x, y)$, for the argument x is defined as the smallest integer y such that $P(x, y)$ is true.

Lemma 5. With each universal Turing machine U and each (partial) recursive function $g(x)$, there exists an argument τ such that $\psi_U(\tau) = g(\tau)$.

Let us for the proof consider an arbitrary (partial) recursive function $\zeta(x)$ with Gödel number z, i.e. $[z](x) = \zeta(x)$. By Lemma 1 there is a recursive encoding function $\phi(z, x)$ such that $\psi_U(\phi(z, x)) = \zeta(x) = [z](x)$. Hence for all x and all Gödel numbers z we have $[z](x) = \psi_U(\phi(z, x))$. Consider next another arbitrary (partial) recursive function $g(x)$. Since $\phi(z, x)$ is recursive, the composition $h(z, x) = g(\phi(z, x))$ will be a partial recursive or a recursive function. Hence we can apply Lemma 3 on $h(z, x)$ to obtain $[e](x) = h(e, x)$. Thus: $[e](x) = \psi_U(\phi(e, x)) = h(e, x) = g(\phi(e, x))$, i.e. there is a τ $[= \phi(e, x)]$ such that $\psi_U(\tau) = g(\tau)$. This completes the proof of Lemma 5.

Notice that if the (partial) recursive function g appearing in the last three lemmas is partial recursive, then the numbers $\langle e, x\rangle$ or τ may not necessarily belong to the domain of g (and hence not either to the domain of ψ_{U_0} or ψ_U). But if g is recursive, it is total, i.e. defined for all arguments. Hence we can combine the last two lemmas in the following more restrictive form.

Lemma 6. With each universal Turing machine U and each recursive function $g(x)$, there exists an argument τ of the domain of g such that $\psi_U(\tau) = g(\tau)$.

Finally we need the following lemma.

Lemma 7. Let U be an arbitrary universal Turing machine, and x an arbitrarily given tape expression for U. Then there exists a computable number ξ such that $s(\xi, U)$ is larger than x.

This lemma follows from the fact that the computable numbers are denumerably infinite. For any U there corresponds with each integer ξ (which is indeed computable) a unique $x = s(\xi, U)$, and no two distinct ξ-numbers can have the same x-number with reference to U. Hence, for any given x there must be an unlimited number of ξ-integers such that $s(\xi, U)$ is larger than x.

We are now in a position to prove theorems about the function $s(\xi, U)$.

Theorem 1. For no universal Turing machine U, $s(\xi, U)$ is a recursive function of ξ. (For each U there exists a ξ whose shortest U-form is not effectively knowable.)

The proof will be a reductio ad absurdum on the following assumption, whose negation is equivalent with Theorem 1. Assumption (to be proved false): there is a universal Turing machine U for which $s(\xi, U)$ is a recursive function of ξ. Let us for the proof consider the function:

$$g(x) = \min_\xi [s(\xi, U) > x].$$

The form of $g(x)$, together with Lemma 7 and the assumption made, immediately reveals that $g(x)$ is recursive. Hence, by Lemma 6, there exists an initial tape expression τ such that $\psi_U(\tau) = g(\tau)$. Thus U, starting from τ, produces a number $\xi_0 = \min_\xi [s(\xi, U) > \tau]$ whose shortest U-form is larger than τ. This contradiction completes the proof of Theorem 1.

Theorem 2. For each universal Turing machine U, the set of shortest U-forms of the computable numbers is not recursively enumerable.

The proof will be a reductio ad absurdum on the following assumption, whose negation is equivalent with Theorem 2. Assumption (to be proved false): for some U there is a recursive enumerating function $e(n, U)$ that produces all shortest U-forms x_i in some order x_{i_n} without repetition $[e(0, U) = x_{i_0}$, $e(1, U) = x_{i_1}, \ldots]$. Let us for the proof consider the function:

$$g^*(x) = \psi_U\{e[\min_n (e(n, U) > x), U]\}.$$

The form of $g^*(x)$, together with the assumption made, reveals first of all that the argument function $e[\min_n (e(n, U) > x), U]$ is recursive, and that it produces an x_i-number which is larger than x. Since each x_i belongs to the domain of the partially recursive function ψ_U, it follows that $g^*(x)$ is recursive. Hence, by Lemma 6, there exists an initial tape expression τ such that U, starting from τ, produces a number $g^*(\tau)$ whose shortest U-form is larger than τ. This contradiction completes the proof of Theorem 2.

Theorem 3. For each universal Turing machine U, the set of its shortest U-forms is not recursive.

The proof is an easy corollary of Theorem 2, for a set is recursive if and only if both it and its complement are recursively enumerable; compare Davis [2].

Actually, Theorem 1 also follows from Theorem 2, for if $s(\xi, U)$ were recursive it would be a recursive enumerating function for the set of shortest U-forms, contradicting Theorem 2.

Theorem 4. There is no algorithm by which we can generate a computable number, whose shortest U-form is larger than an arbitrarily given tape expression x.

Let us for the proof assume that there is an algorithm, i.e. a recursive function $f(x)$, such that $s(f(x), U) > x$ for all x. By Lemma 6 there is a τ such that $\psi_U(\tau) = f(\tau)$. This means that U, when starting from τ, produces a number $f(\tau)$ whose shortest U-form is larger than τ. This contradiction completes the proof of Theorem 4.

Theorem 5. There is no algorithm by which we can generate a computable number, whose shortest U-form is larger than the shortest U-form of an arbitrarily given computable number ξ.

Let us for the proof assume that there is an algorithm, i.e. a recursive function $u(\xi)$ such that $s(u(\xi), U) > s(\xi, U)$. Then the xth composition $u^{(x)}(\xi)$ is recursive, and obviously there exists a ξ_0 such that the shortest U-form of $u^{(x)}(\xi_0)$ is larger than x. Hence we have generated an algorithm $f(x)$ that contradicts Theorem 4. This completes the proof of Theorem 5.

III. Limitations for the Recognition of U-Order

In the definition of U-order there is an arbitrariness, namely the lexicographical order of the description alphabet α. In order to improve on this point, we can naturally define order-classes $O_v(U)$ such that a pattern ξ belongs to $O_v(U)$ if and only if the length of $s(\xi, U)$ is equal to v. Concerning these order classes we have the following theorem.

Theorem 6. There is no algorithm which, for an arbitrarily given v, generates the order class $O_v(U)$. Nor is there an algorithm that generates any element of $O_v(U)$.

Let us for the proof assume that there is an algorithm that generates an element of $O_v(U)$, i.e., a recursive function $v(v)$ such that $s(v(v), U)$ has the length v. This means that we can select an arbitrary x, say of length v, and generate a $\xi = v(v + 1)$ with a shortest U-form that is larger than x. This contradiction against Theorem 4 completes the proof of Theorem 6.

IV. Limitations for the Generation of Random Patterns

In Section I we motivated the hypothesis that our intuitive notion of randomness be identified with the lack of U-order with respect to some universal Turing machine U. We shall see that this hypothesis implies that random patterns cannot in general be generated effectively, i.e., algorithmically.

Let $\xi(v)$ stand for a pattern of length v. For any fixed v, there is among all $\xi(v)$ one unique pattern $\xi_{rand}(v)$, the random pattern of length v, such that $s(\xi_{rand}(v), U)$ is longer than the shortest U-form of any other $\xi(v)$. Concerning these random patterns we have the following theorem.

Theorem 7. For no universal Turing machine U, there exists an algorithm which, for an arbitrarily given v, generates a random pattern $\xi_{rand}(v)$ with respect to U.

