Jürgen Habermas

Knowledge and Human Interests: A *General Perspective*

I

In 1802, during the summer semester at Jena, Schelling gave his Lectures on the Method of Academic Study. In the language of German Idealism he emphatically renewed the concept of theory that has defined the tradition of great philosophy since its beginnings.

> The fear of speculation, the ostensible rush from the theoretical to the practical, brings about the same shallowness in action that it does in knowledge. It is by studying a strictly theoretical philosophy that we become most immediately acquainted with Ideas, and only Ideas provide action with energy and ethical significance.[1]

The *only* knowledge that can truly orient action is knowledge that frees itself from mere human interests and is based on Ideas —in other words, knowledge that has taken a theoretical attitude.

The word "theory" has religious origins. The *theoros* was the representative sent by Greek cities to public celebrations.[2] Through *theoria*, that is through looking on, he abandoned himself to the sacred events. In philosophical language, *theoria* was transferred to contemplation of the cosmos. In this form, theory already presupposed the demarcation between Being and time that is the foundation of ontology. This separation is first found in the poem of Parmenides and returns in Plato's *Timaeus*. It reserves to *logos* a realm of Being purged of inconstancy and uncertainty and leaves to *doxa* the realm of the mutable and perishable. When the philosopher views the im-

mortal order, he cannot help bringing himself into accord with the proportions of the cosmos and reproducing them internally. He manifests these proportions, which he sees in the motions of nature and the harmonic series of music, within himself; he forms himself through mimesis. Through the soul's likening itself to the ordered motion of the cosmos, theory enters the conduct of life. In *ethos* theory molds life to its form and is reflected in the conduct of those who subject themselves to its discipline.

This concept of theory and of life in theory has defined philosophy since its beginnings. The distinction between theory in this traditional sense and theory in the sense of critique was the object of one of Max Horkheimer's most important studies.[3] Today, a generation later, I should like to reexamine this theme,[4] starting with Husserl's *The Crisis of the European Sciences*, which appeared at about the same time as Horkheimer's.[5] Husserl used as his frame of reference the very concept of theory that Horkheimer was countering with that of critical theory. Husserl was concerned with crisis: not with crises in the sciences, but with their crisis as science. For "in our vital state of need this science has nothing to say to us." Like almost all philosophers before him, Husserl, without second thought, took as the norm of his critique an idea of knowledge that preserves the Platonic connection of pure theory with the conduct of life. What ultimately produces a scientific culture is not the information content of theories but the formation among theorists themselves of a thoughtful and enlightened mode of life. The evolution of the European mind seemed to be aiming at the creation of a scientific culture of this sort. After 1933, however, Husserl saw this historical tendency endangered. He was convinced that the danger was threatening not from without but from within. He attributed the crisis to the circumstance that the most advanced disciplines, especially physics, had degenerated from the status of true theory.

II

Let us consider this thesis. There is a real connection between the positivistic self-understanding of the sciences and traditional ontology. The *empirical-analytic* sciences develop their theories in a self-understanding that automatically generates continuity with the beginnings of philosophical thought. For both are committed to a theoretical attitude that frees those who take it from dogmatic association with the natural interests of life and their irritating influence; and both share the cosmological intention of describing the universe theoretically in its lawlike order, just as it is. In contrast, the *historical-hermeneutic* sciences, which are concerned with the sphere of transitory things and mere opinion, cannot be linked up so smoothly with this tradition—they have nothing to do with cosmology. But they, too, comprise a *scientistic consciousness*, based on the model of science. For even the symbolic meanings of tradition seem capable of being brought together in a cosmos of facts in ideal simultaneity. Much as the cultural sciences may comprehend their facts through understanding and little though they may be concerned with discovering general laws, they nevertheless share with the empirical-analytic sciences the methodological consciousness of describing a structured reality within the horizon of the theoretical attitude. Historicism has become the positivism of the cultural and social sciences.

Positivism has also permeated the self-understanding of the *social sciences*, whether they obey the methodological demands of an empirical-analytic behavioral science or orient themselves to the pattern of normative-analytic sciences, based on presuppositions about maxims of action.[6] In this field of inquiry, which is so close to practice, the concept of value-freedom (or ethical neutrality) has simply reaffirmed the ethos that modern science owes to the beginnings of theoretical thought in Greek philosophy: psychologically an unconditional commitment to theory and epistemologically the severance of knowledge from interest. This is represented in logic by the distinction between descriptive and prescriptive statements, which makes grammatically obligatory the filtering out of merely emotive from cognitive contents.

Yet the very term "value freedom" reminds us that the postulates associated with it no longer correspond to the classical meaning of theory. To dissociate values from facts means counterposing an abstract Ought to pure Being. Values are the nomi-

nalistic by-products of a centuries-long critique of the emphatic concept of Being to which theory was once exclusively oriented. The very term "values," which neo-Kantianism brought into philosophical currency, and in relation to which science is supposed to preserve neutrality, renounces the connection between the two that theory originally intended.

Thus, although the sciences share the concept of theory with the major tradition of philosophy, they destroy its classical claim. They borrow two elements from the philosophical heritage: the methodological meaning of the theoretical attitude and the basic ontological assumption of a structure of the world independent of the knower. On the other hand, however, they have abandoned the connection of *theoria* and *kosmos*, of *mimesis* and *bios theoretikos* that was assumed from Plato through Husserl. What was once supposed to comprise the practical efficacy of theory has now fallen prey to methodological prohibitions. The conception of theory as a process of cultivation of the person has become apocryphal. Today it appears to us that the mimetic conformity of the soul to the proportions of the universe, which seemed accessible to contemplation, had only taken theoretical knowledge into the service of the internalization of norms and thus estranged it from its legitimate task.

III

In fact the sciences had to lose the specific significance for life that Husserl would like to regenerate through the renovation of pure theory. I shall reconstruct his critique in three steps. It is directed in the first place against the objectivism of the sciences, for which the world appears objectively as a universe of facts whose lawlike connection can be grasped descriptively. In truth, however, knowledge of the apparently objective world of facts has its transcendental basis in the prescientific world. The possible objects of scientific analysis are constituted a priori in the self-evidence of our primary life-world. In this layer phenomenology discloses the products of a meaning-generative subjectivity. Second, Husserl would like to show that this productive subjectivity disappears under the cover of an objectivistic self-understanding, because the sciences have not radically freed themselves from interests rooted in the primary life-world. Only phenomenology breaks with the naive attitude in favor of a rigorously contemplative one and definitively frees knowledge from interest. Third, Husserl identifies transcendental self-reflection, to which he accords the name of phenomenological description, with theory in the traditional sense. The philosopher owes the theoretical attitude to a transposition that liberates him from the fabric of empirical interests. In this regard theory is "unpractical." But this does not cut it off from practical life. For, according to the traditional concept, it is precisely the consistent abstinence of theory that produces action-orienting culture. Once the theoretical attitude has been adopted, it is capable in turn of being mediated with the practical attitude:

> This occurs in the form of a novel practice . . . , whose aim is to elevate mankind to all forms of veridical norms through universal scientific reason, to transform it into a fundamentally new humanity, capable of absolute self-responsibility on the basis of absolute theoretical insight.

If we recall the situation of thirty years ago, the prospect of rising barbarism, we can respect this invocation of the therapeutic power of phenomenological description; but it is unfounded. At best, phenomenology grasps transcendental norms in accordance with which consciousness necessarily operates. It describes (in Kantian terms) laws of pure reason, but not norms of a universal legislation derived from practical reason, which a free will could obey. Why, then, does Husserl believe that he can claim practical efficacy for phenomenology as pure theory? He errs because he does not discern the connection of positivism, which he justifiably criticizes, with the ontology from which he unconsciously borrows the traditional concept of theory.

Husserl rightly criticizes the objectivist illusion that deludes the sciences with the image of a reality-in-itself consisting of facts structured in a lawlike manner; it conceals the constitution of these facts, and thereby prevents consciousness of the

interlocking of knowledge with interests from the life-world. Because phenomenology brings this to consciousness, it is itself, in Husserl's view, free of such interests. It thus earns the title of pure theory unjustly claimed by the sciences. It is to this freeing of knowledge from interest that Husserl attaches the expectation of practical efficacy. But the error is clear. Theory in the sense of the classical tradition only had an impact on life because it was thought to have discovered in the cosmic order an ideal world structure, including the prototype for the order of the human world. Only as cosmology was *theoria* also capable of orienting human action. Thus Husserl cannot expect self-formative processes to originate in a phenomenology that, as transcendental philosophy, purifies the classical theory of its cosmological contents, conserving something like the theoretical attitude only in an abstract manner. Theory had educational and cultural implications not because it had freed knowledge from interest. To the contrary, it did so because it derived *pseudonormative power from the concealment of its actual interest.* While criticizing the objectivist self-understanding of the sciences, Husserl succumbs to another objectivism, which was always attached to the traditional concept of theory.

IV

In the Greek tradition, the same forces that philosophy reduces to powers of the soul still appeared as gods and superhuman powers. Philosophy domesticated them and banished them to the realm of the soul as internalized demons. If from this point of view we regard the drives and affects that enmesh man in the empirical interests of his inconstant and contingent activity, then the attitude of pure theory, which promises *purification* from these very affects, takes on a new meaning: disinterested contemplation then obviously signifies emancipation. The release of knowledge from interest was not supposed to purify theory from the obfuscations of subjectivity but inversely to provide the subject with an ecstatic purification from the passions. What indicates the new stage of emancipation is that catharsis is now no longer attained through mystery cults but established in the will of individuals themselves by means of theory. In the communication structure of the polis, individuation has progressed to the point where the identity of the individual ego as a stable entity can only be developed through identification with abstract laws of cosmic order. Consciousness, emancipated from archaic powers, now anchors itself in the unity of a stable cosmos and the identity of immutable Being.

Thus it was only by means of ontological distinctions that theory originally could take cognizance of a self-subsistent world purged of demons. At the same time, the illusion of pure theory served as a protection against regression to an earlier stage that had been surpassed. Had it been possible to detect that the identity of pure Being was an objectivistic illusion, ego identity would not have been able to take shape on its basis. The repression of interest appertained to this interest itself.

If this interpretation is valid, then the two most influential aspects of the Greek tradition, the theoretical attitude and the basic ontological assumption of a structured, self-subsistent world, appear in a connection that they explicitly prohibit: the connection of knowledge with human interests. Hence we return to Husserl's critique of the objectivism of the sciences. But this connection turns *against* Husserl. Our reason for suspecting the presence of an unacknowledged connection between knowledge and interest is not that the sciences have abandoned the classical concept of theory, but that they have not completely abandoned it. The suspicion of objectivism exists because of the *ontological illusion of pure theory* that the sciences still deceptively share with the philosophical tradition *after casting off its practical content.*

With Husserl we shall designate as objectivistic an attitude that naively correlates theoretical propositions with matters of fact. This attitude presumes that the relations between empirical variables represented in theoretical propositions are self-existent. At the same time, it suppresses the transcendental framework that is the precondition of the meaning of the validity of such propositions. As soon as these statements are understood in

relation to the prior frame of reference to which they are affixed, the objectivist illusion dissolves and makes visible a knowledge-constitutive interest.

There are three categories of processes of inquiry for which a specific connection between logical-methodological rules and knowledge-constitutive interests can be demonstrated. This demonstration is the task of a critical philosophy of science that escapes the snares of positivism.[7] The approach of the empirical-analytic sciences incorporates a *technical* cognitive interest; that of the historical-hermeneutic sciences incorporates a *practical* one; and the approach of critically oriented sciences incorporates the *emancipatory* cognitive interest that, as we saw, was at the root of traditional theories. I should like to clarify this thesis by means of a few examples.

V

In the *empirical-analytic sciences* the frame of reference that prejudges the meaning of possible statements establishes rules both for the construction of theories and for their critical testing.[8] Theories comprise hypothetico-deductive connections of propositions, which permit the deduction of lawlike hypotheses with empirical content. The latter can be interpreted as statements about the covariance of observable events; given a set of initial conditions, they make predictions possible. Empirical-analytic knowledge is thus possible predictive knowledge. However, the *meaning* of such predictions, that is their technical exploitability, is established only by the rules according to which we apply theories to reality.

In controlled observation, which often takes the form of an experiment, we generate initial conditions and measure the results of operations carried out under these conditions. Empiricism attempts to ground the objectivist illusion in observations expressed in basic statements. These observations are supposed to be reliable in providing immediate evidence without the admixture of subjectivity. In reality basic statements are not simple representations of facts in themselves, but express the success or failure of our operations. We can say that facts and the relations between them are apprehended descriptively. But this way of talking must not conceal that as such the facts relevant to the empirical sciences are first constituted through an a priori organization of our experience in the behavioral system of instrumental action.

Taken together, these two factors, that is the logical structure of admissible systems of propositions and the type of conditions for corroboration suggest that theories of the empirical sciences disclose reality subject to the constitutive interest in the possible securing and expansion, through information, of feedback-monitored action. This is the cognitive interest in technical control over objectified processes.

The *historical-hermeneutic sciences* gain knowledge in a different methodological framework. Here the meaning of the validity of propositions is not constituted in the frame of reference of technical control. The levels of formalized language and objectified experience have not yet been divorced. For theories are not constructed deductively and experience is not organized with regard to the success of operations. Access to the facts is provided by the understanding of meaning, not observation. The verification of lawlike hypotheses in the empirical-analytic sciences has its counterpart here in the interpretation of texts. Thus the rules of hermeneutics determine the possible meaning of the validity of statements of the cultural sciences.[9]

Historicism has taken the understanding of meaning, in which mental facts are supposed to be given in direct evidence, and grafted onto it the objectivist illusion of pure theory. It appears as though the interpreter transposes himself into the horizon of the world or language from which a text derives its meaning. But here, too, the facts are first constituted in relation to the standards that establish them. Just as positivist self-understanding does not take into account explicitly the connection between measurement operations and feedback control, so it eliminates from consideration the interpreter's pre-understanding. Hermeneutic knowledge is always mediated through this pre-understanding, which is derived from the interpreter's initial situation. The world of traditional meaning discloses itself to the interpreter only to the extent that his own world becomes clari-

fied at the same time. The subject of understanding establishes communication between both worlds. He comprehends the substantive content of tradition by *applying* tradition to himself and his situation.

If, however, methodological rules unite interpretation and application in this way, then this suggests that hermeneutic inquiry discloses reality subject to a constitutive interest in the preservation and expansion of the intersubjectivity of possible action-orienting mutual understanding. The understanding of meaning is directed in its very structure toward the attainment of possible consensus among actors in the framework of a self-understanding derived from tradition. This we shall call the *practical* cognitive interest, in contrast to the technical.

The systematic sciences of *social action*, that is economics, sociology, and political science, have the goal, as do the empirical-analytic sciences, of producing nomological knowledge.[10] A critical social science, however, will not remain satisfied with this. It is concerned with going beyond this goal to determine when theoretical statements grasp invariant regularities of social action as such and when they express ideologically frozen relations of dependence that can in principle be transformed. To the extent that this is the case, the *critique of ideology*, as well, moreover, as *psychoanalysis*, take into account that information about lawlike connections sets off a process of reflection in the consciousness of those whom the laws are about. Thus the level of unreflected consciousness, which is one of the initial conditions of such laws, can be transformed. Of course, to this end a critically mediated knowledge of laws cannot through reflection alone render a law itself inoperative, but it can render it inapplicable.

The methodological framework that determines the meaning of the validity of critical propositions of this category is established by the concept of *self-reflection*. The latter releases the subject from dependence on hypostatized powers. Self-reflection is determined by an emancipatory cognitive interest. Critically oriented sciences share this interest with philosophy.

However, as long as philosophy remains caught in ontology, it is itself subject to an objectivism that disguises the connection of its knowledge with the human interest in autonomy and responsibility (*Mündigkeit*). There is only one way in which it can acquire the power that it vainly claims for itself in virtue of its seeming freedom from presuppositions: by acknowledging its dependence on this interest and turning against its own illusion of pure theory the critique it directs at the objectivism of the sciences.[11]

VI

The concept of knowledge-constitutive human interests already conjoins the two elements whose relation still has to be explained: knowledge and interest. From everyday experience we know that ideas serve often enough to furnish our actions with justifying motives in place of the real ones. What is called rationalization at this level is called ideology at the level of collective action. In both cases the manifest content of statements is falsified by consciousness' unreflected tie to interests, despite its illusion of autonomy. The discipline of trained thought thus correctly aims at excluding such interests. In all the sciences routines have been developed that guard against the subjectivity of opinion, and a new discipline, the sociology of knowledge, has emerged to counter the uncontrolled influence of interests on a deeper level, which derive less from the individual than from the objective situation of social groups. But this accounts for only one side of the problem. Because science must secure the objectivity of its statements against the pressure and seduction of particular interests, it deludes itself about the fundamental interests to which it owes not only its impetus but *the conditions of possible objectivity* themselves.

Orientation toward technical control, toward mutual understanding in the conduct of life, and toward emancipation from seemingly "natural" constraint establish the specific viewpoints from which we can apprehend reality as such in any way whatsoever. By becoming aware of the impossibility of getting beyond these transcendental limits, a part of nature acquires, through us, autonomy in nature. If knowledge could ever outwit its innate human interest, it would be by comprehending that

the mediation of subject and object that philosophical consciousness attributes exclusively to its own synthesis is produced originally by interests. The mind can become aware of this natural basis reflexively. Nevertheless, its power extends into the very logic of inquiry.

Representations and descriptions are never independent of standards. And the choice of these standards is based on attitudes that require critical consideration by means of arguments, because they cannot be either logically deduced or empirically demonstrated. Fundamental methodological decisions, for example such basic distinctions as those between categorial and noncategorial being, between analytic and synthetic statements, or between descriptive and emotive meaning, have the singular character of being neither arbitrary nor compelling.[12] They prove appropriate or inappropriate. For their criterion is the metalogical necessity of interests that we can neither prescribe nor represent, but with which we must instead *come to terms*. Therefore my first thesis is this: *The achievements of the transcendental subject have their basis in the natural history of the human species.*

Taken by itself this thesis could lead to the misunderstanding that reason is an organ of adaptation for men just as claws and teeth are for animals. True, it does serve this function. But the human interests that have emerged in man's natural history, to which we have traced back the three knowledge-constitutive interests, derive both from nature and *from the cultural break* with nature. Along with the tendency to realize natural drives they have incorporated the tendency toward release from the constraint of nature. Even the interest in self-preservation, natural as it seems, is represented by a social system that compensates for the lacks in man's organic equipment and secures his historical existence *against* the force of nature threatening from without. But society is not only a system of self-preservation. An enticing natural force, present in the individual as libido, has detached itself from the behavioral system of self-preservation and urges toward utopian fulfillment. These individual demands, which do not initially accord with the requirement of collective self-preservation, are also absorbed by the social system. That is why the cognitive processes to which social life is indissolubly linked function not only as means to the reproduction of life; for in equal measure they themselves determine the definitions of this life. What may appear as naked survival is always in its roots a historical phenomenon. For it is subject to the criterion of what a society intends for itself as *the good life*. My *second thesis* is thus that *knowledge equally serves as an instrument and transcends mere self-preservation.*

The specific viewpoints from which, with transcendental necessity, we apprehend reality ground three categories of possible knowledge: information that expands our power of technical control; interpretations that make possible the orientation of action within common traditions; and analyses that free consciousness from its dependence on hypostatized powers. These viewpoints originate in the interest structure of a species that is linked in its roots to definite means of social organization: work, language, and power. The human species secures its existence in systems of social labor and self-assertion through violence, through tradition-bound social life in ordinary-language communication, and with the aid of ego identities that at every level of individuation reconsolidate the consciousness of the individual in relation to the norms of the group. Accordingly the interests constitutive of knowledge are linked to the functions of an ego that adapts itself to its external conditions through learning processes, is initiated into the communication system of a social life-world by means of self-formative processes, and constructs an identity in the conflict between instinctual aims and social constraints. In turn these achievements become part of the productive forces accumulated by a society, the cultural tradition through which a society interprets itself, and the legitimations that a society accepts or criticizes. My *third thesis* is thus that *knowledge-constitutive interests take form in the medium of work, language, and power.*

However, the configuration of knowledge and interest is not the same in all categories. It is true that at this level it is always illusory to suppose an autonomy, free of presuppositions, in which knowing first grasps reality theoretically, only to be taken subsequently into the service of interests alien to it. But the mind can always reflect back upon the interest structure

that joins subject and object a priori: this is reserved to self-reflection. If the latter cannot cancel out interest, it can to a certain extent make up for it.