Let us for the proof consider an arbitrary universal Turing machine U with a tape alphabet of, say, n symbols. Then there are n^v distinct $\xi(v)$-patterns.* Let x be an arbitrarily given U-form and select a v such that $n^v > x + 1$, i.e., such that the number of distinct $\xi(v)$-patterns is larger than $x + 1$. What can now be said about $\xi_{rand}(v)$? By definition $s(\xi_{rand}(v), U)$ is larger than the shortest U-form of each of the $n^v - 1(> x)$ other $\xi(v)$-patterns, i.e., $s(\xi_{rand}(v), U) > x$. Hence, if there were an algorithm for the generation of $\xi_{rand}(v)$, there would be an algorithm producing a computable number, whose shortest U-form is larger than arbitrarily given x. This contradiction against Theorem 4 completes the proof of Theorem 7.

Obviously Theorem 7 remains valid even if $\xi_{rand}(v)$ is more generously defined as a pattern of length v, such that its shortest U-form is larger than the shortest U-form of $f(v)$ other $\xi(v)$-patterns, where $f(v)$ $(\leq n^v - 1)$ is a recursive, increasing, function of v.

Again it is obvious how to modify the proof of Theorem 7 to obtain the following theorem, valid for both definitions of $\xi_{rand}(v)$.

Theorem 8. Let v be an arbitrary integer and $\mu(v)$ any particular integer such that $\mu(v) > v$. Then, for no universal Turing machine U and for no μ-function, there exists an algorithm for the generation of a random pattern $\xi_{rand}(\mu(v))$ with respect to U.

Theorem 8 is intuitively easy to understand. As we have argued in Section I, it is the regularities of very large patterns (sequences) that enable us to give manageably short effective descriptions of them. Since a random pattern by definition lacks regularities, it is most reasonable that there cannot be a communicable prescription for how to generate arbitrary large random patterns.

It should be noticed that all previous conclusions are valid for all universal Turing machines U and hence also for any particular choice of U, which may be considered best with respect to some particular context.

In an early paper by Church [1], another definition of randomness is suggested, which is also related to von Mises' principle of the excluded gambling system. It invokes the notion of effective computability, but in a way *that is different* from ours. An infinite sequence (pattern) is accepted as random if it passes a certain nonrecursively enumerable set of effective tests. Church concludes that sequences satisfying his definition of randomness cannot be

* Sometimes the ξ-pattern of a Turing machine is produced as a string of only 1's. In this case, of course, we would here identify such a string with the number of 1's that it contains, counted to the base n.

effectively constructed, because such a construction would require the set of tests to be recursively enumerable.*

V. Order and Evolution

Let A be a physical object (pattern) and B a surrounding to A. Let us assume that A has some order and that B has some order such that there is a description $d(A, B)$ of A with respect to B. Let us further assume that this description is sufficiently short to be contained in A, even if A has the further property of providing B with a copy of $d(A, B)$. Then we shall say that A is self-reproductive with respect to B. For if we start with B alone and thereafter provide B with an A, then A will cause B to construct another A and so forth.

It is not difficult to give examples of surroundings B with very small order (regularities; far below the universality of a U-machine), which still will perform as described above, i.e. which are capable of allowing some A to be self-reproductive with respect to B.

Notice, however, that if B is an autonomous automaton that produces A-patterns, then we cannot say that A is self-reproductive with respect to B, for here B alone will produce an A. If A really were self-reproductive with respect to B, one A has to be present from the beginning and be the cause of B constructing another A.

It is immediately clear that if B is a surrounding with no regularities whatsoever, then no A can be explained† to be self-reproductive with respect to B.

If, however, both A and B possess sufficiently large order such that A is self-reproductive with respect to B, then it can happen that A changes the surrounding. The surrounding of a particular A may after a while be more ordered than it was before (due to the appearance of new A's). Order may thus feed new order. It may happen that the surrounding B will be changed to B^*, permitting another pattern A^* to be self-reproductive with respect to B^*. If at least one A^* were present in the original B, it can thus happen that we reach a situation with two self-reproductive species (A and A^*). "Natural selection" may occur and an evolutionary process may start. Eventually it will lead to self-reproducing automata like ourselves, who find it impossible to figure out the order of some patterns. This does not at all contradict the popular form of the natural-selection principle, which says that "nature does not plan its own development."

It should be noticed that an order-feeding-order process requires energy. For one thing a surrounding B has to extract information from the description $d(A, B)$. And every extraction of information requires energy.

The properties that permit a surrounding to serve as a breeding place can, as we have outlined, be produced by other self-reproductive processes in the surrounding. Thus it is natural to consider an hierarchy of self-reproductive processes like the one we find in our own biological bodies. This hierarchy has a natural beginning with unicellular organisms. But we also have symbiosis, where both parts take advantage of each other.

It is most interesting also to notice the "natural-selection" and evolution-like character of thought processes. A theory, for example, can develop as a consequence of a competition between previously existing theories. If these have an overlapping range, where they predict different results, it can happen that one of them will become extinct; for example be forgotten. Coexistence inside one brain can occur. But in the brain of a professional scientist it frequently happens that a coexistence of conflicting theories leads to an open conflict, that results in a new theory. Again two theories may compete for elegance (shortness), even if they do not predict different results.

The discussion may point on a natural extension of cybernetics, from communication and control in animal and machine, towards a study of invariances of evolution characteristics over animal and machine (or over animal formation and theory formation). The control concept requires externally defined norms to control after, whereas such norms are automatically generated in a "natural-selection" process.

* With the motivation that all sequential tests, of use in statistical practice, are of the recursively enumerable type or even of a much simpler type, a recursively enumerable set of tests for "randomness" (thus a conception that is different both from Church's and from ours) is suggested in a very recent paper by P. Martin-Löf [The definition of random sequences. Research Report from the Institution of Mathematical Statistics of the University of Stockholm, Sweden, 1966]. In this report the author also refers to two independently suggested complexity measures, somewhat similar to our shortest U-form, $s(\xi, U)$. One is due to A. Kolmogorov [Three approaches to the definition of the concept of "quantity of information" (in Russian). *Problemy Peredači Informacii* 1, 3–11 (1965)]. The other is due to R. Solomonoff [A formal theory of inductive inference. Part I. *Information and Control* 7, 1–22 (1964)]. Here, however, the interests are focused upon other problems than the question if the measures suggested are effective in any sense.

† In this case, however, it is possible to give a set-theoretic interpretation of self-reproduction that is consistent with for example the von Neumann-Bernays-Gödel set theory. In this model self-reproduction implies the negation of the axiom of Fundierung and is implied by for example the Quine Theory of Individuals. The conclusions are, as will be shown in a forthcoming paper, that the existence of an automaton A, that is self-reproducing in an unordered surrounding B, can be axiomatized as well as the nonexistence of such an automaton.

REFERENCES

[1] Church, A., On the concept of a random sequence. *Bull. Amer. Math. Soc.* **46**, 130–135 (1940).
[2] Davis, M., "Computability and Unsolvability." McGraw-Hill, New York, 1958.
[3] Davis, M., A note on universal Turing machines, *in* "Automata Studies" (C. E. Shannon and J. McCarthy, eds.). Princeton Univ. Press, Princeton, New Jersey, 1956.
[4] Popper, K., "The Logic of Scientific Discovery." Basic Books, New York, 1961.

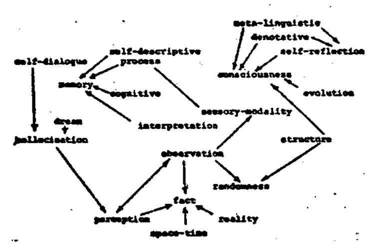
Some of the work on this paper was done at the University of Illinois under AFOSR Grant 7-67.

RANDOMNESS

We have no chance for generating or noticing random events or processes.