It is no accident that the standards of self-reflection are exempted from the singular state of suspension in which those of all other cognitive processes require critical evaluation. They possess theoretical certainty. The human interest in autonomy and responsibility is not mere fancy, for it can be apprehended a priori. What raises us out of nature is the only thing whose nature we can know: *language*. Through its structure, autonomy and responsibility are posited for us. Our first sentence expresses unequivocally the intention of universal and unconstrained consensus. Taken together, autonomy and responsibility constitute the only Idea the we possess a priori in the sense of the philosophical tradition. Perhaps that is why the language of German Idealism, according to which "reason" contains both will and consciousness as its elements, is not quite obsolete. Reason also means the will to reason. In self-reflection knowledge for the sake of knowledge attains congruence with the interest in autonomy and responsibility. The emancipatory cognitive interest aims at the pursuit of reflection as such. My *fourth thesis* is thus that *in the power of self-reflection, knowledge and interest are one.*

However, only in an emancipated society, whose members' autonomy and responsibility had been realized, would communication have developed into the non-authoritarian and universally practiced dialogue from which both our model of reciprocally constituted ego identity and our idea of true consensus are always implicitly derived. To this extent the truth of statements is based on anticipating the realization of the good life. The ontological illusion of pure theory behind which knowledge-constitutive interests become invisible promotes the fiction that Socratic dialogue is possible everywhere and at any time. From the beginning philosophy has presumed that the autonomy and responsibility posited with the structure of language are not only anticipated but real. It is pure theory, wanting to derive everything from itself, that succumbs to unacknowledged external conditions and becomes ideological. Only when philosophy discovers in the dialectical course of history the traces of violence that deform repeated attempts at dialogue and recurrently close off the path to unconstrained communication does it further the process whose suspension it otherwise legitimates: mankind's evolution toward autonomy and responsibility. My *fifth thesis* is thus that *the unity of knowledge and interest proves itself in a dialectic that takes the historical traces of suppressed dialogue and reconstructs what has been suppressed.*

VII

The sciences have retained one characteristic of philosophy: the illusion of pure theory. This illusion does not determine the practice of scientific research but only its self-understanding. And to the extent that this self-understanding reacts back upon scientific practice, it even has its point.

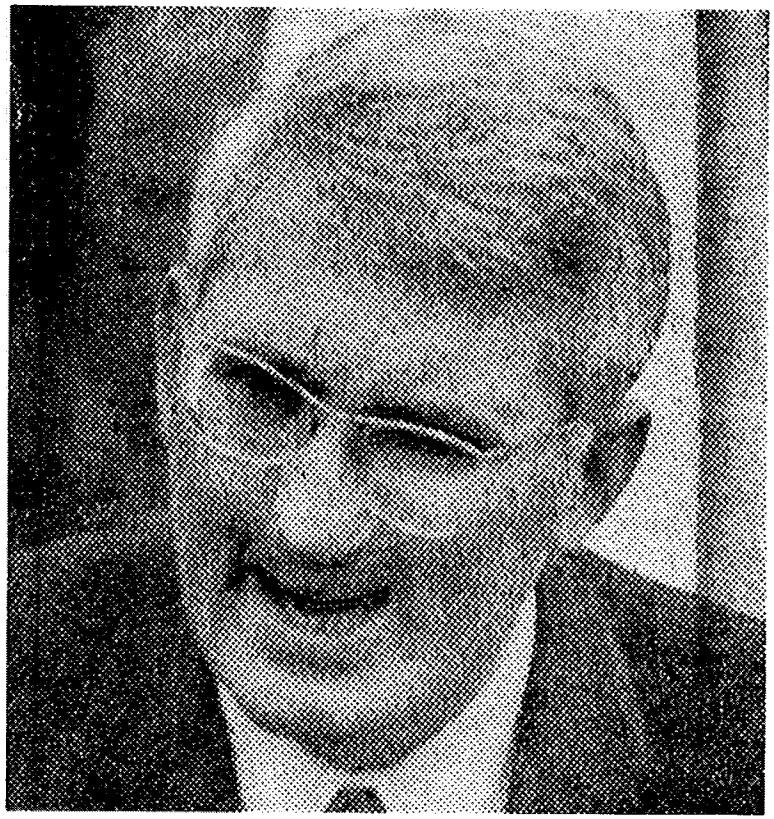
The glory of the sciences is their unswerving application of their methods without reflecting on knowledge-constitutive interests. From knowing not what they do methodologically, they are that much surer of their discipline, that is of methodical progress within an unproblematic framework. False consciousness has a protective function. For the sciences lack the means of dealing with the risks that appear once the connection of knowledge and human interest has been comprehended on the level of self-reflection. It was possible for fascism to give birth to the freak of a national physics and Stalinism to that of a Soviet Marxist genetics (which deserves to be taken more seriously than the former) only because the illusion of objectivism was lacking. It would have been able to provide immunity against the more dangerous bewitchments of misguided reflection.

But the praise of objectivism has its limits. Husserl's critique was right to attack it, if not with the right means. As soon as the objectivist illusion is turned into an affirmative *Weltanschauung*, methodologically unconscious necessity is perverted to the dubious virtue of a scientistic profession of faith. Objectivism in no way prevents the sciences from intervening in the conduct of life, as Husserl thought it did. They are integrated into it in any case. But they do not of themselves de-

velop their practical efficacy in the direction of a growing rationality of action.

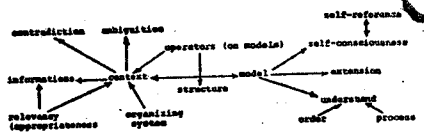
Instead, the positivist self-understanding of the *nomological sciences* lends countenance to the substitution of technology for enlightened action. It directs the utilization of scientific information from an illusory viewpoint, namely that the practical mastery of history can be reduced to technical control of objectified processes. The objectivist self-understanding of the *hermeneutic sciences* is of no lesser consequence. It defends sterilized knowledge against the reflected appropriation of active traditions and locks up history in a museum. Guided by the objectivist attitude of theory as the image of facts, the nomological and hermeneutical sciences reinforce each other with regard to their practical consequences. The latter displace our connection with tradition into the realm of the arbitrary, while the former, on the levelled-off basis of the repression of history, squeeze the conduct of life into the behavioral system of instrumental action. The dimension in which acting subjects could arrive rationally at agreement about goals and purposes is surrendered to the obscure area of mere decision among reified value systems and irrational beliefs.[13] When this dimension, abandoned by all men of good will, is subjected to reflection that relates to history objectivistically, as did the philosophical tradition, then positivism triumphs at the highest level of thought, as with Comte. This happens when critique uncritically abdicates its own connection with the emancipatory knowledge-constitutive interest in favor of pure theory. This sort of high-flown critique projects the undecided process of the evolution of the human species onto the level of a philosophy of history that dogmatically issues instructions for action. *A delusive philosophy of history, however, is only the obverse of deluded decisionism.* Bureaucratically prescribed partisanship goes only too well with contemplatively misunderstood value freedom.

These practical consequences of a restricted, scientistic consciousness of the sciences[14] can be countered by a critique that destroys the illusion of objectivism. Contrary to Husserl's expectations, objectivism is eliminated not through the power of renewed *theoria* but through demonstrating what it conceals: the connection of knowledge and interest. Philosophy remains true to its classic tradition by renouncing it. The insight that the truth of statements is linked in the last analysis to the intention of the good and true life can be preserved today only on the ruins of ontology. However even this philosophy remains a specialty alongside of the sciences and outside public consciousness as long as the heritage that it has critically abandoned lives on in the positivistic self-understanding of the sciences.

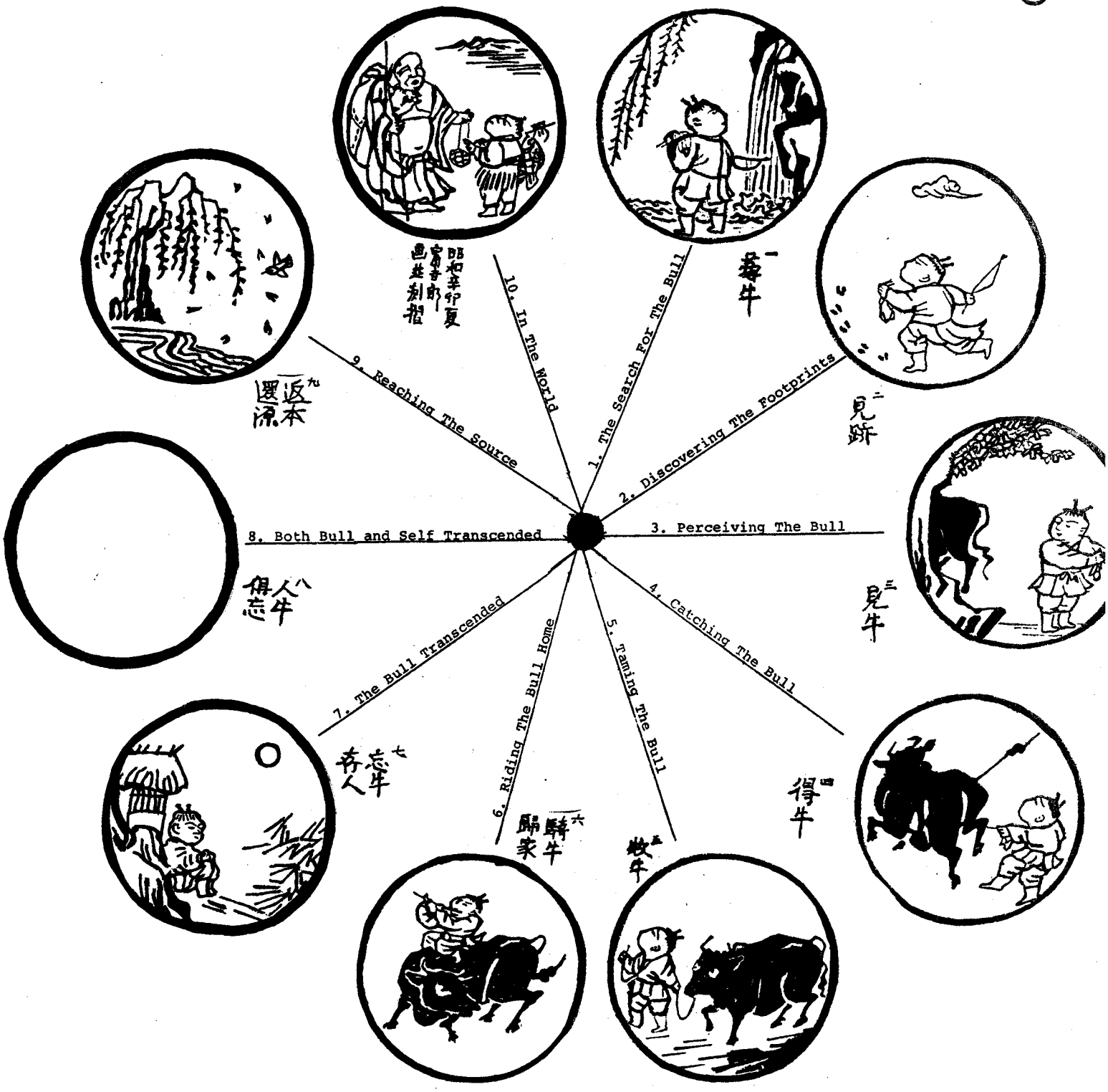


UNDERSTANDING

The name given that aspect of MODELLING WITHING CONTEXTS of a modelling system (such as the brain) which develops the patch-quilts of sub-models that allow the system to fine-tune the contexts of predicting the unknown (be it the unknown of a process in nature or the metaphysical unknowns such as man's relation to the order of the universe and of life and death).



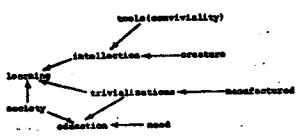
[J.K.]



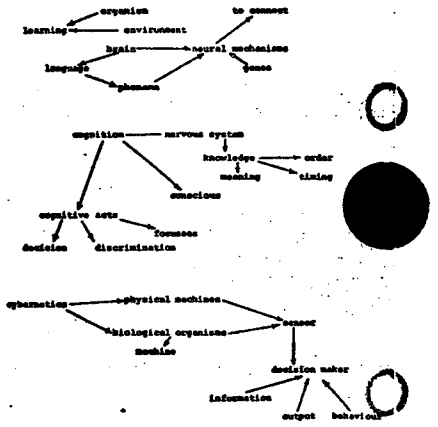
LEARNING

The balance of knowledge is determined by the ratio of two kinds of learning in a society. The first is a result of intellection which is a creative action of people on their environment. The second represents the result of man's trivialization by means of a manufactured milieu. The first kind of learning takes place in the personal involvement of people with each other and from their use of convivial tools; the second accrues to them as a result of purposeful and programmed training to which they are subjected. The relation between what can be learned from ordinary living and what must be learned as a result of intentional teaching differs widely with place and time.

Total learning expands when the range of spontaneous learning widens along with access to an increasing number of taught skills: both liberty and discipline then flower. This expansion of the balance of learning cannot go on forever. It is self-limiting. It can be optimized, but it cannot be forcibly extended. One reason it that man's life-span is limited. Another, just as inexorable, is that the specialization of tools and the division of labor reinforce each other: when centralization and specialization grow beyond a certain point, they require highly programmed operators and clients. In a society where this occurs, more of what each man must know is then due to what another man has designed and has obtained the power to force on him. The balance of learning deteriorates: it is skewed in favor of "education". People come to feel that they need "education". Learning thus becomes a commodity and, like any commodity that is marketed, it becomes scarce. The prevalence of education in the balance of learning paralyzes man's poetic ability, his power to endow the world with his personal meaning. [I.I.]

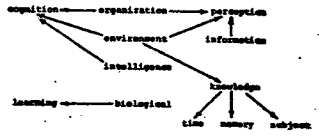


LEARNING



Learning is a word used to describe some of those situations where an organsim behaves differently before and after an exposure to the environment. I once put down everything I knew about learning in 2 articles which you can read if you like. Both are in The Neurosciences; in the 1967 volume (Quarton, Melnechuk, Schmitt, eds.) see p. 637; in the 1970 volume (Schmitt, ed.) see p. 161 and p. 289. [R.G.]

LEARNING



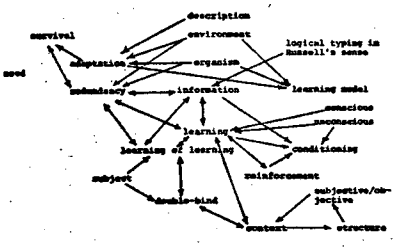
A term covering essentially all types of irreversible biological plasticity and rigidity at all levels of biological behavior organization (from micro to macro levels; from milliseconds to years). [K.W.]

LEARNING MODEL

My requirements of specifications for such a model include such notions as the following:

1. That the model shall accomodate interaction (mutual adaptation or the reverse) between organism and environment. "Learning" shall not be restricted to apply only to changes in the adaptive-maladaptive adaptation between organism and environment.

2. "Learning" shall include all receipt of information from the environment or receipt of information by one part of the brain (or body) from another part.



3. The information received shall be defined as to logical typing in Russell's sense.

4. The neurological level of receipt and storage shall be defined. [G.B.]

Asta su Abuelo.

Anti-Hodmanship : A Report on the State and Prospects of CAI

Gordon Pask

Dear Editors,

In reply to your inquiry, the art of computer-assisted instruction (henceforward CAI) is in a pretty dowdy state; a sort of depression. But the condition is improving quite rapidly and the recovery should be complete. Perhaps I can say this without seeming to be unduly critical just because I know of no one older in this (fortunately youthful) field, thus combining the advantage of a historical perspective with the culpability of a participant. At any rate these remarks, though occasionally acid, are intended constructively; neither as indicators of derogation nor exasperation.

1. The maladies are as follows:

(1) Research in this area has been chronically underfinanced in Europe and (less obviously so) in the U.S.A. and the U.S.S.R. Moreover, a rather large proportion of the money allocated to research/development has been mis-spent on respectable looking but pedestrian head counting and mark counting projects, or studies in which one juggernaut of a system is compared with another (on grounds that cannot, in the nature of the problem, be fully specified). These symptoms are nowadays less obtrusive; on the one hand, because of a growing realisation that CAI is important and that it is a " hard " science: on the other, because many of us have learned the salutary lesson that it is better to say what you want your system to do before investing in the machinery to do (or not do) it.

(2) Over the last 15 years, the CAI sub-culture has been obsessed with cost effectiveness criteria which, though reasonable in themselves, are consistently misinterpreted (just how, we shall see later). Since there is an internal misperception about the scope of CAI and thus the benefits likely to accrue from its implementation it is not surprising that the cases made to administrators have been, and to some extent still are, improperly stated, so that the standards of effectiveness are commonly inappropriate. For example, it is rarely meaningful to make estimates in terms of student hours or " station " hours or stored " words " per student hour. More subtle indices are needed and are nowadays available.

(3) Pragmatism is a good idea; take it for granted. But there are two species of pragmatist, at least. One interprets pragmatism as the valuable but prosaic solution of problems formulated on the basis of conventional wisdom. His occupational hazard is to mistake the regurgitation of solution methods for problem solving. The other type of pragmatist is an inventor. The trouble with him is that he poses more problems than he solves. Both kinds have their place in things, and they operate at all levels (theoretical, scientific, and technical). But, in a developing field like CAI, the innovative pragmatist is essential. Unfortunately, his activities have been discouraged in (rightly) demanding a pragmatic approach so that the first type of pragmatist came to dominate the experimental field; to the extent of being accepted as the *only type* of pragmatist.

One symptom of this bias is that our sub-culture used to suffer from the time-consuming pursuit of technical detail; for instance, studies of frankly outmoded terminals; trying to make sensible systems work in unsuitable computer-oriented languages or user-oriented languages (like the earlier members of the *Coursewriter* series) built for the realisation of programmed instruction alone. Somehow these pursuits were conceived as worthy; nowadays, they are generally agreed to be fatuous as well.

(4) Another fairly ubiquitous activity, also deemed worthy, was the habit of smudging empirical data about learning with various statistics. In the condensed records that emerge from this process, any information that may have been obtainable about learning as an explanatory, strategic, creative, or idiosyncratic process is obfuscated. Naturally, the data are well suited to the " respectable looking " studies, criticised in (2) above, but they have little bearing on education, either in theory or in practice.

(5) For a period, the sub-cultures of CAI and " Artificial Intelligence," so clearly companions in cognitive psychology and applied epistemology, were separated. Gladly, their estrangement has ended in reunion rather than divorce.

The change in thinking came about for several reasons, but one of them was the resolution of a prevailing confusion between computation science (cybernetics, system theory or just computation unqualified) and the operation of existing computing machines. It is true that computation science often uses computers as tools. But its subject-matter is much broader. Computation science deals with relational networks and processes that may represent concepts: with the structure of knowledge and the activity of real and artificial minds. Computation science lies in (even *is*) the kernel of CAI; it lends stature to the subject and bridges the interdisciplinary gap, between philosophy, education, psychology and the mathematical theory of organisations. Computer techniques, in contrast, bear the same relation to CAI as instrument making to physics or reagent manufacture to chemistry.

These facts, above all others perhaps, have changed the rather gloomy picture painted in (1), (2), (3) and (4).

In the context of a unified science of education, many aspects of which call for computer assistance if they are to be practicably realised, it is possible to resolve the dilemmas of (1) and (2): to say what we are really up to and to make an honest case to our patrons and sponsors. It is usually possible, in this context, to " say what the system should do " (often after some highly sophisticated and mechanised but small-scale experimentation) and thus avoid an embarrassing clutter of smart but misbegotten facilities. Likewise (3) becomes outmoded as a criticism, along with the terminals. Though there is a legacy of unsuitable hardware and software it is realised that fiddling with its quirks is " making the best of existing equipment," and not a serious pursuit in its own right. Finally, the condensed results of (4) (not, please, the data which may be excellent), stand out as the irrelevant piffle they are.

2. That is the present *state* of the art. Let us turn to the aims of CAI and the body of knowledge on which its growth as an art or a science or a technology is founded.

Here, the viewpoint will be idiosyncratic; some people share it, others have different ideals and divergent, though equally defensible, intentions.

(1) Most people in the educational profession (teachers, psychologists, university professors, curriculum designers, ETV producers, graduate students, and CAI merchants) have a sense of vocation. If you do not have, you will find it hard to make sense of the remaining paragraphs. Many of these practitioners owe their vocations to having at least glimpsed some moment of excellence: the power of evolving symbolism, the sheer joy of comprehension. The phenomena in question are varied. A child suddenly learns to learn, and you account for it by some sort of

neurophysiological change; but you know that the explanation is phoney and really you saw a miracle. A culture is engendered by a project; sometimes by an idea; or an adult who seems to have died in his twenties comes alive again. A design class in California, where the students were ignorant of electronics, gets, uses and innovates with, laser technology; all in a week. Von Foerster has over 100 well-documented examples, Illich more than 1,000, Papert has legions of instances, my own students have many. I know full well that a new renaissance is going on around us and see the signs of it as much in the ikons and the rituals as in the Ph.D. theses.

Though diverse, these phenomena are all evolutionary rather than revolutionary (a half-truth; they are non-destructive). They are all dramatic, large in magnitude and unmistakeable. They are all inexplicable within the standard scientific framework attuned to rationalising more picayune events (that is, a framework more or less copied from macrophysics, where it is apposite, to psychology, where it is not). Hence, we are trained not to notice them; but do so, and marvel.

Because there *are* such moments of excellence it is not absurd to escape from the fashionable pessimism spawned by " world dynamics " and other simple-minded forecasting schemes and feel a vocation to do something (but what?) about it all.

In general life, the phenomena dubbed " moments of excellence " are rare; a circumstance that is only in part attributable to the dissonance established by conventional training in scientific *techniques*. The chief reason for their rarity can be uncovered and represented formally; I shall gloss a lengthy argument by saying the world of learning and knowledge does not contain enough situations that count in a valid, non-trivial, very profound sense[1] as *conversations*; for there, and only there, are such moments of excellence manifest. On that account the new renaissance is a local phenomenon; symbolic evolution does not take off. It is not simply that moments of excellence are infrequently observed but also that, except under special and propitious conditions, neither are they evident to the participants.

How could moments of excellence be made more frequent (and, I wager on good grounds, more excellent)? There are several methods. The danger with most of them (Essalyn, the group techniques, the magician) is that the conversations induced with their assistance degenerate into the exercise of empathy as a surrogate for intellect. The perfectly predictable reason for this pathology is that the discourse is not effectively coupled to the growth of knowledge and is apt to be dissociated from cultural (and familial) tradition, hence the need for an order in which moments of excellence become immutable is half satisfied by rituals born of the limbic system, not of civilisation. As catalysts of conversation, immune to these defects, we have left the priest, the guru, the theatre and (odd though it may sound) the human use of CAI.

That, I think, is CAI's important role: to foster conversation which is coupled to a corpus of wisdom (some of it encoded no doubt, but some of it not) and thus to increase the frequency with which moments of excellence occur. The other legitimate agents (priests and dramatists) may use it in conjunction with educators, as a major tool; an amplifier and, quite possibly, a spur to innovation in its own right.

[1]Please pardon the emphasis; but I am not (whatever else) talking about the so-called " conversational " interaction between a man and an on-line computing machine. The word is used in all earnestness with its full meaning. On the other hand conversations need not be strictly interpersonal and certain man-machine interactions (of a not very well publicised type and mostly using visual rather than verbal utterances) also count. I am quite prepared to justify this at first sight outrageous claim to any interested reader.

Why should that be so important? Well, just because education is all about learning to learn, or teaching to learn or learning to teach, to create, to evolve. Apart from the alphabets (of letters and numerals), some notions of right and left, perhaps a bit of the multiplication table, facts can be stored in an engine that fits in your pocket. It is a waste of that most precious commodity, time, to store them in your brain. Moreover, a plethora of facts permanently inscribed in the brain may impede those mental reconstructions, explanations and explorations of concepts that constitute *memory* (parenthetically *memory* is not at all the same as storage, *that* at least becomes obvious on turning from *computer operation* to *computation science*).

Moreover, the following statements are true. Education is a regulatory system of civilisation. The current dimensions of communication, transportation and government render *it* the primary regulator. Even if society is decentralised, as the more thoughtful ecologists recommend, an educational system between five and 1,000 times more efficient than any one commonly and currently available is needed to avert, quite literally, collapse. That need is recognised in countries (like Mexico) with a high growth rate and a conventional school system that cannot possibly expand fast enough; there, they are using unconventional means to solve a pressing problem. But any means able to give the requisite magnitude of enhancement must, I submit, make each lesson into a focus of those currently rare " moments of excellence." If I did not believe that CAI could do that (it is a belief, though a reasonably founded one) I should not be in the field.

If you subscribe to this dogma then the criticism of section 1 (2) is stripped of its superficial perversity. Current evaluation schemes are based on the idea that CAI is " no worse than " other methods; here it must be vastly better to merit consideration at all. Also the criticism of section 1 (3) is far from capricious; the criteria being condemned are precisely those that level excellence into a higher mediocrity and obscure the exercise of individual strategies for learning or teaching which, in this context, prove all-important determinants of efficacity.

(2) It will be obvious that CAI, viewed in this light, does not resemble an elaborate branching program; nor is it necessarily concerned with operating booths and student stations.

Further, it is clear that some major revisions of thinking, very fundamental ones, must underlie the contention if it is to be taken seriously, as I hope it will be. CAI itself is the handmaiden of these innovations, but also, in a curious way, the progenitor of some of them.

In the remaining paragraphs I shall do my best to sketch (in most cases from their historical origins) the developments that culminated in a complete revision of thinking about such matters as learning and teaching: also to outline the features of those CAI systems able to sustain tutorial conversation and thus to magnify the force and scope of the educational process.

In the interests of brevity, these comments are slanted towards work in my own laboratory and few specific references are given. I appreciate, of course, that other people have worked concurrently along the same lines: their work is not stated explicitly. The reader anxious to remedy this imbalance (whilst retaining essentially the same stance) is referred to papers (listed at the end of this article) in which I have reviewed the subject giving proper and necessarily lengthy acknowledgements to other workers in this field. The volumes in which these papers appear cover CAI quite comprehensively and often from different points of view.[2]

[2]Good abstract services that concentrate on CAI are provided by the following agencies--Entelek; OECD; Training Research Abstracts; Enfield College of Technology; School of Education, Malmo, Sweden; Dept. of Commerce Washington (Translated Reviews of U.S.S.R.) and Council of Europe, Strasbourg, Survey Project 1970.

(3) Many operations are performed by a real life CAI system. For example, it may act like a library access device to retrieve data or instructions from a storage bank. It may record and aggregate a student's performance; it may construct gradings for a class, and so on. Though often of great importance, these features are peripheral to its main and minimal function: to interact with the student(s) and either to *teach* or else to *guide* a learning process. Concentrating on this function (for the other features can usually be added as required) let us start with the first instance of a teaching operation that called for computation on the part of a *machine*.[3]

During the late 1950s a number of adaptive teaching systems, mostly using special purpose computers, were employed for instructing both intellectual and perceptual motor skills; some of them being commercially exploited. These devices increase problem difficulty to compensate for a student's increasing proficiency (which they continually sense) and, vice versa, reduce it, if a student runs into trouble. Unlike the simple feedback arrangements embodied in a programmed text, or the slightly more sophisticated feedback tricks of a branching program, these CAI systems adjust and aim to optimise the criteria which govern adaptation. Moreover, they are generally multidimensional in so far as they compensate differentially for distinct response components and/or error factors. By about 1964, when similar arrangements were embodied in computer programs, it was clear that learning can be controlled and, in a limited sense, optimised by these devices, if, and only if, certain conditions are satisfied.

(a) There must be an adequate model for how the student solves problems and engages in the higher level problem-solving which is one facet of learning; otherwise there are no grounds for rating problems as more or less difficult (a cue, for example, may actually be misleading). (b) There must be a model for the subject-matter. (c) There must be a class of justifiable teaching strategies (together with means for measuring and predicting proficiency). (d) The man-machine interaction must be rich enough to sustain a dialogue.

For some situations, these requirements are satisfied and, if so, the man-machine interaction resembles a rather restricted conversation; it is, for example, entrapping. If not, the system's behaviour is manifestly instable. But, even when the conditions are not satisfied, it is still useful to regard all modes of adaptation as regulators of a student's uncertainty and to note that uncertainty regulation, by one means or another, is a prerequisite for effectual learning.

Except in the latter form, the adaptive paradigm fails to encompass realistically large learning situations, though it has a local utility attested by a great deal of experimental data, some of it quite recently garnered. Nevertheless the problem areas delineated by (a), (b), (c), (d), still provide a framework, which is used in section 2 (5) for discussing present-day issues.

(4) Around this period five separable lines of development got under way, using various subject-matters and categories of student:

(A) Production of systems for general tutorial conversation that could adumbrate topics of educational interest and a theory to go with them.

(B) Implementation of learning groups, in which students teach one another under the surveillance and guidance of a monitoring machine.

[3]After all, the adult participant in one of Vygotsky's "paired learning" experiments, computes the child's expected mode of behaviour; the author of a "structural communication" text or a mathetics text computes also, though he relegates on-line condition testing to the student himself.

(C) Scrutiny of the ways that a computer can be used in teaching. For example, it may be employed directly (as in section 2 (3) but usually augmented by an inquiry facility). Or, it may be used as a simulation tool, for instance, to present a country's economy in the form of a dynamic microcosm. Again, it can be used as a laboratory (*e.g.* in mathematics) with or without the pattern recognition capability to discern *what* or *how* the student makes laboratory models.

(D) Design of hardware and software systems, some of them (like PLATO or IMPACT) very elegant ones, for interfacing the tutorial.

(E) Development of means for accessing student terminals and/or data storage which are economically feasible for large student number and/or storage capacity (for example, ALF, which though rather inflexible as a student facility, has impressive characteristics in this respect).

Of these developments, very little will be said about (D) or (E). Item (B) can be treated alongside (A) and (C). The teach-yourself idea is powerful and the notion of a "learning group" is quite distinct from the notion of "many students treated individually by a single processor" (a more familiar plan). But it is approximately true to say that any conversational system can be extended, in principle, to deal with an n-person conversation where $n > 2$. The practical difficulties involved in instrumenting the scheme are likely to decrease as concurrently operating processors come into vogue as they are beginning to do.

(5) To take up the theme of section 2 (3). Item (a): a system that adequately models a student must be able to *learn* in a non-trivial sense, not merely to *adapt*.

First, the CAI system must learn because there are salient individual differences between students and changes in problem solving method that occur in the course of learning. Fortunately, students fall into theoretically predictable (and empirically verified) categories on a given occasion, *i.e.* any student who is free to do so adopts one *class* of learning strategy in respect of a certain subject-matter; moreover, there is a general individual *disposition* to adopt one or other class at the outset. Very crudely, there are students who have no overall concept of the subject-matter and must be instructed in order to learn (these can be eliminated from the population by using a special pretraining technique that shows them how to learn, *i.e.* to acquire a "learning set.") Of the remainder, some students prefer to move one step at a time and typically to isolate a topic about which they become *certain* before proceeding. Other students prefer to tackle the matter holistically; to access many topics in pursuit of their goal and, typically, to show confidence about solving problems under this goal long before they are certain of what to do.

Next, the CAI system must learn in order to interact with the student in his own terms. There is a current dispute about how much liberty a student should be allowed and the amalgamated result indicates "not too much, but not too much restriction either." The fact is that *if* a student is given liberty to explore, he (any pretrained student or anyone with the requisite foresight and self-awareness to begin with) *can* learn fast. But *if* he is given liberty, then the CAI system must match its strategy to his style and it can only do this if *its* data structure images *his* concept of the material.

All of this depends upon a model for the subject-matter (item (b) of section 2 (3)). Simple hierarchical schemes in which topics are treated like nouns *do* work to *some* extent (so-called knowledge structures) but a much more complicated picture of the subject-matter is mandatory for serious work. In it, topic names are treated like verbs and the entire mesh is cyclic in form.

Its canonical representation, as a graph, is called an entailment structure, and represents what *may be known*. It is crucial that the entailment structure depicts many permissible and possible ways of getting to know the subject-matter and, for each topic, at least one way of reconstructing the requisite concept from the concepts required to deal with topic relations that are already understood.

Although an entailment structure is not generally learned *by* the CAI system (*i.e.* it is prepared by subject-matter experts and analysts beforehand), it must be available to the system in order to accommodate the " learning about the student " already deemed mandatory. For example, the distinction between strategic types can only be exhibited in respect of an entailment structure for the subject-matter (on which any strategy appears as a certain cluster of *markings*); any message or advice delivered to the student on the basis of knowledge about his strategic type and competence is also contingent upon the entailment structure.

The teaching strategies (item (c) of section 2 (3)) come into the same domain as the learning strategies which would be generated by a self-aware (if need be, a pretrained) student in the absence of tuition. Many variants are possible, of course. But it can be shown that at least one condition is essential, *i.e.* any teaching strategy employed to enforce or to guide learning must be matched to a learning strategy (I) that this individual is prone to adopt and (II) that he is competent to execute if left on his own.

Regarding item (d) of section 2 (3) it is generally recognised that the man-machine interaction language should have many of the capabilities of natural language. There are numerous technical difficulties in the way of programming computers to interpret and manipulate full natural language dialogue. But it is important to see the wood for the trees. Roughly, the essential features of an acceptable language are that it should accommodate (α), asking questions that call for explanations in reply and which may legitimately treat many forms of explanation as correct (in contrast, most current languages, though they seem to be more powerful, are only able to comprehend multiple-choice or list questions and to adjudicate an answer or a list of answers as correct or not); (β) the language should be capable of representing analogy or, in literary terms, metaphor (either loose or structured).

It is hard to satisfy these requirements in the context of typed or spoken utterances and the current difficulties seem to present an insuperable barrier. But, to some extent, the problem is spurious. Using a visual display modality, and the modelling facilities mooted in section 2 (4) (C), it is possible to elicit modelling or constructive operations that are, within the terms of reference, one (or sometimes many) modes of explanation; further these can be " pattern recognised " and judged as satisfactory or not with respect to the entailment structure for the subject-matter.[4] By the same token, such models *can* be interpreted as expressing analogies (though not, as yet, verbal metaphors).

The point is important because it turns out that though responses to multiple-choice or list questions furnish valuable indices of a student's uncertainty and/or belief, they give little information about whether or not he possesses a concept and (more important) is able to reproduce it or reconstruct it as a memory. On the other hand, replies of an explanatory type do provide this information. Under appropriate circumstances a student can be said to *understand* a concept if he is able to explain it and to explain how he achieved it. One way of putting this point is to say that he could teach the concept to another student. If the necessary facilities are made available for this purpose then they can be employed, also, for the *laboratory* use of the computer; as a playground in which the student can construct and innovate.

Amongst the possible teaching strategies (section 2 (3) (c)) that have been mooted there is at least one class that shows the following properties and these strategies (by their nature) permit the student a constrained degree of liberty. These, rather than the adaptive paradigm (section 2 (3)), are now dubbed conversational.

(6) (i) As a pre-requisite for executing these strategies, the CAI system must furnish the student with a copy of the entailment structure for the subject-matter: we use a large map-like display. The points representing separate topics must be accessible and must be marked to indicate, at any instant, the topics the student is exploring; that he appreciates or is aiming to learn; the topics (usually several) he is currently working on; and the topics he already understands.

(ii) Any exploration strategy (a plan for learning whatever the entailment structure permits) is selected as a compromise between the student's option and the machine's (biased according to the desired mode) but so that the student's preferred learning strategy is matched (Condition (I) and Condition (II) of section 2 (5)); and so that the student shows evidence that he *can* learn in the prescribed fashion.

(iii) Matching secures a situation in which the student's gross level of uncertainty is regulated· but

(iv) Local conditions are adjusted (by a modified adaptive routine) so that his uncertainty in respect of the currently aimed-for topic and the cluster of currently worked-on topics is held within limits.

(v) Once committed to a topic, the student must explain it. Some CAI systems are able to interpret symbol string explanations; we use and prefer an explanation that is physically modelled by the student on a laboratory facility, inspected by the machine. A topic is marked *understood*, if and only if the student can both explain it and show how he learns to explain it.

(7) These are not the only teaching strategies, of course, but they are, in practice, remarkably effective; yielding an enhancement of learning between two and more than ten times the standard value (depending upon whether free study or rigidly controlled instruction is taken as the basis for comparison). There is appreciable transfer of " learning to learn " and good retention. The data obtainable from such a system give a detailed picture of progress, are more revealing than a gross learning and retention score and have obvious diagnostic value. The price paid for all this is quite large. In addition to tutorial text and graphics it is necessary to display a dynamically marked representation of the entailment structure and to construct a task specific simulator for modelling and eliciting explanations. On the other side of the coin, these facilities can be designed (and have been built) for subject-matters as diverse as statistics, applied science and history.

My own, experimentally deployed, equipment for executing conversational strategies (or any degenerate variants) has the acronym CASTE (Course Assembly System and Tutorial Environment). But we know that CASTE can be replicated on four student stations' worth of PLATO and hope to realise a transcription onto some such large-scale CAI facility in the near future.

(8) You may regard section 2 (1) as a euphoric vision; an obverse to the earlier deprecated depression. Or you may regard it (as I do) like the image on a jigsaw puzzle box, awaiting the pieces to make the picture real.

[4] Strictly, with respect to a *task structure* or *command graph* one of which is linked to each node representing a topic in the entailment structure.

The pieces needed to reify moments of excellence are given, in part, by actualising a tutorial conversation. But if they are to form a coherent pattern they must also have formal status both as psychological and computational units. It is at this level that CAI becomes an enthralling science. CASTE for example, was built in well-informed faith. It worked (hence, it is a pragmatically justifiable tool) and it furnished empirical data. But *subsequent* analysis of its operation has revealed that it only *can* work if the following statements are true: each one (it is a partial list only) tags one of the integrable units required to make sense of section 2 (1) at its face value and is also needed to justify some apparently sloppy arguments (especially for example, the cavalier approach to n-person conversations in section 2 (4)).

(a) A knowable topic is a relation.

(b) A *concept* is the reconstruction (technically, the reproduction) of a topic.

(c) A *memory* is the reproduction of a concept.

(d) An *individual* is a class of self-reproducing memories, viable in the surroundings afforded by suitably (and definably) related topics. Some sociologists would call it a *role*.

So, as characterised in a tutorial conversation, an individual may be correlated with one man, one brain; or it may be immanent in a group of men; or it may be that there are several in a single brain (as, when you " learn alone " you really house a distinguishable " teacher " and " learner "; both individuals).

From this point, we can assemble a theory of education which is somewhat beyond the scope of this article, though it is extremely relevant to the development of CAI. Instead, I shall conclude with one remark, emerging from this theory, which bears on all manner of CAI systems and provides a canon for appraising them. If a system is legitimately said to *teach*, then it must be able to *learn from* its *student* who may reverse the roles to play at *teacher*. That is what tutorial conversation *means*; I submit it is what *teaching* (in contrast to indoctrination, instruction, or ill-disciplined cavorting with knowledge) really means.

There is one interesting corollary. Whatever may be learned (the entity called an entailment structure in section 2 (5)) is fixed only for convenience. In general, it is open to growth and that is both theoretically the case and factually so (the course assembly aspect of CASTE which has not been discussed). In a very formal sense, there are gaps in knowledge and (as a conjecture) there are some unknowables; but to learning there is no end at all.