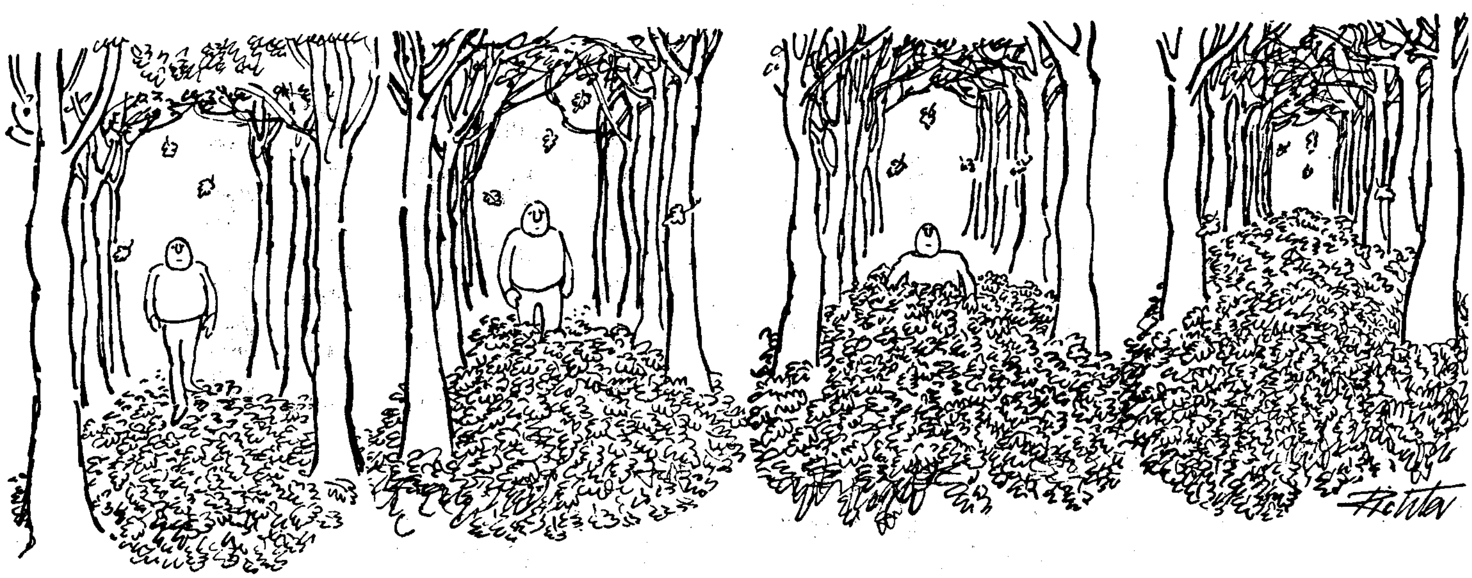
One of the principles which 'works' against randomness is the processing mechanism of our nervous system. At its periphery the nervous system extracts non-random features from its own excitatory processes and transforms these inputs to coded interpretations or outputs.

The non-random outcome of the matching of output and input, or the perceptual transformation according to unique numerical base systems is represented in Stephens' 'power law', says Baird, who also notes that subjects when generating clearly 'prefer' certain frequency categories over others. Apparently the subjects perform a 'spacing along an internal response scale.



What we call Laws of Nature, are the functional limitations imposed by our own structure, i.e. the laws of our own nature.

[R.F.]



IF YOU SMILE AT ME I WILL UNDERSTAND.... Glenn Kowack

Cybernetician #5 generated a great deal of discussion when a response to an assignment was taken out of context. During a discussion of context, Bob Rebitzer suggested that a way may be found to design a composition so that: 1. No part could be taken out of context, or 2. That any part taken out of context would have no meaning other than the one intended. I have formed the following observations:

Typically, observations are made in preference for least organization and least energy. That is, in a given situation, observations made first are those that require the least thought, time, information or effort. An individual who walks into a room will first perceive a bright flashing light before he sees that that light is slowly changing frequency. This is only a generalized analogy. "Least"organization and energy are dependent

HZXTNY
GRISPX
ZTNYE
TENOIA
BRSBY

upon the particular observer. The act of taking something out of context can be the result of this principle. In designing a format so that no parts are taken out of context, it is therefore necessary to create that composition in such a way that it would require less energy and organization to be observed in context than not. Also, it would be advantageous to in-context observation if the composition was arranged so that if the particular organization necessary for the understanding were not used the parts viewed would be meaningless.

I draw the distinction between two types of compositions. The first is form without specific content, and that content is provided by the observer. I Ching, some types of poetry, music, abstract art, can fit under this heading. The second provides both content and form, such as a technical paper, newpaper articles. The key characteristic of this composition being that it is necessarily made of commonly accepted symbols, words, or units of meaning. I consider these two composition types to be on the opposite ends of a continuum, and, that either one cannot be totally devoid of the other (all words, etc. have some subjectivity, and I doubt if completely content-less forms can exist).

In designing a HZXTNY[1] composition, I am concerned only with those compositions that use standard units of meaning (such as cybernetic articles). These units are, essentially, a code. Since this code is known by everyone (English language), they may pull any part out of context. It is therefore necessary that a composition generate its own unique code. By inserting this code in proper intervals the GRISPX[2] of context could be solved. This is a very specific ZTNYE[3] for only one type of composition, which I hope to TENOIA[4] into a more comprehensive BRSBY[5]. The next step, appears to be incorporation of new coding into the body and process of the composition, rather than an addendum.

[1]context-only

[2]problem

[3]technique

[4]develop

[5]system

Packaged Entropy: Testing the A-Bomb, near Alamogordo, New Mexico, May 1945.

ENTROPY

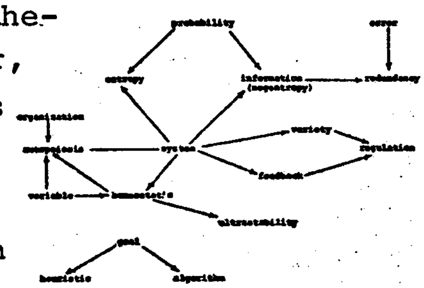
A thermodynamic quantity (S) correlated with the negative of the quantity H for the ensemble used to represent the system of interest. The constant k is needed to allow for choice of units. $S = -kH$. The ensemble which is chosen to represent the system of interest is determined by the observations we have made on the condition of that system. Thus the value of H will be determined by the nature of our knowledge of the condition of the system. Since the value of H is lower the less exactly the state of the system is specified, the entropy of a system is a measure of the degree of our ignorance as to its condition. Thermodynamic entropy is defined by its relation to heat (Q) and temperature (T) and is statistically interpreted with the help of H.

$$\Delta S \geqslant \int \frac{\delta Q}{T}$$

[S.U.]

ENTROPY

In cybernetics, entropy is generalized to measure the tendency of any closed system to move from a less to a more probable state, using the same mathematical apparatus as in physics. If however, the system is open to information, then this tendency may be arrested. This is because, mathematically speaking, information can be defined precisely as negative entropy (often called negentropy). [S.B.]



NEGENTROPY

(See S.B.: "Entropy".)