Acknowledgement

The research reported in this document has been supported (under various contracts and grants) by the Social Science Research Council of Great Britain or the U.S.A.F. Office of Aerospace Research through its European Office and carried out at System Research Ltd.

I wish to thank Mr. D. Kallikourdis and Mr. B. C. E. Scott for reading through and commenting on this paper.

References

Pask, G. and Scott, B. C. E. (1972) " Uncertainty Regulation in Learning Applied to Procedures for teaching concepts of Probability," Final Scientific Report SSRC Research Grant HR 1203/1, System Research Ltd., January 1972. To be published in *International Journal of Man-Machine Studies.*

Pask, G. and Scott, B. C. E. (1971) " Learning Strategies and Individual Competence," SSRC Final Scientific Report HR 983/1, System Research Ltd., January 1971. To be published in *International Journal of Man-Machine Studies.*

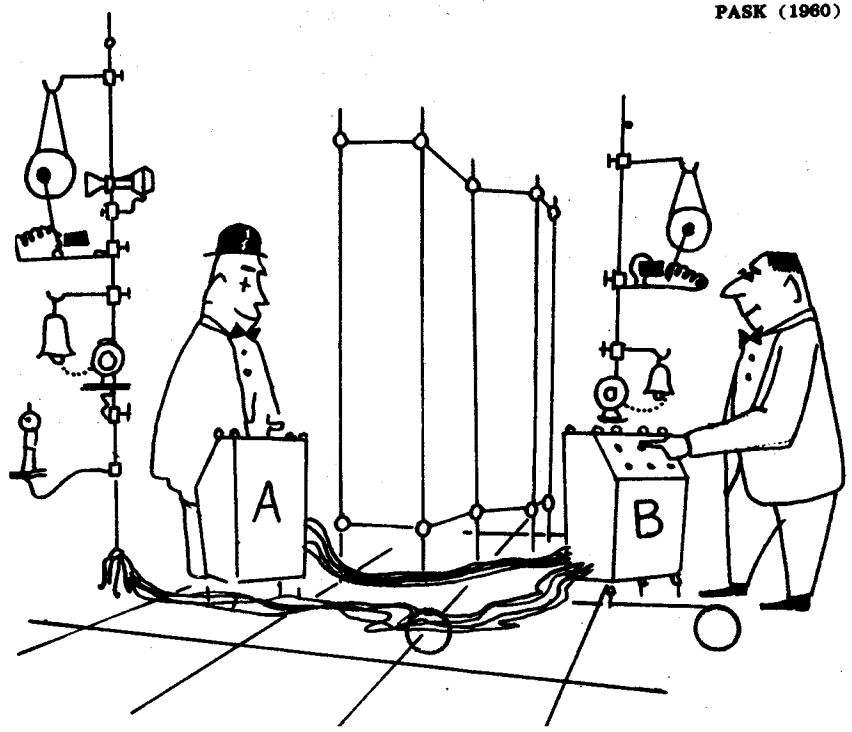
Pask, G. (1970) " Fundamental Aspects of Educational Technology, illustrated by the principles of Conversational Systems," in *Proceedings IFIP World Conference on Computer Education*, Vol. 1. *Invited papers*, R. Sheepmaker (Ed.). pp. 1-29 to 1-52. Amsterdam: IFIP.

Pask, G. (1970) " Computer Assisted Learning and Teaching," in *Proceedings of the Leeds Seminar on Computer Based Learning*. J. Annett and J. Duke (eds.) pp. 50–63, NCET.

Pask, G. (1972) " The Automization of Instruction and the Nature of Learning," in *Proc. Lehersysteme 72 Berlin*, P. Weltner (ed.). Munich Ehlehwirph Verlag, Revised version to be published in *International Journal of Man-Machine Studies.*

Gordon Pask is Director of Research at System Research Ltd., he is also Professor in the Department of Cybernetics at Brunel University and Professor in the Institute of Educational Technology at the Open University.

PASK (1960)



INNOVATION, IMAGINATION AND EDUCATION: A RECAPITULATION OF GORDON PASK ON LEARNING

(S. Sloan)

1. INTRODUCTION

The paradox of American institutes of learning is that they don't seem to concern themselves with the process of learning. They are interested in what people learn but disregard whether the process is interesting, enjoyable, or effective. People are forbidden the basis of education: the ability to recognize and solve problems, to critically analyze systems and to generate new ones. In its failure to provide people with the opportunity to develop these abilities, American education represents mediocrity and stagnation.

I have written this recapitulation of some of Gordon Pask's work in the hope it will serve as an alternative way of looking at the learning process: a way in which it is regarded as a heuristic process of imagination and innovation.

2. EXPERIMENTAL WORK

The following experiments performed by Gordon Pask and his co-workers provide an opportunity to examine the function of the environment in the learning process.

2.1 Experiment 1

In one experiment, subjects were presented with a board of randomly flashing lights. This activity usually fostered imagination, for even if the experimenter said that there were no rules (no relations), the subjects imagined their existence. However, this situation did not foster innovation; once the subjects generated a set of rules, they never changed them.

2.2 Experiment 2

In a second experiment which did foster innovation, the experimenters took black and white checkerboards, sixty-four squares each, rearranged the squares to produce various patterns, and projected them into a screen for .05 seconds each.

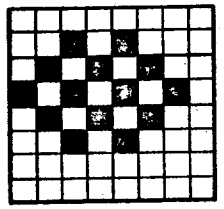


Figure 1. Sample figure

To generate these patterns, experimenters made up a list of slides, selecting patterns randomly and repeating those patterns which had been selected.

The subject was given a series of unlabeled buttons. He agreed to categorize these by generating labels referring to the properties of the projected patterns. Thus when a pattern appeared he either labeled and pushed a button, pushed a button he had previously labeled, or both. Labels were initially adjectives like "round", "spiked", "thin"; as the experiment proceeded, higher order predicates such as "limp-haired",

"spiked thin", and "turkey egg" became prevalent. In general, subjects used a maximum of four or five labels to describe all patterns.

When a subject classified a particular pattern, x, at a level of self-consistency specified by the experimenter, x was replaced by another pattern, m. This replacement was made at various degrees of self-consistency.

The interpretation of the subject's activities was carried out using the concepts of "redundancy" (R) and "uncertainty" (H), which are related by the formula

$$R = 1 - H/H_{max}$$

where H/H_{max} is the ratio of H, the subject's uncertainty with regard to a given state of affairs, to H_{max}, the maximum possible uncertainty with regard to that state of affairs. Thus when $H = H_{max}$, R is 0, a random state, whereby the shortest description of a particular pattern would be the pattern itself. As H approaches 0, R tends to unity, whereby one element of a pattern determines all elements of that pattern. Learning can be treated as a decrease in uncertainty.

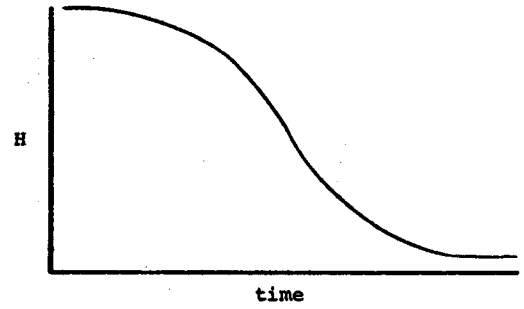


Figure 2.

Results can be summarized as follows:

Case I: The experimenters observed that if they set the replacement threshold too low (changed patterns after low recognition ability displayed by the subject), the subject became confused, frustrated, and opted out of the experiment (Figure 3).

Case II: If the threshold was set too high, recognition of patterns came easily so that subjects soon lost interest and became bored (Figure 4).

Case III: If, however, the threshold was set at a value which was neither so low that the checkerboard universe appeared chaotic and there seemed to be nothing to learn nor so high that a single set of coherent rules was discovered, the subjects became involved in the innovative process and reported experiencing a sense of euphoria during and immediately after the experiment (Figure 5). (It became necessary to detain subjects who were driving home for one hour after the completion of the experiment to cure a rash of automobile accidents involving euphorized subjects.)

A problem is solved as the subject makes the jump from appreciation to understanding. Pask suggested the interpretation of these graphs shown in Figure 6. Dr. Humberto Maturana once described this process as follows:

> The subject attempts to be consistent by placing an order on the universe until he observes that the order is no longer valid. He has the ability to invent a new system (make new representations) whenever he is willing to do so. So that in order to remain in a system, one must make an effort. So that if one wants to innovate, it is necessary to make an effort to stop making an effort to remain in the system and simply accept being outside of it. By staying in the system we simply confirm our need to do so.

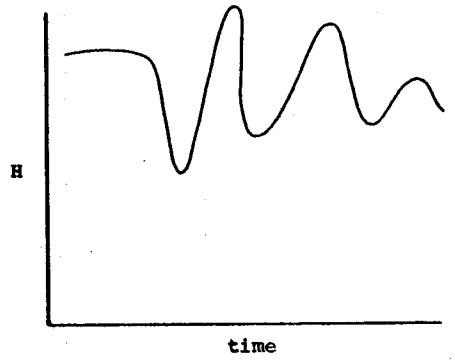


Figure 3

Case I: The subject becomes confused, frustrated and his attention fluctuates.

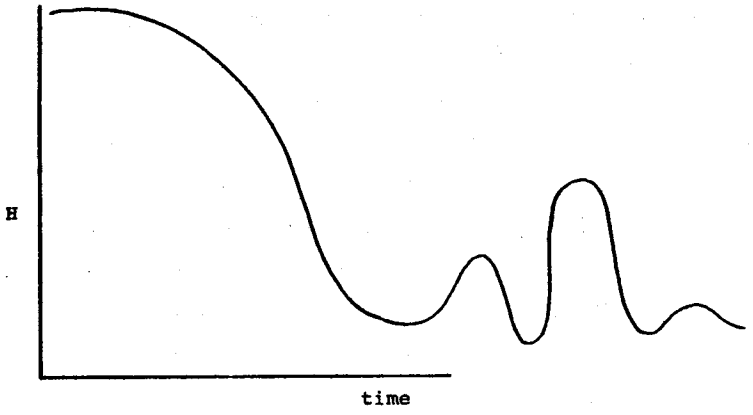


Figure 4

Case II: The subject learns quickly, loses interest, becomes bored and his attention fluctuates.

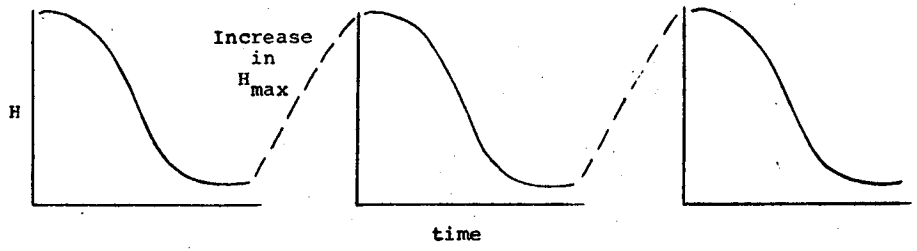


Figure 5

Case III: By increasing H_{max} at crucial times subject can be kept at desired level of uncertainty at which he is continuously reducing H. If a high anxiety factor is introduced, subject becomes stuck in one problem.

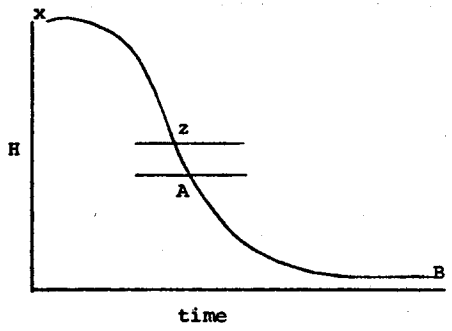


Figure 6

Point X -- Point of confusion; subject does not know how to proceed.

Region XZ -- Region of appreciation; to appreciate is to know of something.

Region ZA -- Operating region; region in which problem solving (innovating) occurs.

Region AB -- Region of understanding; to understand is to know; to be able to replicate; to be able to explain.

Point B -- Everything about a particular state is known; nothing left to learn.

3. STUDENTS, TEACHERS, AND LEARNING

Innovation depends on the learner's uncertainty with respect to a state of affairs; this uncertainty is often dependent on his interaction with others. In describing these interactions, it is convenient to make the distinction between student and teacher. However, both the student and the teacher are learners.

When learning is viewed as problem solving, the "teacher" is a simplifier (Figure 7).

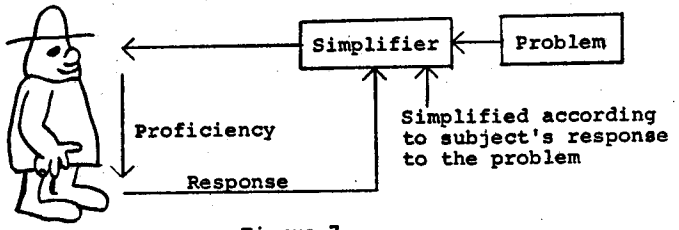


Figure 7.

Difficulty is defined to be 1 - Simplicity. The graph of difficulty and simplicity vs. time exhibits the following relationship (Figure 8).

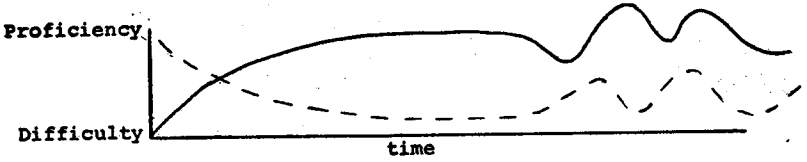


Figure 8.

From Figure 8, we can observe that we must allow for continuous simplification in order to facilitate problem solving. This can only be accomplished when the teacher and the student are responding to and learning from each other.

Learning is not simply stimulus-response. Learning resides in interpersonal relationships. To discuss learning adequately, one needs a linguistic hierarchy that discusses not only the universe the student observes but also the model the student makes of the teacher and the model the teacher makes of the student (Figure 9).

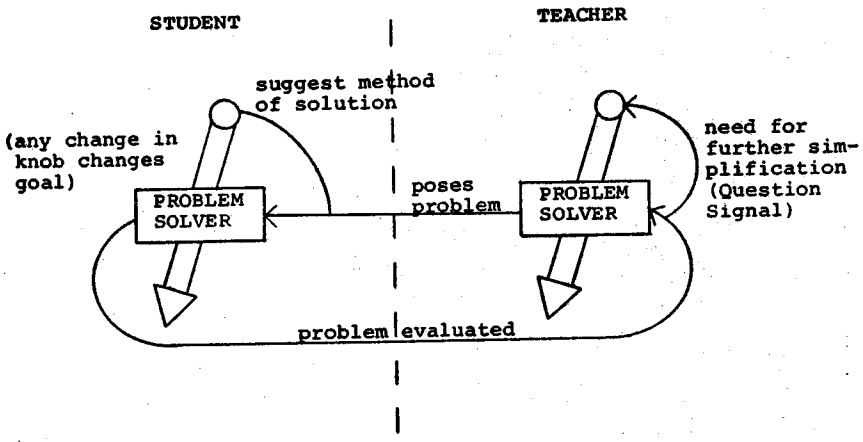


Figure 9.

If after an initial response by the student the problem is not yet solved, there is then a problem for which there is no problem solver. The student and the teacher must build models (representations) of each other at which levels of discourse can occur (Figure 10, where $REP_S(P)$ is the representation of P made by the student and $REP_T(P)$ is the representation of P made by the teacher).

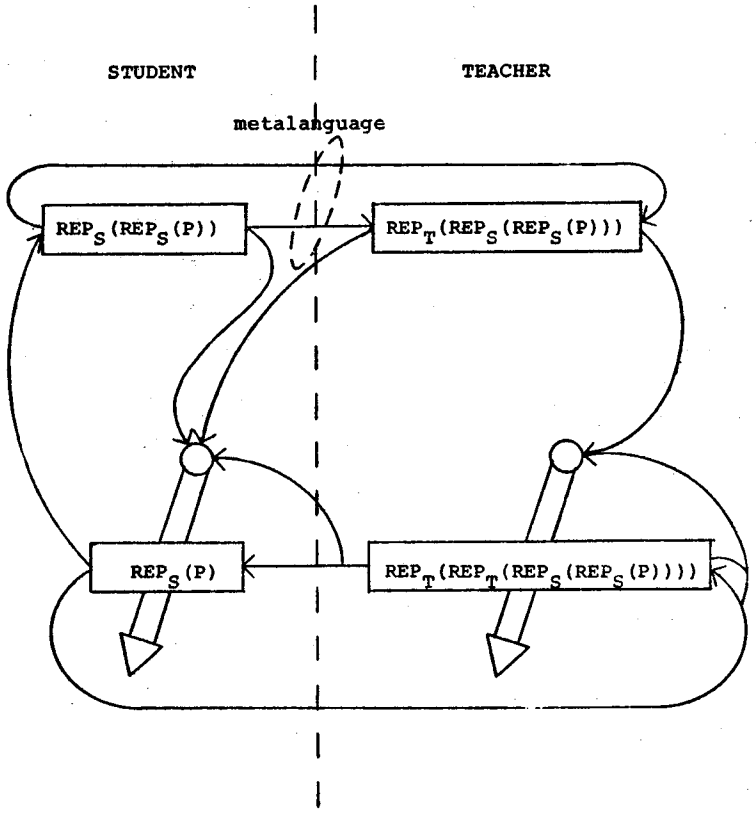


Figure 10.

In all cases, the participants should attempt to remain in the operating region between appreciation and understanding. Learning thus becomes the problem solving of problem solving.

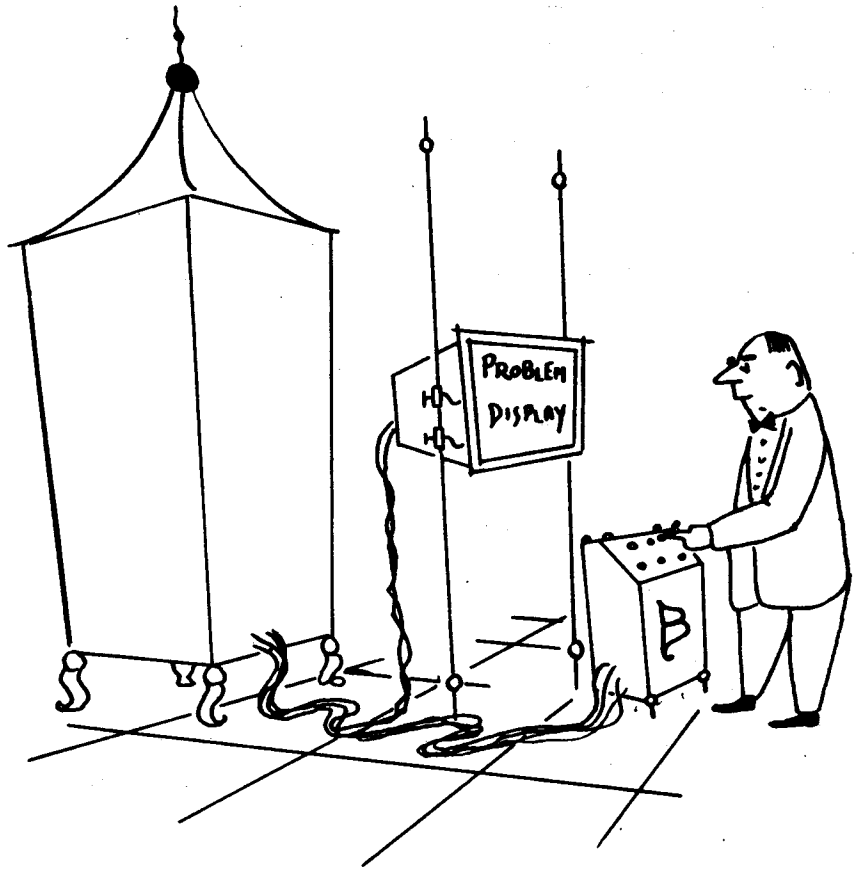
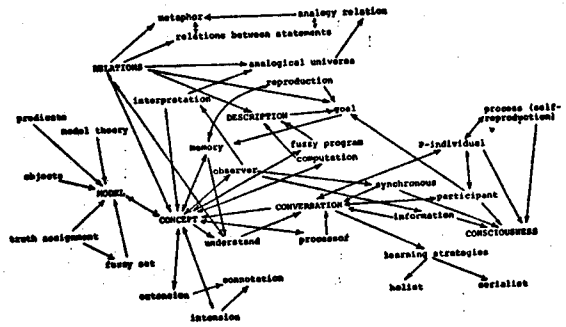


FIG. II. The situation in Figure I with Subject A enclosed in a box to simulate a teaching mechanism.

Si sabrà mas el discipulo?

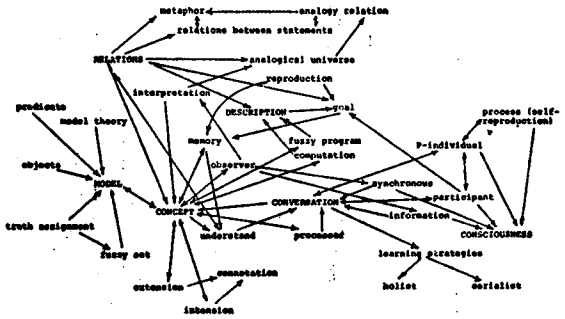
HOLIST

The holist and serialist strategies are two kinds of learning strategy that are initially exclusive under the conditions of a strict conversation anchored upon a given domain. [G.P.]



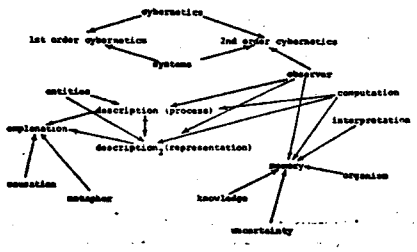
MEMORY

A memory is a concept that reconstructs or stablises or reproduces other concepts; by extrapolation (legitimised in view of the established existence of Fuzzy reproductive programs) learning is the construction or production of a concept. [G.P.]



MEMORY

The irreducible uncertainity of an observer with incomplete knowledge of the present internal state of a non-trivial machine (say, a living organism), which the observer interprets as a property of the machine. [H.V.F.]



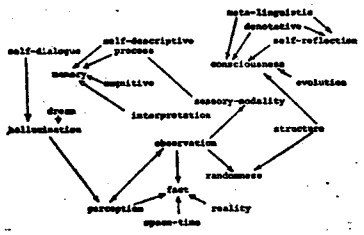
BY RELATING MACHINE AND TRANSFORMATION WE ENTER THE DISCIPLINE THAT RELATES THE BEHAVIORS OF REAL PHYSICAL SYSTEMS TO THE PROPERTIES OF SYMBOLIC EXPRESSIONS, WRITTEN WITH PEN ON PAPER

MEMORY

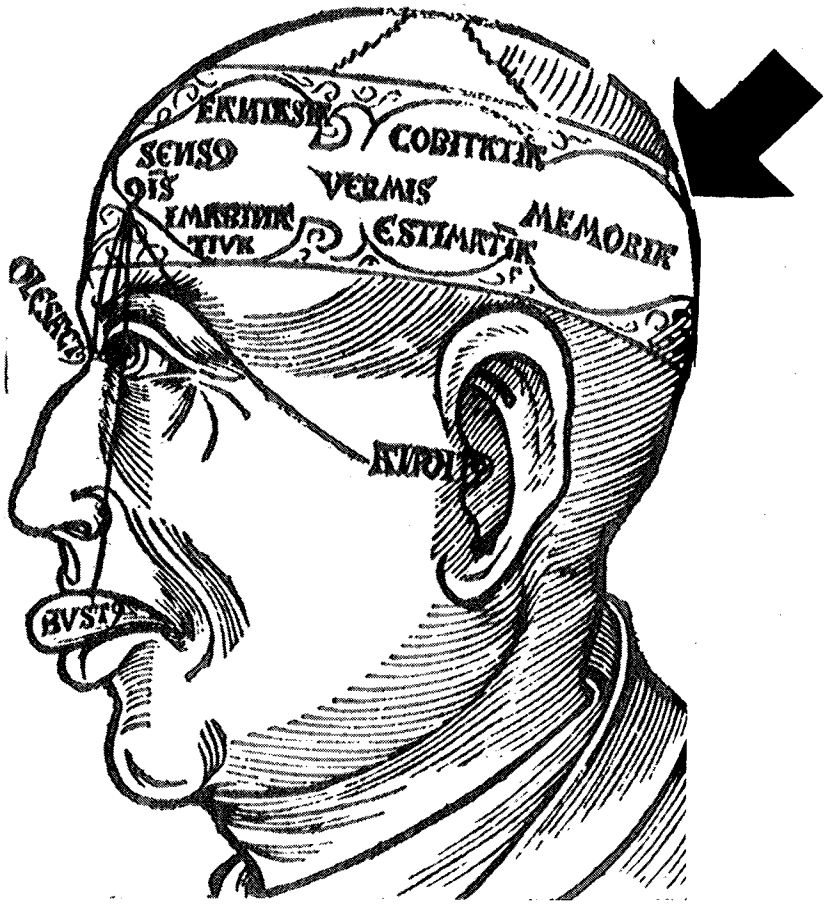
Conscious experience is the cortical (cognitive) interpretation of a particular level of subcortical arousal. Hence experience is state-bound (or memorized) since it can be recalled either by inducing a particular level of arousal or by presenting some symbol (image, melody, etc.) of its interpretation.

Memories, then are 'flashbacks' either on an individual level or on that of the species. More specifically, memory refers to that Platonic knowledge, "which is already there", the human 'program' laid down in not more than a dozen or so (archetypal) great stories, pictures, statues, songs, etc. which are re-created, written, composed, sculpted by each generation. Thus creativity re-affirms the self-descriptive nature of a process which reflects by reflecting itself.

In this state-bound sense all art is good art insofar as it can re-induce an experiential 'high' which closely resembles that level of arousal which prevailed in the composer, painter or poet 'once upon a time'. [R.F.]



THE LOGIC OF THE WORLD IS THE LOGIC OF DESCRIPTIONS (OF THE WORLD)



MEMORY WITHOUT RECORD

Heinz Von Foerster

VON FOERSTER:* Perhaps, I should make my position clear by opening with a metaphor. Let me confess that I am a man who is weak in properly carrying out multiplications. It takes me a long time to multiply a two or three digit number, and, moreover, when I do the same multiplication over and over again most of the time I get a different result. This is very annoying, and I wanted to settle this question once and for all by making a record of all correct results. Hence, I decided to make myself a multiplication table with two entries, one on the left (X) and one at the top (Y) for the two numbers to be multiplied, and with the product (XY) being recorded at the intersection of the appropriate rows and columns (Table 15).

TABLE 15

X·Y		Y									
		0	1	2	3	4	5	6	7	.	. .
	0	0	0	0	0	0	0	0	0	.	. .
	1	0	1	2	3	4	5	6	7	.	. .
	2	0	2	4	6	8	10	12	14	.	. .
X	3	0	3	6	9	12	15	18	21	.	. .
	4	0	4	8	12	16	20	24	28	.	. .
	5	0	5	10	15	20	25	30	35	.	. .
	6	0	6	12	18	24	30	36	42	.	. .
	7	0	7	14	21	28	35	42	49	.	. .
	.	.	.	.	.	.	.	.	.	.	. .
	.	.	.	.	.	.	.	.	.	.	. .

*Some of the ideas presented in this paper are the results of work sponsored in part by AF-OSR under Grant G-7-63 and by N.I.H. Grant 10718-01.

In preparing this table I wanted to know how much paper I need to accommodate factors X, Y up to a magnitude of, say, n decimal digits. Using regular-size type for the numbers, on double-bond sheets of 8 1/2 x 11 in, the thickness D of the book containing my multiplication table for numbers up to n decimal digits turns out to be approximately

$$D = n \cdot 10^{2n-6} \text{ cm.}$$

For example, a 100 x 100 multiplication table ($100 = 10^2$; $n = 2$) fills a "book" with thickness

$$D = 2 \cdot 10^{4-6} = 2 \cdot 10^{-2} = 0.02 \text{ cm} = 0.2 \text{ mm.}$$

In other words, this table can be printed on a single sheet of paper.

PRIBRAM: I thought you said you couldn't multiply?

VON FOERSTER: That is true. Therefore, I manipulate only the exponents, and that requires merely addition.

Now, I propose to extend my table to multiplications of ten-digit numbers. This is a very modest request, and such a table may be handy when preparing one's Federal Income Tax. With our formula for D, we obtain for $n = 10$:

$$D = 10 \cdot 10^{20-6} = 10^{15} \text{ cm.}$$

In other words, this multiplication table must be accommodated on a bookshelf which is 10^{15} cm long, that is, about 100 times the distance between the sun and the earth, or about one light-day long. A librarian, moving with the velocity of light, will, on the average, require a 1/2 day to look up a single entry in the body of this table.

This appeared to me not to be a very practical way to store the information of the results of all ten-digit multiplications. But, since I needed this information very dearly, I had to look around for another way of doing this. I hit upon a gadget which is about 5 x 5 x 12 in in size, contains 20 little wheels, each with numbers from zero to nine printed on them. These wheels are sitting on an axle and are coupled to each other by teeth and pegs in an ingenious way so that, when a crank is turned an appropriate number of times, the desired result of a multiplication can be read off the wheels through a window. The whole gadget is very cheap indeed and, on the average, it will require only 50 turns of the crank to reach all desired results of a multiplication involving two ten-digit numbers.

The answer to the question of whether I should "store" the information of a $10^{10} \times 10^{10}$ multiplication table in the form of a 8 1/2 x 11 in book 6 billion miles thick, or in the form of a small manual desk computer, is quite obvious, I think. However, it may be argued that the computer does not "store" this information but calculates each problem in a separate set of operations. My turning of the crank does nothing but give the computer the "address" of the result, which I retrieve at once—without the "computer" doing anything—by reading off the final position of the wheels. If I can retrieve this information, it must have been put into the system before. But how? Quite obviously, the information is stored in the computer in a structural fashion. In the way in which the wheels interact, in cutting notches and attaching pegs, all the information for reaching the right number has been laid down in its construction code, or, to put it biologically, in its genetic code.

If I am now asked to construct a "brain" capable of similar, or even more complicated stunts, I would rather think in terms of a small and compact computing device instead of considering tabulation methods which tend to get out of hand quickly.

During this Conference, it has been my feeling that, in numerous examples and statements, you gentlemen have piled up considerable evidence that the nervous system operates as a computer. However, to my great bewilderment—in so far as I could comprehend some of the points of discussion—in many instances you seemed to have argued as if the brain were a storehouse of a gigantic table. To stay with my metaphor, your argument seems to have been whether the symbols in my multiplication table are printed in green or red ink, or perhaps in Braille, instead of whether digit-transfer in my desk computer is carried out by friction or by an interlocking tooth.

I have to admit that, as yet, my metaphor is very poor indeed, because my computer is a deterministic and rigid affair with all rules of operation a priori established. This system cannot learn by experience, and, hence, it should not have been brought up in a conference on "memory."

I shall expand my metaphor, then, by proposing to build a computer that has first to learn by experience the operations I require it to perform. In other words, I posed to myself the problem of constructing an adaptive computer. However, before I attempt to suggest a solution for this problem, permit me to make a few preliminary remarks.

The first point refers to the temptation to consider past experience, again, in terms of a record. This approach offers itself readily because of the great ease with which a cumulative record can be manufactured. One just keeps on recording and recording, usually completely neglecting the problems that arise when attempting to utilize these tabulations. Ignoring this ticklish point for the moment, arguments of how to record may arise. Again, to use my metaphor, one may consider whether ink should be used which fades away after a certain time if it is not reinforced, or whether valid or invalid entries should be made with attached + or - signs, or whether the print should be fat or thin to indicate the importance of the entry, etc. Questions of this sort may

come up if we have the production of the big table in mind. But the kinds of problems are, of course, of an entirely different nature, if we consider building an adaptive computer device in which the internal structure is modified as a consequence of its interaction with an environment.

I believe that many of the remarks made during this meeting addressed themselves to the problem of how to write the record, instead of how to modify the structure of a computer so that its operational modality changes with experience. The interesting thing for me, however, is that these remarks were usually made when the speaker referred to how the system "ought" to work, and not when he referred to how the system actually works. This happened, for instance, when I understood Sir John to have made the remark that, for learning, we need an increase in synaptic efficacy with use. However, as he showed in his interesting example of the reflex action of a muscle that was for a while detached from the bone, these wretched cells would just work the other way round: Efficacy increased with duration of rest. Or, if I remember Dr. Kruger's point correctly, that in order to account for forgetting we need degeneracy of neurons. From his remarks, I understand that it seems to be very hard to get these cells to die.

It is perfectly clear that the comments about what these components ought to do are suggested by the idea that they are to be used in an adaptive recording device. I propose to contemplate for a moment whether the way these components actually behave is precisely the way they ought to behave, if we use them as building blocks for an adaptive computer device. For instance, degeneracy of neurons—if it occurs —may be an important mechanism to facilitate learning, if learning is associated with repression of irrelevant responses, as Sir John pointed out earlier. On the other hand, increased efficacy of a junction, caused by a prolonged rest period, may be used as a "forgetting" mechanism when associated with some inhibitory action.

McCONNELL: I sort of hate to ruin your lovely analogy, but aren't there many cases where a table would be more efficacious? For example, prime numbers?

VON FOERSTER: Correct. But please permit me to develop my story a bit further. I shall soon tighten the constraints on my computer considerably and your comment will be taken up later. Forgive me for developing my metaphor so slowly.

McCONNELL: We're getting caught in your strand.

VON FOERSTER: You shouldn't. Just wait a little and then hit me over the head. I shall give you many occasions, I think.

My second preliminary remark with regard to the construction of an adaptive computer is concerned with the choice of a good strategy of how to approach this problem. Fortunately, in the abstract that was distributed prior to the meeting, Sir John gave us an excellent guideline: "Learning involves selectivity of response, and presumably inhibition would be significantly concerned in the repression of irrelevant response." In other words, learning involves selective operations; but, in order to obtain the results of these operations, a computing device is needed to carry out these selective operations.

PRIBRAM: Did you know, Sir John, what you were saying?

ECCLES: No. (Laughter.)

VON FOERSTER: How would you do it, otherwise? How would you select? How would you have selective manipulation?

ECCLES: How does an animal do it?

VON FOERSTER: I think an impressive array of suggestions have come up during this meeting. I have in mind, for instance, Sir John, your demonstration of changes in the efficacy of synaptic junctions as a result of various stimulations. Instead of interpreting these changes as storage points for some fleeting events, I propose to interpret them as alterations in the transfer function in a computer element. In other words, I propose to interpret these local changes as modifications—admittedly minute—of the response characteristic of the system as a whole. In Dr. Uttley's presentation, he considered each junction as acting as a conditional probability computer element. Dr. Hydén's most sophisticated scheme is to compute the appropriate proteins with the aid of coded DNA and RNA templets. Of course, the biochemist would probably use the terms "to form" or "to synthesize," instead of "to compute," but in an abstract sense these terms are equivalent.

Let me return to the original problem I posed, namely, the construction of a computing device that changes its internal organization as a consequence of interaction with its environment. I believe it is quite clear that, in order to make any progress in my construction job, I have to eliminate two questions: First, what is this environment to which my computer is coupled? Second, what is my computer supposed to learn from this environment?

As long as I am permitted to design the rules that govern the events in this environment and the task my computer has to master, it might not be too difficult to design the appropriate system. Assume for the moment that the environment is of a simple form that rewards my computer with an appreciable amount of energy whenever it comes up with a proper result for a multiplication problem posed to it by the environment. Of course, I could immediately plug my old desk calculator into this environment, if it were not for one catch: The number system in which the environment poses its questions is not specified a priori. It may be a decimal system, a binary system, a prime-number product representation, or—if we want to be particularly nasty—it may pose its problems in Roman numerals. Although I assume my computer has the Platonic Idea of multiplication built in, it has to learn the number system in order to succeed in this environment.

My task is now sufficiently specified; I know the structure of the environment, I know what my system has to learn, and I can start to think of how to solve this problem.

Instead of amusing ourselves with solving the problem of how to construct this mundane gadget, let me turn to the real problem at hand, namely, how do living organisms succeed in keeping alive in an environment that is a far cry from being simple.

The question of the environment to which our systems are coupled is now answered in so far as it is nature, with all her unpredictabilities, but also with her stringent regularities which are coded in the laws of physics or chemistry.

We are now ready to ask the second question: What do we require our organisms to learn? Perhaps this question can be answered more readily if we first ask: "Why should these organisms learn at all?" I believe that if we find a pertinent answer to this question we will have arrived at the crux of the problem which brought us here. With my suggestion of how to answer this question, I will have arrived at the central point of my presentation. I believe that the ultimate reason these systems should learn at all is that learning enables them to make inductive inferences. In other words, in order to enhance the chance of survival, the system should be able to compute future events from past experience; it should be an "inductive inference computer." On the other hand, it is clear that only a system that has memory is capable of making inductive inferences because, from the single time-slice of present events, it is impossible to infer about succeeding events unless previous states of the environment are taken into consideration.

I have now completed the specifications of my task: I wish to construct an inductive inference computer whose increase of internal organization should remove uncertainties with respect to predictions of future events in its environment.

Having reached this point, let us look back to the position where we still pitched a computing device against a recording device in order to tackle our memory problem. It is clear from the task I just described that a record of the past, as detailed and as permanent as one may wish, is of no value whatsoever. It is dead. It does not give us the slightest clue as to future events, unless we employ a demon that permanently zooms along this record, computes with lightning speed a figure of merit for each entry, compares these figures in a set of selective operations, and computes from these the probability distribution of the next future events. He must do all this between each instant of time. If we insist on making records, we transfer our problem of memory to the potentialities of this demon who now has the job of acting as an inductive inference computer. Consequently, I may as well throw away the record and consider the construction of this demon who does not need to look at the record of events, but at the events themselves.

I have now concluded the metaphorical phase of my presentation which, I hope, has in some qualitative way outlined my position and my problem. I propose now to consider the problem from a quantitative point of view. In other words, before approaching the actual construction of such an inductive inference computer, it may be wise to estimate, in some way or another, how much internal organization we expect our computer to acquire during its interaction with the environment, and how much uncertainty with respect to future events it is able to remove by its acquisition of higher states of order.

It is fortunate that two decisive concepts in my argument can be defined in precise, quantitative terms. One is the concept of uncertainty, the other, the concept of order. In both cases it is possible to define appropriate measure functions which allow the translation of my problem into a mathematical formalism. Since the whole mathematical machinery I will need is completely developed in what is known today as "Theory of Information," it will suffice to give references to some of the pertinent literature (13, 2, 4) of which, I believe, the late Henry Quastler's account (11) is the most appealing one for the biologically oriented. I have the permission of the Chairman to redefine some of the basic concepts of this theory for the benefit of those who may appreciate having their memories refreshed without consulting another source.

With my apologies to those who will miss rigor in the following shadow of an outline of Information Theory, let me quickly describe some of its basic vocabulary.

For the first step in this development, namely the derivation of measure-function for uncertainty, called "Entropy", denoted by H and defined by

$$H = -\sum_{1}^{n} p_i \log_2 p_i$$

where the quantities p_i $(i = 1 \to n)$ represent the probabilities of the occurence of states S_i of the system under consideration, is to be found on p.184 as introduction for the section on Tables in Information Theory and Combinatorics.