H

ENTROPY

INFORMATION THEORY PARTITIONS, COALITIONS TABLES

Bob Feldman, Bruce Garret, Jeff Glassman, Glenn Kowack, Debby Langerman, Leslie Murphy, Stephen Sloan, Heinz Von Foerster

INFORMATION THEORY

(AN ELEMENTARY INTRODUCTION)

The most fundamental step in a mathematical theory of information is the development of a measure for the amount of uncertainty of a situation as a whole, or—as I shall put it—for the uncertainty of a "well-defined universe." The definition of this universe under discourse can be done on several levels. The first step in its definition is to associate with this universe a finite number of distinguishable states which are, also, all the states that the universe can assume. To use some worn-out but illustrative examples, a die with its six faces, or a coin with its two sides, may be considered as such a universe. The face or the side that comes up after the die or the coin is tossed, represents a distinguishable state in these respective "universes." Due to the distinguishability of the individual states, it is possible to label them, say, S_1; S_2; S_3; etc. In general, we may call a state S_i, where i goes through all integers from 1 to n, if our universe is defined by precisely n states. Thus, for the coin:

$$n = 2; \qquad S_1 = \text{heads}, \quad S_2 = \text{tails},$$

and, similary, for the die, $n = 6$, with the names of states S_1 corresponding—for simplicity—to the eyes, i, shown by the die.

As long as we do not deal with a completely deterministic universe, that is, a universe in which for each state there exists one, and only one possible successor state (and our previous examples are certainly not of this type), we intuitively associate with such indeterministic universes a certain amount of uncertainty. For instance, in the coin situation we may say that we are unable to predict the outcome of a particular toss, but we are much less uncertain with respect to the situation as a whole if we compare it with the die situation with its superior variety of possible outcomes. The question of <u>how much</u> uncertainty can be associated with these various situations leads to the second step in the defintion of our universes. Since, clearly, probability considerations will determine these uncertainties, I propose to associate with each state S_i in our universe the probability p_i of its occurrence. Since our universe consists of precisely n states, and hence must be in one of these states at any instant of time, we have, of course, certainty that it is at a given instant of time in any one of its states:

$$p_1 + p_2 + \ldots\ldots + p_n = \sum_1^n p_i = 1. \qquad (1)$$

In the simple situation of a universe in which all probabilities p_i are alike, say, p_u -- as it is the case for an "honest" coin or an "honest" die -- the equation above is simply

$$np_i = 1,$$

and the probability for an individual state S_i is just the inverse of the number of states:

$$p_u = p_i = \frac{1}{n},$$

or, for our two examples:

$$p_{coin} = \frac{1}{2}, \text{ and } p_{die} = \frac{1}{6}.$$

If we wish now to associate with each universe a measure of uncertainty, it appears, at least, to be plausible that this measure has to take into consideration the probabilities, or the uncertainties, if you wish, of all states that define this universe. In other words, the measure of uncertainty—usually denoted by H—of a particular universe should be a function of all p_i:

$$H = H(p_1;\ p_2;\ p_3;\ \dots\ p_i\ \dots\ p_n). \qquad (2)$$

Since there are infinitely many functions from which we may choose one, as, for example,

$$\sum_1^n p_i^2\ ;\ \sum_1^n e^{p_i}\ ;\ \prod_1^n \log p_i\ ;\ \text{etc., etc., } \dots.$$

we are in a position to introduce certain conditions which we intuitively like to see fulfilled in our final choice for a measure of the uncertainty of a universe. It cannot be stressed strongly enough that the choice of these conditions is more or less arbitrary, their justifications being solely confined to their implications.

One of these conditions may reasonably be that the measure of these uncertainties should, in a sense, reflect our intuition of the amount of these uncertainties. In other words, more uncertainty should be represented by a higher measure of uncertainty.

Another condition may be that the measure of uncertainty should vanish (H = 0) for a deterministic universe, that is, for a universe in which there are no uncertainties.

Finally, we may propose that the measure of uncertainty for two independent universes, U_1 and U_2, should be the sum of the measures of uncertainty of each universe separately. In mathematical language, this is

$$H(U_1 \ \&\ U_2) = H(U_1) + H(U_2)\ . \qquad (3)$$

From this last condition we may get a hint as to the form of our measure function. Consider for a moment our two examples, the coin and the die. We wish that the measure that expresses the uncertainity of a universe composed of a coin and a die equals the sum of the measures that express the uncertainties of the coin-universe and of the die-universe. Since, in the combined universe, the number of states n_{CD} is the product of the number of states of the component universes n_C, n_D:

$$n_{CD} = n_C \cdot n_D = 2 \cdot 6 = 12;$$

and since all probabilities are equal:

$$p_{CD} = \frac{1}{n_{CD}} = \frac{1}{n_C \cdot n_D} = p_C \cdot p_D\ ,$$

we have with a postulated addition theorem of independent universes:

$$H(p_{CD}) = H(p_C \cdot p_D) = H(p_C) + H(p_D)\ .$$

This relation states that the desired measure function, H, taken of the product of two factors, should equal to the sum of the measure function of the factors There is, essentially, only one mathematical function that fulfills this condition, namely, the logarithmic function. Consequently, we put tentatively:

$$H(p) = k \cdot \log(p)\ ,$$

with k being an as yet undetermined constant. We verify the relation above in $H(p_{CD})$:

$$H(p_{CD}) = k \log(p_{CD}) = k \log(p_C \cdot p_D)$$
$$= k \log(p_C) + k \log(p_D)$$
$$= H(p_C) + H(p_D)\ .$$

Q. E. D.

Since, on the other hand, we intuitively feel that a universe that can assume more states than another is also associated with a higher degree of uncertainty, we are forced to give our as yet undetermined constant k a negative sign:

$$H(p) = -k \cdot \log(p) = -k \cdot \log\left(\frac{1}{n}\right) = k \cdot \log(n),$$

In the simple situation of universes whose states are equiprobable, we have come to the conclusion that an adequate function that can be used as a measure of uncertainty is of logarithmic form:

$$H(p) = k \cdot \log(n) = -k \cdot \log(p). \qquad (4)$$

However, situations in which all states are equiprobable are relatively rare, and we have to consider the general case in which each state S_i is associated with a probability p_i which may or may not be equal to the probability of other states. Since, in the equiprobable situation the uncertainty measure turned out to be proportional to the log of the reciprocal of the probability of a single state, it is suggestive to assume that, in the general case, the uncertainty measure is now proportional to the mean value of the log of the reciprocal of each individual state:

$$H = -k \cdot \overline{\log p_i}, \qquad (5)$$

the bar indicating the mean value of log (p_i) for all states n.

The calculation of a mean value is simple. Take N sticks consisting of n groups, each of which contains N_i sticks of length ℓ_i. What is the mean value of their length? Clearly, the total length of all sticks, divided by the number of sticks:

$$\bar{\ell} = \frac{\sum_1^n N_i \ell_i}{N}$$

Call p_i the probability of the occurrence of a stick with length l_i:

$$p_i = N_i \;/\; N,$$

and the expression for the mean length becomes

$$\bar{\ell} = \sum_1^n \frac{N_i}{N} \cdot \ell_i = \sum_1^n p_i \cdot \ell_i \, .$$