O O O

The next step in my development is to use these quantitative concepts to construct still another measure function, this time a measure of "order." Again, I shall be guided by intuitive reasoning when selecting from the vast amount of possible measure functions the one that fulfills some desired criteria. First, I would like to suggest that whenever we speak of "order" we mean it in a relative sense, that is, we refer to the state of order of a particular universe. We say that a room is in various states of order or disorder, or that a desk is a mess, and so on. Hence, if we wish to state that a given "universe" is in complete disorder, our function representing a measure of order should vanish. Conversely, perfect order may be represented by a value of unity. Consequently, various states of order of a given universe may be represented by any number between zero and one. Further hints as to the general form of the function that expresses an order measure may be taken from the truism that our uncertainty H about a completely disordered universe is maximum ($H = H_{max}$), while, in a deterministic universe, ($H = 0$) order is perfect. This suggests that a measure of order an observer may associate with a particular universe is just the difference between his actual and maximum uncertainty of this universe in reference to maximum uncertainty. Accordingly, we define tentatively, Ω , a measure function of order by:

$$\Omega = \frac{H_{max} - H}{H_{max}} = 1 - \frac{H}{H_{max}} \qquad (7)$$

Clearly, the two conditions as discussed above are fulfilled by this function, because for $H = H_{max}$ the order measure will vanish ($\Omega = 0$), while for perfect order the uncertainty vanishes and the order measure approaches unity ($\Omega = 1$).

I propose to further test this expression, now under addition. Assume two universes, U_1 and U_2, with which we may associate measures of actual and maximum uncertainty, H_1, H_{m1} and H_{m2}. We further assume both universes to be in the same state of order:

$$\Omega_1 = 1 - \frac{H_1}{H_{m1}} = \Omega_2 = 1 - \frac{H_2}{H_{m2}} ,$$

or

$$\frac{H_1}{H_{m1}} = \frac{H_2}{H_{m2}} .$$

This condition is fulfilled if

$$H_2 = k\, H_1 ,$$

and

$$H_{m2} = k\, H_{m1} .$$

I now propose to drop the distinction between the two universes and treat both as parts of a large universe. What is the measure of order for this large universe? With Equation (7) defining Ω, Equation (3) the addition theorem for H , and the above identities, we have:

$$\Omega = 1 - \frac{H_1 + H_2}{H_{m1} + H_{m2}} = 1 - \frac{H_1 (1 + k)}{H_{m1} (1 + k)}$$

$$= 1 - \frac{H_1}{H_{m1}} = \Omega_1 = \Omega_2 .$$

In other words, by combining two equally ordered universes, the measure of order remains unchanged, as it intuitively should be.

This measure of order will be helpful in making quantitative estimates of the changes in the internal organization of our inductive inference computer during its interaction with the environment. Presently, let me give you a few examples of systems whose order increases at the cost of external or internal energy consumption. I have already given one example: the magnetic die. Table 17 lists the values of Ω for the various time intervals and uncertainties given in Table 16.

TABLE 17

TIME	t_0	t_1	t_2	t_3	t_4
H	2.582	2.519	2.249	1.582	0.000
Ω	0.000	0.025	0.129	0.387	1.000

Another example may be the growth of crystals in a supersaturated solution. Again, the organization of the system increases as the diffused molecules attach themselves to the crystal lattice, reducing in this process, however, their potential energy.

PRIBRAM: This poses a problem, because the biologist is less interested in this sort of order than he is in the one that involves sequential dependencies. This is why a hierarchical measure of some sort would be more appropriate, unless you can show how you can derive it from your equation.

VON FOERSTER: My measure of order is so general that, I believe, there shouldn't be any difficulty in dealing with the type of order that arises from sequential dependence. Permit me to suggest how I think this may be done. Sequential dependencies express themselves in the form of transition probabilities p_{ij}, that is, the n^2 probabilities for a system that is in state S_i to go into state S_j. If the states of a system are independent of previous states, as is the case with the coin and the die, all p_{ij}'s are, of course, just the p_j's. However, a system that learns will develop strong sequential dependencies as you suggested, and, hence, will shift the p_{ij}'s away from the p_j's. Again, a measure of uncertainty H can be defined for this state of affairs simply by working with the mean value of the various uncertainties H_i, as they can be computed for all states that immediately follow state S_i. Since

$$H_i = - \sum_{j=1}^{n} p_{ij} \log_2 p_{ij} ,$$

we have our rule of developing a mean-value:

$$H = \bar{H}_i = \sum_{1}^{n} p_i H_i .$$

If n, the number of states of the system, remains the same, $H_{max} = \log_2 n$ is unchanged and we may watch how Ω increases steadily as the sequential constraints are going up—that is, as H is going down—while the animal is being trained.

Since changing the internal organization of a system, whether in a spatial or temporal sense, to higher and higher levels of organization is a crucial point in my description of so-called "self-organizing systems" that map environmental order into their own organization, let me establish the criteria which have to be satisfied if we wish our system to be such a self-organizing system. Clearly, for such systems, Ω should increase as time goes on, or:

$$\frac{d\Omega}{dt} > 0.$$

Since our measure of order is a function of H and H_{max}, both of which may or may not be subjected to changes, we obtain the desired criterion by differentiating Equation (7) with respect to time:*

$$\frac{d\Omega}{dt} = - \frac{H_m \frac{dH}{dt} - H \frac{dH_m}{dt}}{H_m^2} > 0.$$

This expression can be transformed into something more tangible. First, we note that for all systems of interest, $H_m > 0$, because only for systems capable of precisely one state $H_m = \log_2 1 = 0$. Second, we divide both sides of the unequality with the product $H \cdot H_m$ and obtain the important relation:

$$\frac{1}{H_m} \frac{dH_m}{dt} > \frac{1}{H} \frac{dH}{dt} \qquad (8)$$

This says that if, and only if, the relative increase of maximum uncertainty is larger than the relative increase of the actual uncertainty, then our system is in the process of acquiring higher states of internal organization.

BOWER: Are empirical considerations operating here?

VON FOERSTER: No, not a single empirical consideration. This is a straightforward derivation starting from one definition and using one criterion.

*For typographical reasons H_{max} will be written H_m.

BOWER: I can think of several counter-examples in which, although something is being learned, the behavior is becoming increasingly more random or disordered, in your sense. For example, I am sure I could train a rat to vary his behavior from trial to trial in a way that is completely unpredictable to me; one simply reinforces variance differentially according to some criterion—for example, by reinforcing a rat's lever press only if its latency differs by at least 2 sec from the preceding lever press. Second, in extinction of learned behavior, the reference response declines in strength and occurs less certainly. You would describe it as increasing disorder, yet it is a lawful and uniform process, particularly so when competing responses are recorded.

VON FOERSTER: I would say that in such a situation training enlarges the behavioral capacities of the rats by creating new states in the rat's behavioral universe. Hence, you operate on H_{max} such that dH_m/dt is larger than zero. If these animals do not deteriorate otherwise, by letting H go up too fast, you have indeed taught them something.

JOHN: Would you define H_{max}, please?

VON FOERSTER: H_{max} can be defined simply as the uncertainty measure of a system with equiprobable states that are also independent of their precursor states. Under these conditions, as we have seen earlier, H_{max} is just $\log_2 n$, where n is the number of states.

JOHN: Precisely. Therefore, it seems to me that the derivative of H_m with respect to time must always be zero.

VON FOERSTER: That is an excellent suggestion. You pointed out one of the fascinating features of this equation, namely, that it accounts for growth. You have already anticipated the next chapter of my story.

To see immediately that dH_m/dt must not necessarily always be zero, let me replace H_m by $\log_2 n$, or, for simplicity, by a ln n, where $a = 1/\ln 2$ is a scale factor converting base 2 log's into natural log's. Then:

$$\frac{dH_m}{dt} = a \frac{d \ln n}{dt} = \frac{a}{n} \frac{dn}{dt} . \qquad (9)$$

Consider for the moment an organism that grows by cell division. In the early stages of development the number of cells usually grows exponentially:

$$n = n_o e^{\lambda t} ,$$

and, with

$$H_m = a \ln n = a \ln n_o + a \lambda t , \qquad (10)$$

the rate of change of maximum entropy becomes

$$\frac{dH_m}{dt} = a \lambda ,$$

which is a positive constant! This means that the actual entropy H of the system must not necessarily decrease in order for the system to acquire higher states of organization. According to Equation (8), it is sufficient that the relative rate of change of H remains just below that of H_{max}. An observer who pays attention only to the increase of the actual uncertainty H may get the impression that his system goes to pot. If we look at our cities today we may easily get this impression. However, if we consider their rapid growth, they represent centers of increasing organization. But, to go back to the population of dividing and differentiating cells: Organized growth of tissue represents considerable constraints on the possible arrangement of cells, hence, the probability distribution of their position is far from uniform. Consequently, H changes during the growth phase of the organism very slowly indeed, if at all. Nevertheless, a conservative model for a tentative expression for the growth of H after Equation (10) is:

$$H = a \ln n_o + \mu a \lambda t , \qquad (11)$$

where $\mu < 1$. The measure of order for the growing organism becomes, with Equations (10) and (11),

$$\Omega = 1 - \frac{\ln n_o + \mu \lambda t}{\ln n_o + \lambda t} .$$

At early stages of its development ($t = 0$), we have

$$\Omega(0) = 0 ,$$

while at its mature state ($t \to \infty$):

$$\Omega(\infty) = 1 - \mu > 0 ;$$

that is, the organism is indeed an "organism."

The possibility of accounting for the acquisition of higher states of organization by incorporation of new states into the system—provided that this incorporation takes place in an orderly fashion—has been illuminated for me by the beautiful work reported here by Dr. Hydén. Let us take the neuron nucleus as the universe under consideration. The organizational increase can be quite formidable, as can be seen from Equation (9), which sets the absolute rate of maximum uncertainty proportional to the percentage rate of increase in the number of ordered elements. Since there are only 80,000 molecules in the nucleus, only a couple of thousand molecules, modified according to Dr. Hyden's ingenious mechanisms, may add substantially to the Ω of the system.

These have been somewhat general remarks about the use of numbers in describing systems in various states of order, expressing an amount of uncertainty, complexity, or "perplexity," as one of my students suggested, associated with these systems and, finally, the amount of information that is required to specify them. I will return now to the topic of our Conference where these numbers may become useful.

At issue is an important property of the functioning of our nervous system. We call it "memory." In looking for mechanisms that can be made responsible for this property, I strongly suggested that we not look upon this system as if it were a recording device. Instead, I have proposed looking at this system as if it were a computer whose internal organization changes as a result of its interaction with an environment that possesses some order. The changes of the internal organization of this computer take place in such a way that some constraints in the environment which are responsible for its orderliness are mapped into the computer's structure. This homomorphism "environment-system" reveals itself as "memory" and permits the system to function as an inductive inference computer. States of the environment which are, so to say, "incompatible with the laws of nature," are also incompatible with output-states of the computer.

I am going to apply now the numerology of information theory to some of the known features of the nervous system, and we shall see what kind of conclusion can be derived from these numbers. The first number I am going to derive is an estimate of the amount of information necessary to specify a "brain." As I pointed out earlier, in order to make any progress in making such estimates, one has first to specify the "universe"—in this case it is the brain—in terms of a finite number of states and the probability of their occurrence. If this is done, H (Brain) can be calculated from Equation (7). To this end, I suggest interpreting "brain" as a set of finite number of elements, the neurons, which are interconnected to each other in some fashion, forming a huge network. I propose to make these connections "directional" by putting imaginary arrows on the connection lines in order to suggest unidirectionality of the propagation of impluses along the axons away from the cell body. The universe under consideration is, then, all possible networks that can be formed by connecting the elements, and a particular state of this universe is a particular network. I have now to estimate the number of states of this universe, in other words, the number of different networks that can be composed by directionally connecting n elements. The question as to what differentiates one network from another can be approached from two different points. One approach is a purely structural one, where the operational modalities of the nodal elements are ignored; the other approach takes these modalities into consideration. I suggest looking first into the purely structural features, and only later into the possible operations that are carried out by each neuron in a structurally defined network.

The problem of counting the number of nets which can be formed by directionally connecting n elements is solved easily with the aid of a connection matrix. This is a square matrix of n rows and n columns labeled according to the label of each element (Fig. 95). If element E_i is connected to element E_j, a "one" is inserted into the i-th row at the intersection of the j-th column. Otherwise, a "zero" is inserted. Thus, the particular way in which "ones" and "zeros" appear in the matrix uniquely determines the corresponding network. Hence $\mathcal{N}$, the number of ways in which "ones" and "zeros" can be distributed over the n^2 entries of the matrix, is also the number of different nets that can be constructed by directionally connecting n elements. With two choices at each entry, this number is

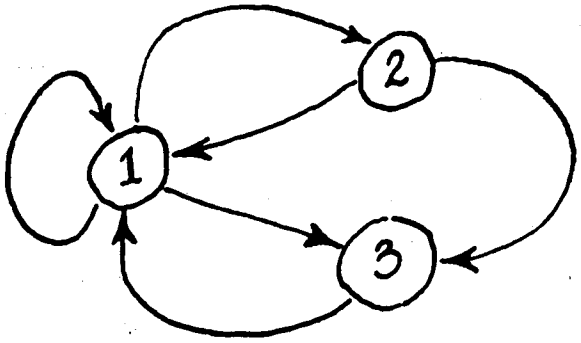
$$\mathcal{N} = 2^{n^2}$$

Since my ignorance is complete in regard to the question of whether one net is more probable than another, my universe is populated by equiprobable states, and, hence, I associate with it an uncertainty:

$$H = \log_2 2^{n^2} = n^2 \text{ bits/net.}$$

In other words, n^2 bits of information are necessary to specify a particular network, as could have been seen directly from the connection matrix where n^2 binary choices had to be made when putting "ones" or "zeros" into the n^2 entries in order to specify a particular network.

Estimates of the number of neurons in a human brain center around 10 billion. Consequently,



NET

	1	2	3
1	1	1	1
2	1	0	1
3	1	0	0

MATRIX

Figure 95. Representations of networks: (a) Graph; (b) Matrix. These representations are equivalent.

$$H = (10^{10})^2 = 10^{20} \text{ bits/brain.}$$

Let us see whether or not the information that is needed to specify just the connection structure of the nervous system—not to speak of the specifications of the operational modalities of its elements—can be genetically determined: Fortunately, there exist good estimates for the information content of the genetic program. The most careful one, I believe, is still the one made by Dancoff and Quastler (3). From various considerations, they arrived at an upper and lower limit for the amount of uncertainty H_G in a single zygote:

$$10^5 < H_G < 10^{12} .$$

In other words, the program that is supposed to define the structure is, by a factor of, say, 10^{10} off the required magnitude! This clearly indicates that the genetic code, which determines far more than just the nervous system, is incapable of programming nets of the unrestricted generality, as considered before. One way out of this dilemma is to assume that, de facto, only an extraordinary small amount in the structure of the nervous system is genetically specified, while the overwhelming portion is left to chance. Although the idea of leaving some space to chance-connections is not to be rejected entirely, it does not seem right to assume that only one hundredth of 1%, or less, of all neurons have specified connectivities. This assumption, for instance, would make neuroanatomy impossible, because the differences in brains would far outweigh their likenesses.

Another way out of this dilemma is to assume that the genetic code is indeed capable of programming a large variety of networks, each of which, however, involves only a small number of neurons, and each of which is repeated in parallel over and over again. Repetition of a particular structure in parallel requires very little information indeed, the only command being: "Repeat this operation until stop." The various kinds of networks may then be stacked in the form of a cascade (Fig. 96). In a very crude way, the appeal of this picture is that there is some resemblance to the various laminate structures that are observed in the distribution of neurons in the outer folds of the brain. Let us see now what numbers we obtain if we assume that the whole system of parallel networks in cascade is specified by the genetic program.

I propose to consider a small elementary net that involves only $2n$ neurons, half of which are located in one layer, say L_1, and the other half in an adjacent layer L_2 (Fig. 96). The axons of neurons in L_1 contact those in L_2, but there are no return pathways assumed in this simple model. Since the total number of neurons located in each layer is supposed to be large, say N, the complete connection scheme for the two layers is established by shifting the elementary network parallel to itself in both directions along the surface of the layers. The number of parallel networks is, thus,

$$P = \frac{N}{n} .$$

Again, a connection matrix for the elementary network can be

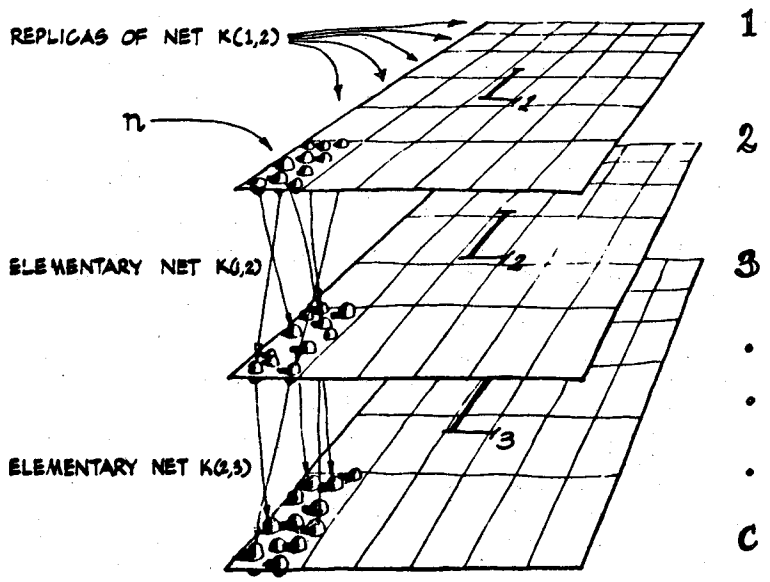


Figure 96. Cascade of C unlike networks composed of P like nets in parallel.

drawn with n rows and n columns, corresponding to neurons in layers L_1 and L_2, respectively, where, at the intersection of a row with a column, the absence or presence of a connection between the appropriate neuron in L_1 with a neuron in L_2 is indicated by a "zero" or a "one." Consequently, the number of nets is again 2^{n^2}, hence, the uncertainty of this elementary network is

$$H_n = n^2.$$

Although P such nets are working in parallel between layers L_1 and L_2, the uncertainty for the whole network connecting these two layers is still only n^2, because there is no freedom for even a single connection in any one of the P networks to change without the corresponding changes in all other nets, since their connectivity is determined by the connection matrix which functions as a genetic mold from which all P nets are cast.

I propose to assume that a different connection matrix controls the connections between the next pair of layers (L_2, L_3), and so on, in the cascade of C layers. The uncertainty of the system as a whole is, therefore,

$$H_S = n^2 C$$

and is assumed to be specified by the genetic code. Hence:

$$H_S = H_G$$

On the other hand, we have to accommodate in the whole system a totality of $\overline{N}$ neurons which are distributed among C layers of nP neurons each:

$$\overline{N} = nPC.$$

Eliminating n, the number of cells in an elementary network, from the two equations above, we obtain a relation between the number of cascades and the number of parallel channels in each cascade

$$H_G = \frac{\overline{N}^2}{CP^2},$$

or

$$P = \frac{\overline{N}}{H_G} \cdot \frac{1}{C}.$$

TABLE 18

H_G bits								
10^6			10^8			10^{10}		
C	P	n	C	P	n	C	P	n
10^2	1.10^6	100	10^2	1.10^5	1000	10^2	1.10^4	10^4
10^3	3.10^5	30	10^3	3.10^4	300	10^3	3.10^3	3000
10^4	1.10^5	10	10^4	1.10^4	100	10^4	1.10^3	1000
10^5	3.10^4	3	10^5	3.10^3	30	10^5	3.10^2	300
10^6	1.10^4	1	10^6	1.10^3	10	10^6	1.10^2	100

Table 18 gives for three reasonable values of the genetic information H_G, a set of five values for triplets C, P, n, which satisfy the above equation.

Among the various choices that are given in Table 18, it seems to me that, for an assumed genetic information of 10^8 bits/zygote, a system that, on the average, consists of 1,000 layers ($C = 10^3$), each layer incorporating 30,000 parallel elementary networks ($P = 3.10^4$) which involve 300 neurons each, is, in the crudest sense, a structural sketch of cortical organization which may, perhaps, not be dismissed immediately for being completely out of the question, quantitatively.

Although this picture is extremely crude, the merit of it—if there is any merit at all—lies purely in its way of suggesting, among the vast amount of possible features, certain ones that may deserve closer inspection.

I would now like to point out a few implications as they seem to me of relevance to our topic. First, the possibility of parallel channels permits us to deal with relatively small nets for which an adequate theory may eventually be devised. I shall report briefly in a moment on the present state of affairs in the theory of small computing nets. Second, a network that consists of periodic repetitions of one and the same elementary network computes on its stimuli the same functions, irrespective of a linear translation of the stimulus distribution. Hence, parallelism implies translational invariance.

I have just referred to the elementary network as "computing nets," and I owe you an explanation for why this may be an appropriate term. At the same time, we shall see what these networks compute, what their computational possibilities are as constituents of large parallel nets, and, finally, how they may modify their operational modalities as a consequence of the results of earlier computations.

Since 1958, at the University of Illinois, we have been looking at the computational possibilities of periodic networks. We have been encouraged in our activity by the various findings of Lettvin (8), Maturana (9), Mountcastle (10), Hubel (5, 6), and others concerning the computation of abstracts in small nets arranged in parallel. The principle idea (1, 7) is to associate geometrical concepts with the connection scheme that prevails between two layers which are the loci of evenly distributed computational elements (Fig. 97).

Assume that from a small area, dA, located at $\vec{r}_1$ in layer L_1, fibers descend in all directions to synapse with elements located in layer L_2. Consider the bundle that synapses with elements located at r_2 in L_2. Along this bundle a certain fraction, λ, of the activity, $\sigma(\vec{r}_1)\,dA$, that prevails in the vicinity of r_1, is passed along and elicits an infinitesimal response, $d\rho(\vec{r}_2)$, in the elements located in the vicinity of $\vec{r}_2$ in L_2. Let this response be proportional to the stimulus activity in $\vec{r}_1$:

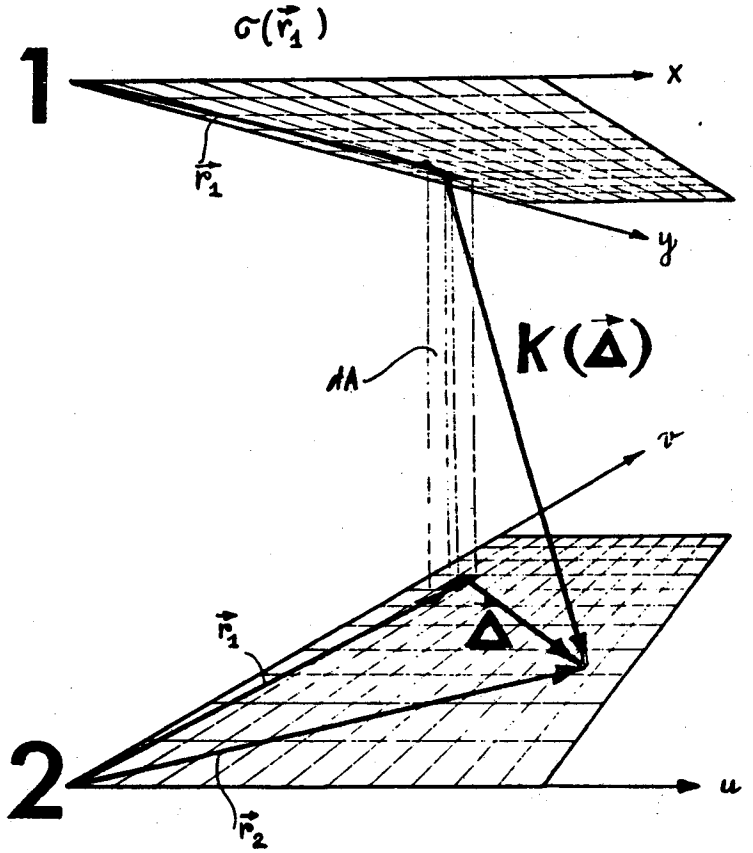


Figure 97. Geometrical relationships in an action network.

$$d\rho(\vec{r}_2) = \lambda K(\vec{r}_1, \vec{r}_2)\,\sigma(\vec{r}_1)\,dA,$$

with $K(\vec{r}_1, \vec{r}_2)$ the proportionality "constant," which may have different values from point to point in the stimulus layer L_1 as well as in the response layer L_2. This is suggested by letting K be a function

of the loci $\vec{r}_1$ and $\vec{r}_2$. Since K is uniquely associated with the bundle of fibers that connect elements at $\vec{r}_1$ with those at $\vec{r}_2$, we have in K the parameter that represents the action of the elementary network which connects the two layers L_1 and L_2. From here on, I shall refer to K as the "action function" of the elementary network. This network is supposed to repeat itself with periodicity, say p, in both directions, hence

$$K\ (x + ip,\ y + jp;\ u + ip,\ v + jp) = K\ (x, y; u, v)$$
$$i, j = 0; \pm 1; \pm 2; \pm 3; \ldots\ldots$$

and, consequently, K is a function of only the distance

$$\vec{\Delta} = \vec{r}_2 - \vec{r}_1$$

between the two points under consideration:

$$K\ (\vec{r}_1, \vec{r}_2) = (\vec{r}_2 - \vec{r}_1) = K\ (\vec{\Delta})\ .$$

The response at $\vec{r}_2$ that is elicited by the contribution from $\vec{r}_1$ is, of course only a fraction of the response elicited at $\vec{r}_2$. In order to obtain the total response at $\vec{r}_2$ we have to add the elementary contributions from all regions in the stimulus layer:

$$\rho(\vec{r}_2) = \int_{L_1} K\ (\vec{\Delta})\ \ \sigma(\vec{r}_1)\ dA. \qquad (13)$$

If the action function $K\ (\vec{\Delta})$ is specified—and my suggestion was that it is specified by the genetic program—then, for a given stimulus distribution, the response distribution is determined by the above expression. The physiological significance of the action function may best be illuminated by breaking this function up into a product of two parts:

$$K\ (\vec{\Delta}) = D\ (\vec{\Delta})\ .\ T\ (\vec{\Delta})\ , \qquad (14)$$

where $D\ (\vec{\Delta})$ represents a change in the density of fibers that arise in $\vec{r}_1$ and converge in $\vec{r}_2$. Hence, $D\ (\vec{\Delta})$ is a structural parameter. $T\ (\vec{\Delta})$, on the other hand, describes the local transfer function for fibers that arise in $\vec{r}_1$ and synapse with neurons in the vicinity $\vec{r}_2$. Hence, $T\ (\vec{\Delta})$ is a functional parameter.

I hope that I have not unduly delayed an account of what is computed by these nets. Unfortunately, I cannot give a detailed account of the various computational results that can be obtained by considering various action functions, K. Let me go only so far as to say that the results of these computations are invariants, or abstracts, of the stimulus distribution. I have already discussed invariance against translation as the prime bonus of using networks in parallel. Furthermore, it may not be too difficult to imagine the kind of abstracts that are computed if the action function, K, possesses certain symmetry properties. Consider, for instance, the three fundamental types of symmetry to hold for three types of action functions: the symmetric, the antisymmetric and the circular symmetric action function defined as follows:

$$K_s\ (\vec{\Delta}) = K_s\ (-\vec{\Delta}),$$

$$K_a\ (\vec{\Delta}) = -K_a\ (-\vec{\Delta}),$$

$$K_c\ (\vec{\Delta}) = K_c\ (|\Delta|).$$

Clearly, K_s gives invariance to reversals of stimuli symmetric to axes $y = 0$ and $x = 0$, that is, a figure 3 into Ɛ, or M into W; while K_a gives invariance to reversals of stimuli symmetric to lines $y = x$, that is, ~ into S, and > into V. Finally, K_c gives invariance to rotations, that is, almost all previous reversals plus N into Z.

Maybe you have recognized in some of the properties of these action functions a resemblance with properties Hubel and others have observed in the response pattern of what they called the "receptor field." There is indeed a very close relation between these two concepts, because knowing the "domain" of the action function, that is, the cells in the target layer that are "seen" by a single cell in the source layer, enables one at once to establish the domain of the "receptor function." That is, the cells in the source layer that are seen by a single cell in the target layer. Let $G(\vec{\Delta})$ be the receptor function, then we have

$$G_s = K_s;\ G_a = -K_a;\ G_c = K_c\ .$$

At this stage of my presentation it may be argued that all this has to do with now well-established filter operations in the nervous system, but what has it to do with memory? It is true that in the foregoing I have indeed attempted to suggest a rigorous framework in which we can discuss these filter operations. But what if they arise from interaction with the environment? How would we interpret the computation

of an invariant if it results from past experience, or if it is the result of some learning process? I venture to say that we would interpret such adaptive computations of abstracts precisely as the functioning of a "memory" which responds with a brief categorization when interrogated with a flood of information. My remembering Karl Pribram cannot consist of interrogating a past record of all attitudes, gestures, etc., that I have stored of him. The probability that I find among these the ones with which I am presented at this moment is almost nil. Most probably, they are not even there. What I have built in, I believe, is a net of computers that compute on the vast input information which is pumped through my visual system all that is "Pribramish" — that is all categories that define him and only him — and come up with a terse name for all these categories which is "Karl Pribram."

In order to support this thesis, my task is now to show that it is indeed possible to teach the kind of computing networks which I have discussed to shift their computational habits from the computation of one type of abstracts to the computation of other abstracts. In the language of the theory of these computing networks, my task is now to show that the action function K, which uniquely determines the computed invariant, is not necessarily an unalterable entity, but can vary under the influence of various agents. In other words, the temporal variation of K does not necessarily vanish:

$$\delta K \neq 0 .$$

Before I go into the details of my demonstration, I have to confess that, to my knowledge, we are today still far from a satisfying theory of adaptive abstracting networks. The kind of mathematical problems which are soon encountered are of fundamental nature and there is as yet little help to be found in the literature. Hence, I will not be able to present spectacular results today. On the other hand, I hope that the following two simple examples will be sufficiently explicit as to indicate the main line of approach.

In my brief outline of some features of parallel networks, I suggested that under certain conditions the genetic program may suffice to specify all networks in the system. In my terminology, a network connecting two layers is specified if the action function, $K(\vec{\Delta})$, for the elementary network is defined. The whole network is then generated by simply shifting the elementary network along all points of the confining layers. In my following discussion of variable action functions, I still propose to think of a genetically programmed action function, say, $K_0(\vec{\Delta})$; but now I consider this action function to be subjected to some perturbation. Due to my earlier suggestion [Equation (14)] that we think about this action function as being composed of two parts, a structural part $D(\vec{\Delta})$ and a functional part $T(\vec{\Delta})$,

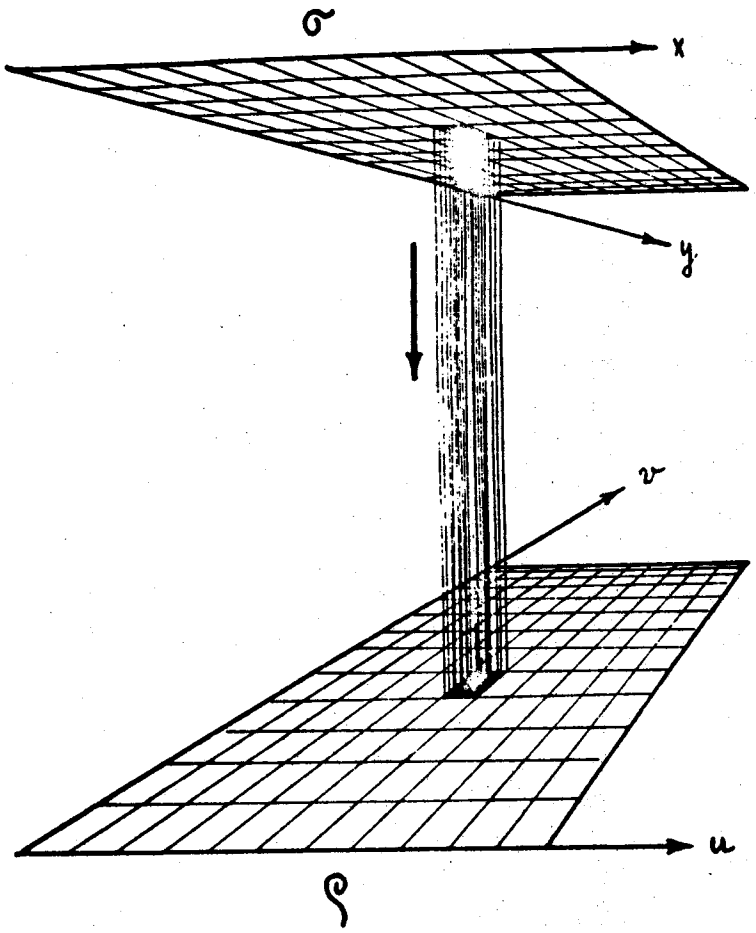


Figure 98. Geometrical relationships in an action function that is defined by an ideal one-to-one mapping.

we may now consider the perturbation to act on either T or D, or on both. Although I realize that the assumption of structural variations in neural nets is not too popular, I will give as my first example the results of just such a structural perturbation which you may, perhaps, accept as a possibility which prevails during the early phases of the formation of these networks.

Assume that the simplest of all possible elementary networks is being programmed genetically (Fig. 96). It is a net consisting of precisely two elements, one in layer L_1, the other in layer L_2. Since $2n = 2$, we have $n = 1$, and the uncertainty for this net is

$$H = 1 \text{ bit/net.}$$

The genetic program says: "Connect these two elements!" In mathematical terms, the mapping function, $D(\vec{\Delta})$, that represents the structure of the net can be expressed by the Dirac delta function

$$D(\vec{\Delta}) = \delta(|\Delta|),$$

where $$\delta(x - x_0) = \begin{cases} 0 \text{ for } x \neq x_0 \\ \infty \text{ for } x = x_0 \end{cases},$$

and $$\int_{-\infty}^{+\infty} \delta(x - x_0)\, dx = 1.$$

For simplicity, let us assume the transfer function $T(\vec{\Delta})$ to be just a constant:

$$T(\vec{\Delta}) = a .$$

With these, the action function is simply

$$K(\vec{\Delta}) = a\,\delta(|\Delta|),$$

and the response in layer L_2 for a given stimulus in L_1 is after Equation (13):

$$\rho(\vec{r}_2) = a\lambda \int_{L_1} \delta(|\Delta|)\, \sigma(\vec{r}_1)\, dA = a\lambda\sigma(\vec{r}_2) .$$

In other words, the response is an identical replica of the stimulus with modified amplitude due to the factor $a\lambda$, as was to be expected by this simple connection scheme.

Assume now that during the process of the realization of this network, it is impossible for the fibers descending from L_1 to make appropriate contacts in L_2, and suffer random deviations due to the presence of the intermediate glia cells. A fiber bundle leaving at $\vec{r}_1$, and destined to arrive at $\vec{r}_2$, will be scattered according to a Gaussian

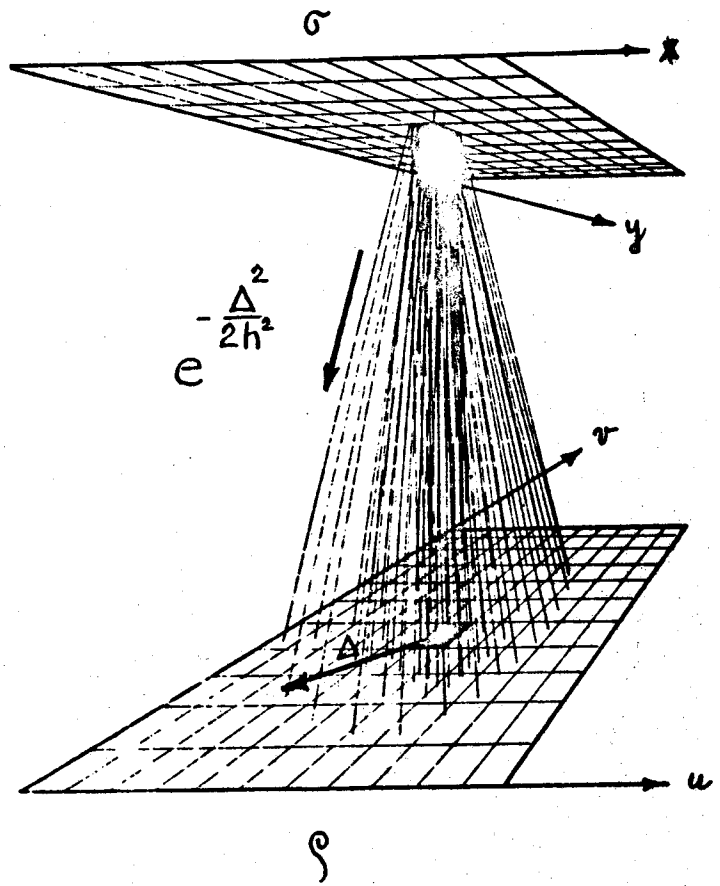


Figure 99. Geometrical relationships in an action function that arises from a random perturbation of an ideal one-to-one mapping.

distribution. Hence, the mapping function becomes:

$$D(\vec{\Delta}) \propto \exp[-\Delta^2/2h^2] ,$$

with h representing the variance of this distribution (Fig. 99).

Assume, furthermore, that the transfer functions for inhibitory and excitation connections are constants, but different for the two kinds,

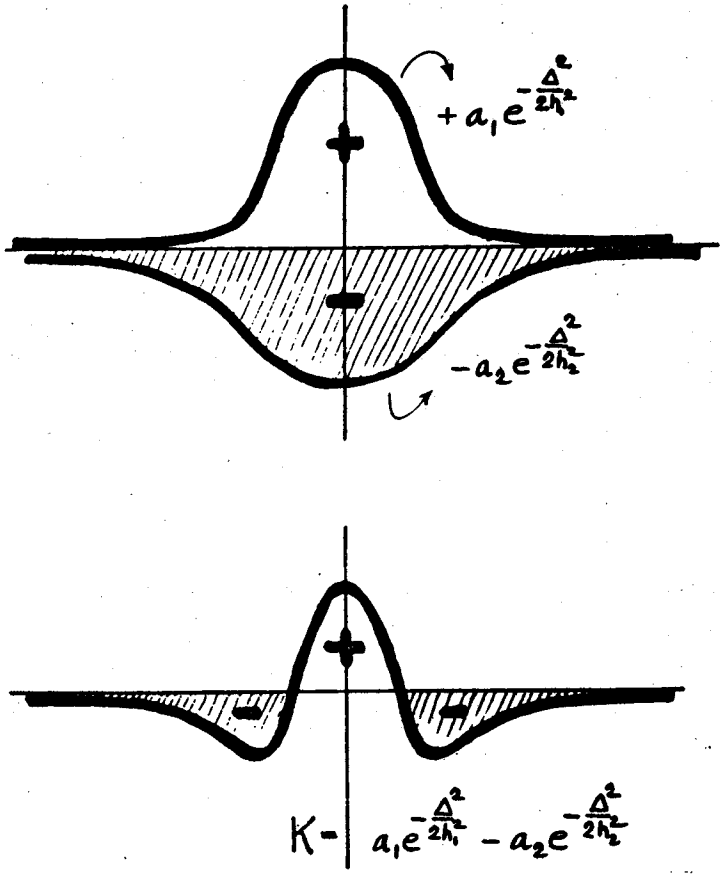


Figure 100. Graphical representation of a Gaussian (Normal) distributed action function. (a) Excitatory and inhibitory distribution separated. (b) Composite of excitatory and inhibitory distribution function.