In other words, the mean value of a set of values is simply obtained by the sum of the products of the various values with the probability of their occurrence. Consequently, the mean value of the various values of log p_i is simply:

$$\overline{\log p_i} = \sum_1^n p_i \log p_i \, ,$$

and the uncertainty measure becomes

$$H = -k \, \overline{\log p_i} = -k \sum_1^n p_i \log p_i$$

This measure function fulfills all that we required it to fulfill. It reduces to the simple equation for equiprobable states, because with

$$p_i = \frac{1}{n} \, ,$$

$$H = -k \sum_1^n \frac{1}{n} \log \frac{1}{n} = -k \, \frac{n}{n} \log \frac{1}{n} = k \log n;$$

consequently, H increases in a monotonic fashion with increasing number of equiprobable states. Furthermore, H vanishes for certainty, which we shall express by assuming only one state to appear with certainty, say $p_1 = 1$, while all others have probability 0. Consequently,

$$H = -k \left[1 \, . \, \log 1 + \sum_2^n 0 \, . \, \log 0 \right] = 0 \, ,$$

because

$$1 \log 1 = 0, \text{ and also}$$

$$0 \log 0 = 0.$$

Since the latter expression is not obvious, because $\log 0 = -\infty$, we quickly show with l'Hospital's rule that

$$\lim_{x \to 0} (x \, . \, \log x) = 0 :$$

$$\lim_{x \to 0} (x \, . \, \log x) = \lim \frac{\log x}{\left(\frac{1}{x}\right)} = \lim \frac{\frac{d \log x}{dx}}{\frac{d}{dx}\left(\frac{1}{x}\right)}$$

$$= \lim \frac{1/x}{-1/x^2} = \lim_{x \to 0} (-x) = 0. \qquad \text{Q. E. D.}$$

Finally, we show that the addition theorem for independent universes is preserved, also. Let p_i and p_j be the probabilities of states of two universes, U_1 and U_2, respectively. The probabilities of the combined universe are

$$p_{ij} = p_i \cdot p_j \,.$$

Let n and m be the number of states corresponding to U_1 and U_2. For U_1 and U_2 the number of states is $n \cdot m$. The uncertainty of the combined universe is

$$H(U_1 \,\&\, U_2) = -k \sum_{ij} p_{ij} \log(p_{ij})$$

$$= -k \sum_{ij} p_i \cdot p_j \log(p_i \cdot p_j)$$

$$= -k \sum_{ij} p_i \cdot p_j \log(p_i) - k \sum_{ij} p_i \cdot p_j \log(p_j)$$

$$= -k \sum_j p_j \sum_i p_i \log(p_i) - k \sum_i p_i \sum_j p_j \log(p_j)$$

Since $\sum_1^n p_i = \sum_1^m p_j = 1$, the above expression becomes

$$H(U_1 \,\&\, U_2) = -k \sum_1^n p_i \log(p_i) - k \sum_1^m p_j \log(p_j)$$

$$= H(U_1) + H(U_2) \,.$$

Q. E. D.

The only point that remains to be settled in our uncertainty measure H, is an appropriate choice of the as yet undetermined constant k. Again, we are free to let our imagination reign in adjusting this constant. The proposition now generally accepted is to adjust this constant so that a "single unit of uncertainty," usually called "one bit," is associated with a universe that consists of an honest coin. A pedestrian interpretation of this choice may be put forth by suggesting that a universe with just two equiprobable states is, indeed, a good standard with some elementary properties of uncertainty. A more sophisticated argument in favor of this choice is connected with problems of optimal coding (11). However, in this framework, I have no justification for elaborating on this issue. Let us, therefore, accept the previous suggestion and give H the value of unity for the measure of uncertainty that is associated with a universe consisting of an honest coin:

$$H_{coin} = 1 = -k \left[\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right]$$

$$= k \log 2 \,.$$

Hence,

$$k = \frac{1}{\log 2} \,,$$

and

$$H = -\frac{1}{\log 2} \sum_1^n p_i \log p_1 \,,$$

or, if we take 2 as the basis of our logarithmic scale:

$$H = -\sum_1^n p_i \log_2 p_i \,. \qquad (6)$$

With this expression, we have arrived at the desired measure of the uncertainty of a universe that is defined by n states S_i, which occur with probability p_i. It may be worthwhile making a few comments to illustrate some properties of this measure function. First, I would like you to appreciate that for a universe with a fixed number of states, n, the uncertainty measure H is maximum, if all states occur with equal probability. A shift away from this uniform probability distribution immediately reduces the amount of H; in other words, reduces the uncertainty of the universe. Let me illustrate this with a die that is born "honest" but "corrupts" as a consequence of its interaction with bad society. My victim is a die made of a hard cubical shell filled with a highly viscous glue, in the center of which is placed a heavy steel ball. Since there is perfect symmetry in this arrangement, when tossed, the die will go with equal probabilities into its six possible states. However, I am going to teach this fellow to show a preference for the side with one eye. To this end, I place under the table an electromagnet and, whenever one eye comes up, I give the die a short magnetic shock. This moves the steel ball slightly toward the bottom, and gravitation will enhance the chance of its falling at the same side the next time. Table 16 lists

TABLE 16

TIME	p_1	p_2	p_3	p_4	p_5	p_6	H bits
t_0	1/6	1/6	1/6	1/6	1/6	1/6	2.582 = $\log_2 6$
t_1	1/4	1/6	1/6	1/6	1/6	1/12	2.519
t_2	1/3	1/6	1/6	1/6	1/6	0	2.249
t_3	2/3	1/12	1/12	1/12	1/12	0	1.582
t_4	1	0	0	0	0	0	0.000

the probability distribution for the various states as it may look in succeeding intervals, t_0, t_1, t_2, t_3, t_4, between shock treatments. The right hand margin gives the values of H, the uncertainty of this universe, corresponding to the probability distribution at successive states, H is given in "bits," and is calculated according to Equation (6).

Another feature of the uncertainty measure H is that changing from a universe with n equiprobable states to another universe with twice as many equiprobable states, 2n, the uncertainty increases exactly one bit:

$$H_1 = \log_2 n ,$$

$$H_2 = \log_2 2n = \log_2 2 + \log_2 n = 1 + H_1 .$$

Thus, a universe with 1 million states has an uncertainty of about 20 bits. Add 1 million states, and this new universe has about 21 bits uncertainty.

Up to this point, I have referred to our measure function H always as a measure of uncertainty. However, a variety of terms are in use which all refer to the same quantity H as defined in Equation (6). These terms are "entropy," "choice," and "amount of information."

To call H the entropy of a universe, or of a system, is justified by the fact that this thermodynamical variable, when expressed in terms of the probability distribution of the molecules comprising a thermodynamical system, is defined by an equation almost identical to our Equation (6) for H. The importance usually given to the concept of entropy stems from one of the consequences of the second law of thermodynamics, which postulates that in a closed thermodynamical system the entropy must either remain constant (for thermal equilibrium), or go up, but it can never decrease. This is an expression of the fact that, in "natural systems," the distribution of the probabilities of states tends to uniformity, as exemplified by a bucket of hot water in a cold room. After a while, the thermal energy of the bucket will distribute itself more or less uniformly over the room (equilibrium; all p_i alike; H is maximum). Consequently, a thermodynamicist, unaware of my magnetic contraption, who watches my die violating the second law of thermodynamics as it slowly, but surely, moves from high values of entropy to smaller and smaller ones, will come to the conclusion that a Maxwellian demon is at work who alters selectively the internal organization of the system. And he is so right! I, of course, am the demon who selectively switches on the magnet whenever the die shows one eye.

Sometimes, H, as defined in Equation (6), is referred to as the amount of choice one has in a universe consisting of n items, in which one is permitted to pick items with a probability p_i associated with item S_i. All the considerations of intuitive nature which helped us to define a measure function for uncertainty—in particular, the addition theorem—can as well be applied to a measure function of choice. Consequently, the resulting function, expressing a measure of choice, is identical with the function expressing a measure of uncertainty—even with regard to the units, if a unit of choice is associated with a well-balanced temptation between either one of two choices, as is illustrated so beautifully by Burdidan's Ass.

Finally, H is also associated with an "amount of information" in a situation in which the actual state of a universe, whose uncertainty is H, is transmitted by an observer to a recipient. Before the recipient is in possession of the knowledge of the actual state of the universe, his uncertainty regarding this universe is H. The question arises as to how much he values the information about the actual state of the universe when transmitted to him by the observer. All the considerations of intuitive nature that helped us to define a measure of uncertainty are applicable here, too, and the resulting function expressing a measure of information is identical with H as defined in Equation (6). Since, in a communication situation, the "states" in question are usually symbols, H is usually referred to in bits per symbol. If the observer transmits symbols at a constant rate, H may also be expressed in bits per second.

If, in our vocabulary—presumably consisting of about 8,000 words—we were to use each word with equal probability, our linguistic universe would have an uncertainty H of 13 bits ($H = \log_2 8192 = 13$). However, due to our using various words with different frequencies, the uncertainty of our linguistic universe is somewhat smaller and has been measured to be about 11 bits (12). Consequently, whenever I utter a word, I transmit to you, on the average, 11 bits of information. Since I utter approximately three words/sec, my rate of generating information is about 33 bits/sec. (I hope nobody will tell me that what I generate is not information but noise.)

I hope that in this short outline I have been able to show that one and the same expression, namely, $-\sum p_i \log_2 p_i$, represents a measure for various concepts in various situations. This is reflected in the various names for this expression as, for instance, uncertainty, entropy, choice, and information. This state of affairs may well be compared to mechanics, where the product of a force and a length represents "work" in one context, but "torque" in another context.

In many cases the probabilities p_i are not given a priori, but are calculated by observing the frequencies N_i of events S_i to occur in a sample of $N = \Sigma N_i$ tests. In such cases, it is convenient to work directly with the observed integer quantities N_i, rather than with the derived quantities $p_i = N_i/N$. We have

$$H = -\sum_1^n p_i \log_2 p_i = -\sum_1^n \frac{N_i}{N}\log_2\frac{N_i}{N}$$

and after some obvious re-arrangements

$$H = \frac{1}{N}[N\log_2 N - \sum_1^n N_i \log_2 N_2]$$

The following Table, taken from H. Tuttle, W.R. Ashby, and K. Kokjer: Entropies of Partitions, BCL Report No. 1.1, Biological Computer Laboratory, University of Illinois, Urbana, Illinois, 61801 gives for integer x the values of expressions:

$x \log_2 x$

$x \log_2 x$

In the calculations of information theory the quantity $x\log_2 x$ appears quite frequently. At times the x involved is a probability, and there are published tables of $p\log_2 p$ (see Quastler, Henry ed. Information Theory in Psychology, Free Press, 1955). At other times, however, the x is an integer and there appear to be no readily available tables of $x\log_2 x$. There are tables of xlnx but they are not too handy since one must apply a multiplying factor (see Kullback, S., Information Theory and Statistics, Wiley, 1959). This table of $x\log_2 x$ was compiled to fill the gap of usable tables.

The table was compiled on the University of Illinois Illiac II computer using the simple formula $x\log_2 x = x\ln x/\ln 2$. Illiac II carries 12 digits in its field and rounds into the last place required. Here we required 9 places, so the tabled value is rounded in the fourth decimal position. The program was run on both Illiac II and the Illinois 7094 and the final table checked against all other runs to try to insure accuracy. The computer printout was copied photographically for the reproduction process so there will be no clerical errors involved.

This table gives values of $x\log_2 x$ for x ranging form 0 to 1009 in increments of one. The table is laid out in the usual manner with unity increments proceeding horizontally and decadic increments vertically.

X LOG X, BASE 2

	0	1	2	3	4	5	6	7	8	9	
0	.0000	-.0000	2.0000	4.7549	8.0000	11.6096	15.5098	19.6515	24.0000	28.5293	0
10	33.2193	38.0537	43.0196	48.1057	53.3030	58.6034	64.0000	69.4869	75.0587	80.7106	10
20	86.4386	92.2387	98.1075	104.0419	110.0391	116.0964	122.2114	128.3820	134.6059	140.8814	20
30	147.2067	153.5801	160.0000	166.4650	172.9737	179.5249	186.1173	192.7498	199.4212	206.1307	30
40	212.8771	219.6596	226.4773	233.3294	240.2150	247.1334	254.0838	261.0657	268.0782	275.1208	40
50	282.1928	289.2937	296.4229	303.5798	310.7639	317.9748	325.2119	332.4747	339.7629	347.0759	50
60	354.4134	361.7750	369.1602	376.5686	384.0000	391.4539	398.9300	406.4280	413.9475	421.4882	60
70	429.0498	436.6320	444.2346	451.8572	459.4995	467.1614	474.8425	482.5426	490.2614	497.9987	70
80	505.7542	513.5279	521.3193	529.1283	536.9547	544.7982	552.6588	560.5361	568.4300	576.3403	80
90	584.2668	592.2093	600.1677	608.1418	616.1314	624.1363	632.1564	640.1915	648.2416	656.3063	90
100	664.3856	672.4794	680.5874	688.7096	696.8457	704.9958	713.1596	721.3370	729.5279	737.7321	100
110	745.9496	754.1802	762.4238	770.6802	778.9495	787.2314	795.5258	803.8327	812.1519	820.4833	110
120	828.8269	837.1825	845.5500	853.9293	862.3203	870.7230	879.1373	887.5630	896.0000	904.4483	120
130	912.9078	921.3784	929.8600	938.3526	946.8560	955.3701	963.8949	972.4304	980.9764	989.5328	130
140	998.0996	1006.6767	1015.2641	1023.8616	1032.4692	1041.0868	1049.7144	1058.3518	1066.9991	1075.6561	140
150	1084.3228	1092.9991	1101.6850	1110.3803	1119.0851	1127.7993	1136.5227	1145.2555	1153.9974	1162.7484	150
160	1171.5085	1180.2776	1189.0557	1197.8427	1206.6385	1215.4432	1224.2565	1233.0786	1241.9093	1250.7486	160
170	1259.5965	1268.4528	1277.3175	1286.1907	1295.0722	1303.9619	1312.8600	1321.7662	1330.6806	1339.6030	170
180	1348.5336	1357.4721	1366.4186	1375.3731	1384.3354	1393.3056	1402.2835	1411.2693	1420.2627	1429.2638	180
190	1438.2726	1447.2889	1456.3128	1465.3442	1474.3831	1483.4294	1492.4831	1501.5442	1510.6126	1519.6883	190
200	1528.7712	1537.8614	1546.9587	1556.0632	1565.1748	1574.2934	1583.4191	1592.5518	1601.6915	1610.8381	200
210	1619.9916	1629.1519	1638.3191	1647.4931	1656.6739	1665.8615	1675.0557	1684.2566	1693.4642	1702.6784	210
220	1711.8991	1721.1265	1730.3603	1739.6007	1748.8475	1758.1008	1767.3604	1776.6265	1785.8989	1795.1777	220
230	1804.4627	1813.7540	1823.0516	1832.3554	1841.6653	1850.9815	1860.3038	1869.6321	1878.9666	1888.3072	230
240	1897.6537	1907.0063	1916.3649	1925.7294	1935.0999	1944.4763	1953.8586	1963.2467	1972.6407	1982.0405	240
250	1991.4461	2000.8574	2010.2745	2019.6974	2029.1259	2038.5601	2048.0000	2057.4455	2066.8966	2076.3533	250
260	2085.8156	2095.2835	2104.7568	2114.2357	2123.7200	2133.2099	2142.7051	2152.2058	2161.7119	2171.2234	260
270	2180.7402	2190.2624	2199.7899	2209.3227	2218.8608	2228.4041	2237.9528	2247.5066	2257.0656	2266.6298	270
280	2276.1992	2285.7738	2295.3535	2304.9383	2314.5282	2324.1232	2333.7232	2343.3283	2352.9384	2362.5535	280
290	2372.1736	2381.7987	2391.4288	2401.0638	2410.7037	2420.3485	2429.9982	2439.6528	2449.3122	2458.9765	290
300	2468.6456	2478.3195	2487.9982	2497.6817	2507.3700	2517.0630	2526.7607	2536.4631	2546.1703	2555.8821	300
310	2565.5986	2575.3197	2585.0455	2594.7759	2604.5109	2614.2505	2623.9947	2721.6854	2731.4790	2741.2771	310
320	2663.0170	2672.7839	2682.5552	2692.3311	2702.1114	2711.8962	2721.6854	2731.4790	2741.2771	2751.0795	320
330	2760.8863	2770.6975	2780.5131	2790.3330	2800.1572	2809.9858	2819.8187	2829.6558	2839.4972	2849.3430	330
340	2859.1929	2869.0471	2878.9056	2888.7682	2898.6351	2908.5061	2918.3814	2928.2608	2938.1443	2948.0320	340
350	2957.9239	2967.8199	2977.7199	2987.6241	2997.5324	3007.4447	3017.3611	3027.2816	3037.2060	3047.1346	350
360	3057.0671	3067.0037	3076.9442	3086.8887	3096.8372	3106.7897	3116.7461	3126.7065	3136.6708	3146.6390	360
370	3156.6111	3166.5872	3176.5671	3186.5509	3196.5385	3206.5300	3216.5254	3226.5246	3236.5276	3246.5345	370
380	3256.5451	3266.5596	3276.5778	3286.5998	3296.6256	3306.6551	3316.6884	3326.7254	3336.7662	3346.8106	380
390	3356.8588	3366.9107	3376.9663	3387.0255	3397.0884	3407.1550	3417.2252	3427.2991	3437.3766	3447.4577	390
400	3457.5425	3467.6308	3477.7228	3487.8183	3497.9174	3508.0201	3518.1264	3528.2362	3538.3495	3548.4664	400
410	3558.5868	3568.7108	3578.8382	3588.9692	3599.1036	3609.2415	3619.3829	3629.5278	3639.6761	3649.8279	410
420	3659.9831	3670.1418	3680.3039	3690.4694	3700.6383	3710.8106	3720.9863	3731.1654	3741.3479	3751.5337	420
430	3761.7229	3771.9155	3782.1114	3792.3107	3802.5132	3812.7191	3822.9284	3833.1409	3843.3567	3853.5759	430
440	3863.7983	3874.0240	3884.2529	3894.4852	3904.7206	3914.9594	3925.2014	3935.4466	3945.6950	3955.9467	440
450	3966.2015	3976.4596	3986.7209	3996.9854	4007.2530	4017.5238	4027.7978	4038.0750	4048.3553	4058.6388	450
460	4068.9254	4079.2152	4089.5081	4099.8041	4110.1032	4120.4054	4130.7107	4141.0192	4151.3307	4161.6453	460
470	4171.9630	4182.2837	4192.6075	4202.9344	4213.2643	4223.5973	4233.9333	4244.2723	4254.6143	4264.9594	470
480	4275.3075	4285.6586	4296.0127	4306.3697	4316.7298	4327.0929	4337.4589	4347.8279	4358.1998	4368.5747	480
490	4378.9526	4389.3334	4399.7171	4410.1038	4420.4934	4430.8859	4441.2814	4451.6797	4462.0810	4472.4851	490
500	4482.8921	4493.3021	4503.7149	4514.1305	4524.5491	4534.9705	4545.3947	4555.8219	4566.2518	4576.6846	500

X LOG X, BASE 2

	0	1	2	3	4	5	6	7	8	9	
510	4587.1203	4597.5587	4608.0000	4618.4441	4628.8910	4639.3407	4649.7933	4660.2486	4670.7067	4681.1676	510
520	4691.6313	4702.0977	4712.5669	4723.0389	4733.5137	4743.9911	4754.4714	4764.9544	4775.4401	4785.9285	520
530	4796.4197	4806.9136	4817.4103	4827.9096	4838.4116	4848.9164	4859.4238	4869.9339	4880.4468	4890.9622	530
540	4901.4804	4912.0013	4922.5248	4933.0510	4943.5798	4954.1113	4964.6454	4975.1822	4985.7216	4996.2636	540
550	5006.8083	5017.3556	5027.9055	5038.4580	5049.0132	5059.5709	5070.1312	5080.6942	5091.2597	5101.8278	550
560	5112.3985	5122.9718	5133.5476	5144.1260	5154.7070	5165.2905	5175.8766	5186.4652	5197.0564	5207.6501	560
570	5218.2463	5228.8451	5239.4464	5250.0502	5260.6566	5271.2654	5281.8768	5292.4907	5303.1070	5313.7259	570
580	5324.3473	5334.9711	5345.5974	5356.2263	5366.8575	5377.4913	5388.1275	5398.7662	5409.4073	5420.0509	580
590	5430.6970	5441.3455	5451.9964	5462.6498	5473.3056	5483.9638	5494.6244	5505.2875	5515.9530	5526.6209	590
600	5537.2912	5547.9639	5558.6390	5569.3166	5579.9965	5590.6788	5601.3634	5612.0505	5622.7399	5633.4317	600
610	5644.1259	5654.8225	5665.5214	5676.2226	5686.9262	5697.6322	5708.3405	5719.0512	5729.7642	5740.4795	610
620	5751.1971	5761.9171	5772.6394	5783.3640	5794.0910	5804.8202	5815.5518	5826.2857	5837.0218	5847.7603	620
630	5858.5011	5869.2441	5879.9894	5890.7370	5901.4869	5912.2391	5922.9936	5933.7503	5944.5092	5955.2705	630
640	5966.0340	5976.7997	5987.5677	5998.3380	6009.1105	6019.8852	6030.6622	6041.4414	6052.2228	6063.0065	640
650	6073.7923	6084.5804	6095.3708	6106.1633	6116.9580	6127.7550	6138.5541	6149.3555	6160.1590	6170.9647	650
660	6181.7727	6192.5828	6203.3951	6214.2095	6225.0262	6235.8450	6246.6660	6257.4891	6268.3145	6279.1419	660
670	6289.9716	6300.8034	6311.6373	6322.4734	6333.3116	6344.1520	6354.9945	6365.8391	6376.6859	6387.5348	670
680	6398.3858	6409.2390	6420.0942	6430.9516	6441.8111	6452.6727	6463.5364	6474.4022	6485.2702	6496.1402	680
690	6507.0123	6517.8865	6528.7627	6539.6411	6550.5215	6561.4041	6572.2887	6583.1753	6594.0641	6604.9549	690
700	6615.8478	6626.7427	6637.6397	6648.5388	6659.4399	6670.3430	6681.2482	6692.1554	6703.0647	6713.9760	700
710	6724.8894	6735.8048	6746.7222	6757.6416	6768.5631	6779.4866	6790.4121	6801.3396	6812.2691	6823.2007	710
720	6834.1342	6845.0698	6856.0073	6866.9469	6877.8884	6888.8320	6899.7775	6910.7250	6921.6745	6932.6260	720
730	6943.5794	6954.5349	6965.4923	6976.4517	6987.4130	6998.3763	7009.3416	7020.3088	7031.2780	7042.2492	730
740	7053.2223	7064.1973	7075.1743	7086.1533	7097.1342	7108.1170	7119.1017	7130.0884	7141.0771	7152.0676	740
750	7163.0601	7174.0545	7185.0508	7196.0491	7207.0492	7218.0513	7229.0553	7240.0612	7251.0690	7262.0787	750
760	7273.0903	7284.1038	7295.1192	7306.1364	7317.1556	7328.1767	7339.1996	7350.2245	7361.2512	7372.2798	760
770	7383.3103	7394.3426	7405.3768	7416.4129	7427.4509	7438.4907	7449.5324	7460.5759	7471.6213	7482.6685	770
780	7493.7176	7504.7686	7515.8214	7526.8760	7537.9325	7548.9908	7560.0510	7571.1130	7582.1768	7593.2425	780
790	7604.3100	7615.3793	7626.4504	7637.5234	7648.5982	7659.6748	7670.7532	7681.8334	7692.9155	7703.9993	790
800	7715.0850	7726.1724	7737.2617	7748.3527	7759.4456	7770.5402	7781.6366	7792.7349	7803.8349	7814.9367	800
810	7826.0403	7837.1456	7848.2528	7859.3617	7870.4724	7881.5848	7892.6991	7903.8151	7914.9329	7926.0524	810
820	7937.1737	7948.2967	7959.4215	7970.5481	7981.6764	7992.8065	8003.9383	8015.0719	8026.2072	8037.3443	820
830	8048.4830	8059.6236	8070.7658	8081.9098	8093.0556	8104.2030	8115.3522	8126.5032	8137.6558	8148.8102	830
840	8159.9662	8171.1240	8182.2835	8193.4448	8204.6077	8215.7724	8226.9387	8238.1068	8249.2765	8260.4480	840
850	8271.6212	8282.7960	8293.9726	8305.1508	8316.3308	8327.5124	8338.6957	8349.8807	8361.0674	8372.2558	850
860	8383.4459	8394.6376	8405.8310	8417.0261	8428.2228	8439.4212	8450.6213	8461.8231	8473.0265	8484.2315	860
870	8495.4383	8506.6467	8517.8567	8529.0684	8540.2818	8551.4968	8562.7135	8573.9318	8585.1517	8596.3733	870
880	8607.5965	8618.8214	8630.0479	8641.2761	8652.5059	8663.7373	8674.9703	8686.2050	8697.4413	8708.6792	880
890	8719.9188	8731.1599	8742.4027	8753.6471	8764.8931	8776.1408	8787.3900	8798.6409	8809.8933	8821.1474	890
900	8832.4031	8843.6603	8854.9192	8866.1797	8877.4418	8888.7055	8899.9707	8911.2376	8922.5060	8933.7761	900
910	8945.0477	8956.3209	8967.5957	8978.8721	8990.1500	9001.4296	9012.7107	9023.9934	9035.2776	9046.5634	910
920	9057.8508	9069.1398	9080.4304	9091.7225	9103.0161	9114.3113	9125.6081	9136.9065	9148.2064	9159.5078	920
930	9170.8108	9182.1154	9193.4215	9204.7291	9216.0383	9227.3491	9238.6614	9249.9752	9261.2906	9272.6075	930
940	9283.9259	9295.2459	9306.5674	9317.8905	9329.2150	9340.5411	9351.8688	9363.1979	9374.5286	9385.8608	940
950	9397.1945	9408.5298	9419.8665	9431.2048	9442.5446	9453.8859	9465.2287	9476.5730	9487.9188	9499.2661	950
960	9510.6150	9521.9653	9533.3171	9544.6705	9556.0253	9567.3817	9578.7395	9590.0988	9601.4596	9612.8219	960
970	9624.1857	9635.5510	9646.9178	9658.2860	9669.6557	9681.0269	9692.3996	9703.7738	9715.1495	9726.5266	970
980	9737.9052	9749.2852	9760.6668	9772.0498	9783.4343	9794.8202	9806.2076	9817.5965	9828.9868	9840.3786	980
990	9851.7719	9863.1666	9874.5627	9885.9604	9897.3594	9908.7600	9920.1619	9931.5653	9942.9702	9954.3765	990
1000	9965.7843	9977.1935	9988.6041	10000.0162	10011.4297	10022.8447	10034.2611	10045.6789	10057.0982	10068.5189	1000

Partitions & Their Entropies

The preceding table provides a considerable short cut in the calculation of entropies

$$H = \frac{1}{N}[N\log_2 N - \sum_1^n N_2 \log_2 N_i]$$

from directly observed frequencies N_i in tests that do not exceed $N = 1009$ observations. Since the cumbersome burden of calculating the values of $N_i \log_2 N_i$ has been taken by the table, it suffices now to look them up and write them down, form the sum of all n terms, subtract this sum from $N\log_2 N$, and divide by N: Presto! The entropy H has been found.

Presto? It still seems--and indeed is--a cumbersome task. Ross Ashby, the practical man, who knew that doing is the best way of learning, and that drudgery is a dispensable companion of doing, suggested--and then proceeded--to eliminate the numerical rigmarol altogether. This gave rise to the following Table of the Entropies of Simple Paritions which was completed by Heath Tuttle*.

The central idea of this table is that the expression

$$N = \sum_1^n N_i$$

is seen as what in number theory is called a "Partition of N into exactly n parts" which, in other words, says that the number N is represented by the sum of the n terms N_i $(i = 1 \to n)$:

$$N = N_1 + N_2 + N_3 + \ldots N_n.$$

For instance:

$$12 = 7+3+1+1$$

is a particular paritions of $N = 12$ into $n = 4$ parts. Of course, there might be tohers, e.g.:

$$12 = 8+2+1+1.$$

In abbreviated form such partitions may be written as (7,3,1,1) or (8,2,1,1) or even more compactly as $7\ 3\ 1^2$ or $8\ 2\ 1^2$ respectively.

*Reproduced here is only a small portion (in fact only the first seven pages) of a 127 page report by Tuttle, H: A Table of the Entropies of Simple Partitions with a Preface by W. Ross Ashby and an Appendix by K. Kokjer, BCL Report No. 1.1, Biological Computer Laboratory, University of Illinois, Urbana, 127 pp. (1970).

If each part (term) in a partition is now interpreted as the frequency N_i of the occurence of a state S_i in a series of N observations, then the entropy of this "universe" can be calculated from the above expression for H.

The following "Table of Entropies for Simple Partitions" is taking the burden of this calculation. Under the heading of N one looks up the particular partition (given "alphabetically" with decreasing numbers) on the left side of the appropriate double column and finds the value of H to the right.

For instance, the values for the changing entropies of a die, growing more and more dishonest with use, as given by an example in the introduction (Table 16, page 188), could have been directly (and more accurately) established from the following table.

Assume 12 tosses had been made during each of the five test-series at t_o, t_1, t_2, t_3 and t_4, and the frequencies of sides showing up were as follows:

N = 12

S_i	⚀	⚁	⚂	⚃	⚄	⚅	
N_i	N_i	N_2	N_3	N_4	N_5	N_6	H
t_o	2	2	2	2	2	2	2.584962
t_1	3	2	2	2	2	1	2.522055
t_2	4	2	2	2	2	0	2.251628
t_3	8	1	1	1	1	0	1.584962
t_4	12	0	0	0	0	0	0.000000

Instead of calculating the probabilities $p_i = N_i/N$, as was done in Table 16, we enter the Table of Entropies at the heading $N = 12$, look up the partition (2,2,2,2,2,2) corresponding to the state of our universe (the die) at t_o and find for $H = 2.584962$ bits; and so on.

Although the abbreviated Table reprinted here goes only up to $N = 16$, due to the identity

$$p_i = N_i/N = kN_i/kN$$

entropies of some partitions for numbers, N, k-times as large as given can be found. For instance, since all terms of the partition

$$60 = 20+16+8+4+4+4+4$$

can be divided by k = 4,

$$H(20,16,8,4,4,4,4) = H(5,4,2,1,1,1,1) = 2.46628$$

under the heading 15 (15 = 60/4). This works also the other way around and may be useful for looking up partitions that are not listed here because of the excessive number of partitions with parts of one element only. For instance the entropy of the partition 8→(2,2,2,1,1), not given under 8, can be found under 16→(4,4,4,2,2) to give H = 2.250000.

TABLES