$(+a_1)$ and $(-a_2)$. Introducing corresponding variances, h_1 and h_2, the action function of the elementary network becomes:

$$K(\vec{\Delta}) = a_1 \exp\left[-\Delta^2/2h_1^2\right] - a_2 \exp\left[-\Delta^2/2h_2^2\right],$$

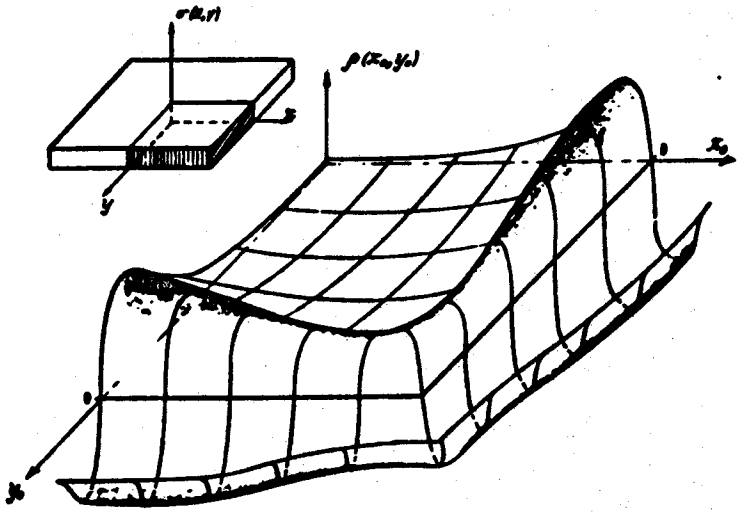


Figure 101. Response distribution elicited by a uniform stimulus confined to a square. Contour detection is the consequence of an action function obtained by superposition of an excitatory and inhibitory Gaussian distribution.

and the stimulus-response relation can be established with Equation (13). For uniform stimulus, $\sigma = \sigma_0$, the resulting response is also uniform:

$$\rho(\vec{r}_2) = \sigma_0 \lambda \quad (a_1 h_1^2 - a_2 h_2^2) = \text{const.}, \qquad (15)$$

and vanishes for

$$a_1 h_1^2 = a_2 h_2^2 . \qquad (16)$$

This condition is assumed in the graphical representation of the Gaussian action function in Fig. 100, and also in Fig. 101, which shows the response activity in layer L_2, if the stimulus pattern is a uniformly illuminated square. The interesting feature in the composite picture of Fig. 100 as a pronounced lateral inhibition of the fiber bundles acting on neurons located in L_2. The consequence of this action function is a computation of a contour by the elements in L_2. This contour is most conspicuously present if uniform stimulation elicits no response, or, in other words, if the condition expressed in Equation (16) is fulfilled.

Since there was nothing in our assumed program that could have specified such an exacting condtion, adaptation is the only mechanism that can be made responsible to accomplish this. In other words, the action function that has evolved so far has to be submitted to further changes, if the ideal connection conditions expressed in Equation (16) have not perchance established themselves spontaneously. This chance, however, is extremely small and, after formation, we may expect the network to be in a state where the equality sign in Equation (16) should be replaced by an inequality sign.

The question now arises as to what is going to change, how is the change to be accomplished, and what is the cause of this change? If I understood correctly some of the remarks that have been made during this meeting, then I am not permitted to change the structure of the network as it has evolved so far, because after the synaptic connections are established they are rigidly maintained. The only escape which is now left to me is to change $T(\vec{\Delta})$, the transfer function, or, as I referred to it earlier, the operational modality of the nodal element in this network.

ROBERTS: Why are you not permitted to change structure?

VON FOERSTER: I am just following the suggestions that were given earlier, I believe, which do not permit me to move synapses around. They are fixed.

SPERRY: But you can't do that because, earlier in the Conference, at least from the point of view of chemistry and microbiology, I didn't hear any loud objections when we were discussing the idea that there could be structural changes at synapses.

VON FOERSTER: Yes, but in my terminology, these would be functional changes of the transfer function, caused by some sub-microscopic changes in the synaptic structure. I don't know whether they can be seen in the microscope. Let me just add that I would be very happy indeed if I were allowed to make changes in the network structure. But what I am trying to say is that, even if no structural changes are permitted in my networks, I am still able to devise a computer that changes its internal organization, because I can now change the transfer function of my elementary components. I believe this is what Sir John was showing when he pointed out the variation in the efficacy of synaptic transmission.

Going back to my action function, I propose now not to change the structural part, $D(\vec{\Delta})$, which established itself as a random distribution of connecting fibers, and to permit the local transfer function, $T(\Delta)$, to be submitted to some perturbation. The transfer function, as I introduced it earlier, was of the most simple kind. It consisted of just two constants, $(+a_1)$ and $(-a_2)$, for excitation and inhibition, respectively. Assume that, after the contacts have been established, these two constants have some initial values, say, $(+a_{10})$ and $(-a_{20})$. However, the crucial condition for optimum contour extraction (Equation 16) is not initially fulfilled:

$$a_{10}h_1^2 \neq a_{20}h_2^2 .$$

Since in this equation the variances of inhibitory and excitory connections are structual parameters (and according to my rules they are tabu), I have to let something operate on either a_{10} or a_{20} in order to obtain the desired equality. Among the many possibilities that offer themselves from a purely hypothetical point of view, let me submit one that has, for me, at least, the ring of plausibility. Let, for example, the left-hand side in the inequality exceed the right-hand side. I shall now assume that at least one of these transfer functions, in this case, a_2, the inhibitory transfer function, will increase its efficacy due to the overall activity of the responding network. In other words, I don't make this feedback loop a local affair, but rather an affair in which the activity of whole cell complexes, plus their surrounding tissue, may be involved. Formulated precisely, I have

$$\frac{\partial a_2}{\partial t} = cR,$$

where R stands for the total response of the layer L_2:

$$R = \int_{L_2} \rho(\vec{r}_2)\, dA .$$

In order to get a rough picture of what happens under these conditions, let the universe of our system be a dull one, with occasional appearances of objects that are "seen" by the elements in L_1, but otherwise is uniformly illuminated. In this case, R can be evaluated at once with the aid of Equation (15):

$$R = (a_{10}h_1^2 - a_2h_2^2) \cdot S(t) ,$$

with $S(t)$ the total stimulation applied to layer L_1 as a function of time. Inserting this into Equation (17), one obtains a differential equation in a_2:

$$\frac{\partial a_2}{\partial t} = c\lambda (a_{10}h_1^2 - a_2h_2^2) \cdot S(t) ,$$

with the initial condition

$$t = 0 \ldots\ldots a_2 = a_{20}$$

and its solution:

$$a_2h_2^2 = a_{10}h_1^2 - (a_{10}h_1^2 - a_{20}h_2^2)\ \exp(-h_2^2c\ \lambda \int_0^t S(t)\,dt)$$

The first term represents the desired condition as expressed in Equation (16), while the second term, due to the decaying exponential eventually must vanish, whatever the history of the stimulus $\int_0^t S dt$ might have been.

Although this is admittedly an almost trivial example of an adaptive network, it enabled me, without too much acrobatics, to get my point across that a computer originally conceived as an ideal repeater, when exposed to the roughness of a real world, transforms itself to a perfect contour detector. This is the more surprising if one considers for a moment that this machine of considerable sophistication grew from a genetic specification that required, as you may remember, only one single bit of information!

With this example, I could retire now if another argument would not loom in the back of my mind: That, again, I could be accused of having merely presented the story of an adaptive filter mechanism, without having even touched the profound problem of memory.

Fortunately, in this dilemma a charming, but not immediately obvious, property of the Gaussian action function comes to my rescue. It turns out that if the condition in Equation (16) is fulfilled, in the vicinity of points $\vec{r}_2$ along the contour in the response layer L_2, this connection scheme produces a mean response density, $\vec{\rho}(\vec{r}_2)$, that is proportional to the curvature of the contour. Take $\underline{R}\ (\vec{r}_2)$ to be the radius of curvature of the contour at point $\vec{r}_2$, and k to be a constant, then

$$\vec{\rho}(\vec{r}_2) = k\ /\ \underline{R}\ (\vec{r}_2)\ .$$

In the first approximation, the total response, R, is the activity integrated over the whole contour

$$R = \int \vec{\rho}(\vec{r}_2)\ ds = k \int \frac{ds}{\underline{R}\ (\vec{r}_2)}\ ,$$

where ds is a line-element along the contour. But

$$ds = \underline{R}\ (\vec{r}_2)\ d\psi\ ,$$

with ψ representing the polar angle from the center of curvature.

Hence,

$$R = k \int_{\psi_1}^{\psi_2} \frac{\underline{R}\ (\vec{r}_2)\ d\psi}{\underline{R}\ (\vec{r}_2)} = k \int_{\psi_1}^{\psi_2} d\psi = \text{size invariant.}$$

In other words, the only function that carries the information about the size of the object, namely $\underline{R}$, cancels out, and the resulting response, R, is invariant with respect to the size of different objects and varies only with their shape.

Imagine now that our system is in contact with a universe which has the peculiar property of being populated by a fixed number of objects of various sizes, but of similar shapes, that move about freely, so that the limited "visual field" of our system perceives at any instant only a certain fraction of the number of objects. Since each object seen will elicit the same response, R_o, Q objects in the visual field will elicit a total response of $Q \cdot R_o$. In other words, our system is able to count! Since nothing of this sort was a property of the system when it was born — we are still dealing with only one bit of genetic information — it must have acquired its mathematical abilities as a result of interactions with its environment. Moreover, one may contrive a large variety of educational methods, that accelerate the process by which the system acquires its knowledge about numbers. When it finally masters this task, a tutor, ignorant about our system's internal workings may speculate about the location of its memory. We, of course, know that it is all over the place, realized in the structure of the connection scheme and in the operational modalities of all nodal elements of this network.

I shall now conclude my remarks by again giving a few numbers. First, let me take our tutor, who does not know the simple evolutionary history of our system, and who wants to study its anatomy, he will find an extraordinary complicated network consisting of n elementary nets, each different from any other. Applying good statistics, he can summarize his data by showing that the connections of all elementary nets have a Gaussian distribution, each network involving on the average, say, m contacted elements. The uncertainty H_e for a single elementary net is approximately*

$$H_e = \log_2 \sqrt{m}\ .$$

*Due to the constraint present in a normal probability distribution, the uncertainty is not $\log_2 m$, but approximately only $\frac{1}{2} \log_2 m$.

and, for the network consisting of N elements,

$$H = N \log_2 \sqrt{m}$$

If I use the figures from Table 18, as suggested earlier, the uncertainty for this system, consisting of only one net (C = 1), is:

$$H = 3 \cdot 10^4 \log_2 \sqrt{300} = 1.2 \cdot 10^5 \text{ bits.}$$

The anatomist may attribute this to the genetic program and will estimate its information content to be about 100,000 bits. We, however, know that the original specifications were written only in terms of one single bit. Where does all this information come from? The answer is, of course, from the "noise" that was introduced into the network when it made its feeble attempt to carry out the genetic command. In this case, as in so many others, it is the noise that enriches a structure that was extremely poor to begin with.

Let us turn now to the performance of our system. With N independent binary elements in its "sensory" layer, L_1, which may be ready for a new sensation in a time interval of t_o seconds, the system's rate of input information is

$$H_{in} = N/t_o \text{ bits/second.}$$

Its output rate depends, of course, upon the demands of our system's environment. Assume that the number of objects seen by our animal varies according to a normal distribution, with a variance of, say, M objects. It reports in intervals of t_o seconds about the state of its visual universe. Hence, its output information rate:

$$H_{out} = \frac{1}{t_o} \log_2 \sqrt{M} ,$$

and the ratio between output and input:

$$H_{out} / H_{in} = (\log_2 \sqrt{M}) / N.$$

However, M is always smaller than N, because one cannot see more objects than there are cones and rods. In the most optimistic case, M = N, and the optimum output-input ratio is

$$\left(\frac{H_{out}}{H_{in}} \right)_{opt} = \frac{\log_2 N}{.2N} .$$

Again, for N = 30,000 this ratio becomes:

$$\left(\frac{H_{out}}{H_{in}} \right)_{opt} = 0.00075 .$$

In other words, the system reduces considerably the uncertainty of its environment by its power of classification and abstraction which, in this case, consists of identifying individual objects.

Up to now, I have only stressed the miracles that are obtained when disorder is introduced into an ordered genetic program. One may ask now: Where is the internal organization that arises in my system from interactions with its environment, as I proclaimed in my earlier statements? It is clear that the consequences of such interactions can only be accounted for after the system is capable of interacting at all; but this is only after the network has established itself. The only quantity that changes after that is the inhibitory transfer function, a_2, which, for simplicity, I shall denote by "a" without index. Let a_{min} be the smallest interval detectable in observing this transfer function, then, n, the number of states of this "universe," is determined by the maximum value a_{max} that this function can assume:

$$n = a_{max} / a_{min} .$$

Consequently, H_{max}, the maximum uncertainty, is

$$H_{max} = \log_2 n.$$

With

$$n_o = a_o / a_{max} \ll n ,$$

its initial uncertainty is

$$H_o = \log_2 (n - n_o),$$

that is, the log of the number of states still available, hence, its initial measure of organization is:

$$\Omega_o = 1 - \frac{H_o}{H_m} = \frac{n_o/n}{\log_2 n} .$$

If, after exposure, the system drifts in its mature state over, say, n_1 distinguishable states, its final state of organization is:

$$\Omega_\infty = 1 - \frac{\log_2 n - \log_2 n_1}{\log_2 n} = \frac{\log_2 (n/n_1)}{\log_2 n}$$

Comparison between initial and final state of organization shows that the measure of the final state of organization is greater than that of the initial state of organization, if

$$\log_2 (n/n_1) > (n_o/n) ,$$

or if, approximately,

$$n > n_o + n_1 .$$

But this is always the case. In fact, if the drift around the final equilibrium state is very small indeed, say $n_1 = 1 + \epsilon$:

$$\Omega_\infty = 1 - \left[\epsilon / \log_2 n \right] \approx 1 ,$$

and the system is in almost perfect order.

I hope that these remarks suggest the possibility to recognize learning and memorizing systems as computers, changing their operational modalities as a consequence of interactions with their environment. Their operations change so as to remove more and more uncertainty of the environment, until the output of these systems keeps them in equilibrium with their universe.

REFERENCES

(Starred references have not been verified.)

*1. Babcock, M. L., A. Inselberg, L. Lofgren, H. Von Foerster, P. Weston, and G. W. Zopf, Jr. 1960. Some principles of preorganization in self-organizing systems. Tech. Rep. #2 Contract

Nonr 1834(21). Electrical Engineering Research Lab., Engineering Experiment Station, Univ. of Illinois, Urbana.

2. Brillouin, L. 1962. Science and Information Theory. Academic Press, New York.
3. Dancoff, S. M. and H. Quastler. 1953. The information content and error rate of living things. In H. Quastler (Ed.), Information Theory in Biology. University of Illinois Press, Urbana. Pp. 263-273.
4. Defares, J. G. and I. N. Sneddon. 1961. The mathematics of medicine and biology. Year Book Medical Publ. Inc., Chicago.
5. Hubel, D. H. and T. N. Wiesel. 1959. Receptive fields of single neurons in the cat's striate cortex. J. Physiol. 148: 547-595.
6. Hubel, D. H. and T. N. Wiesel, 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. 160: 106-154.

*7. Inselberg, A. and H. Von Foerster. 1962. Property extraction in linear networks. Tech. Rep. #2 N. S. F. Grant 17414; Electrical Engineering Research Laboratory, Engineering Experiment Station, Univ. of Illinois, Urbana.

*8. Lettvin, J. Y., H. R. Maturana, W. S. McCulloch and W. Pitts. 1959. What the frog's eye tells the frog's brain. Proc. I. R. E. 47: 1940-1951.

9. Maturana, H. R. and S. Frenk. 1963. Directional movement and horizontal edge detectors in the pigeon retina. Science. 142: 977-978.

*10. Mountcastle, V. B. 1963. Abstract in Symposium on information processing in the nervous system. 22nd international congress of physiological science. Excerpta Medica Foundation, Amsterdam, 1, pt. 2, p. 930.

11. Quastler, H. 1958. A primer on information theory. In H. P. Yockey, R. L. Platzman and H. Quastler (Eds.), Symposium on Information Theory in Biology. Pergamon Press, New York. Pp. 3-49.

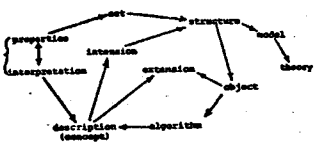
*12. Shannon, C. E. 1951. Prediction and entropy in printed English. The Bell Syst. Tech. J. 30: 50-64.

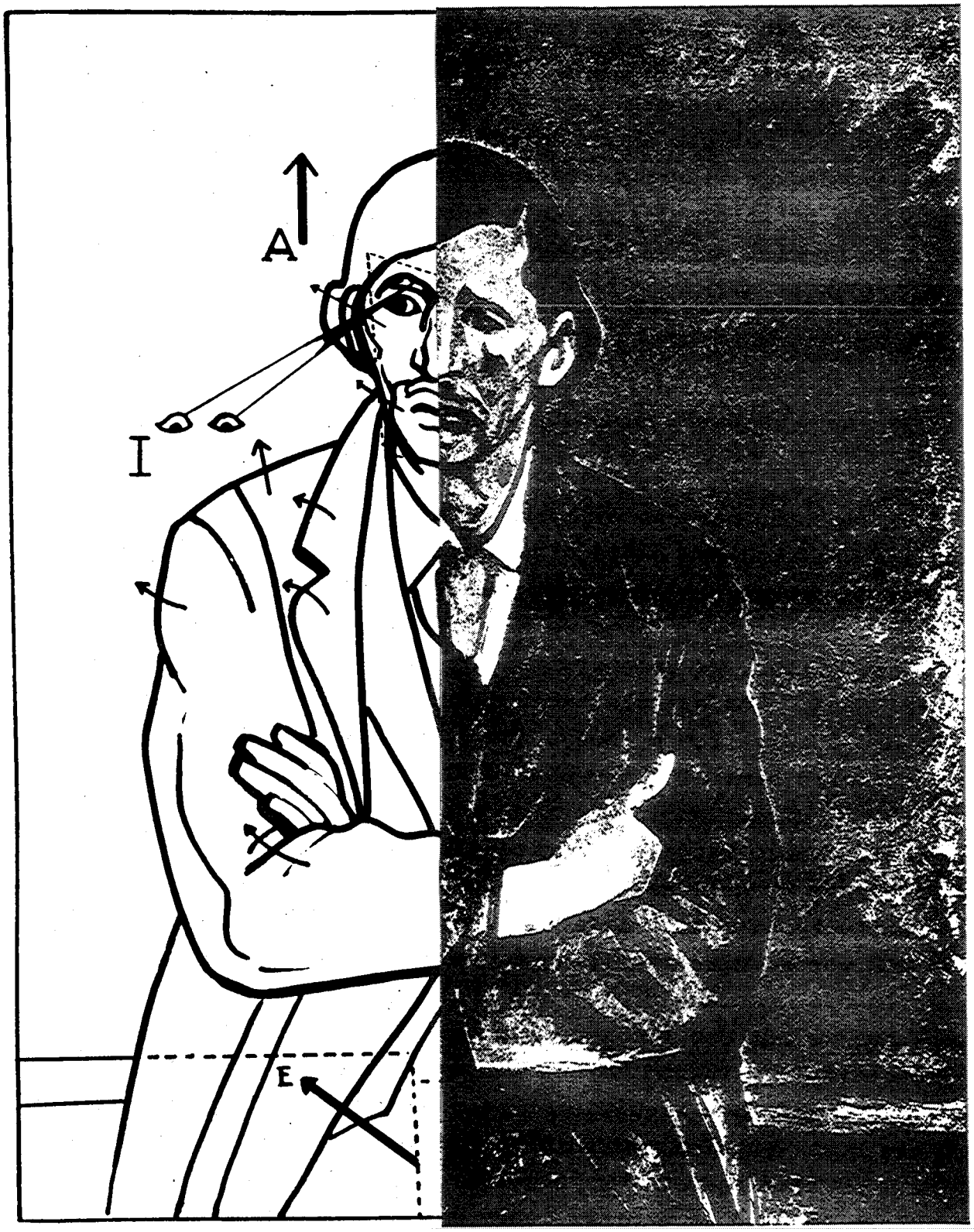
13. Shannon, C. E. and W. Weaver. 1949. The Mathematical Theory of Communication. University of Illinois Press, Urbana.

*14. Von Foerster, H. 1960. On self-organizing systems and their environments. In M. C. Yovits and S. Cameron (Eds.), Self Organizing Systems. Pergamon Press, New York. Pp. 31-50.

MODEL

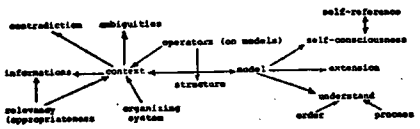
Model of a theory is a structure such that every theorem of the theory holds true in the structure. [L.L.]





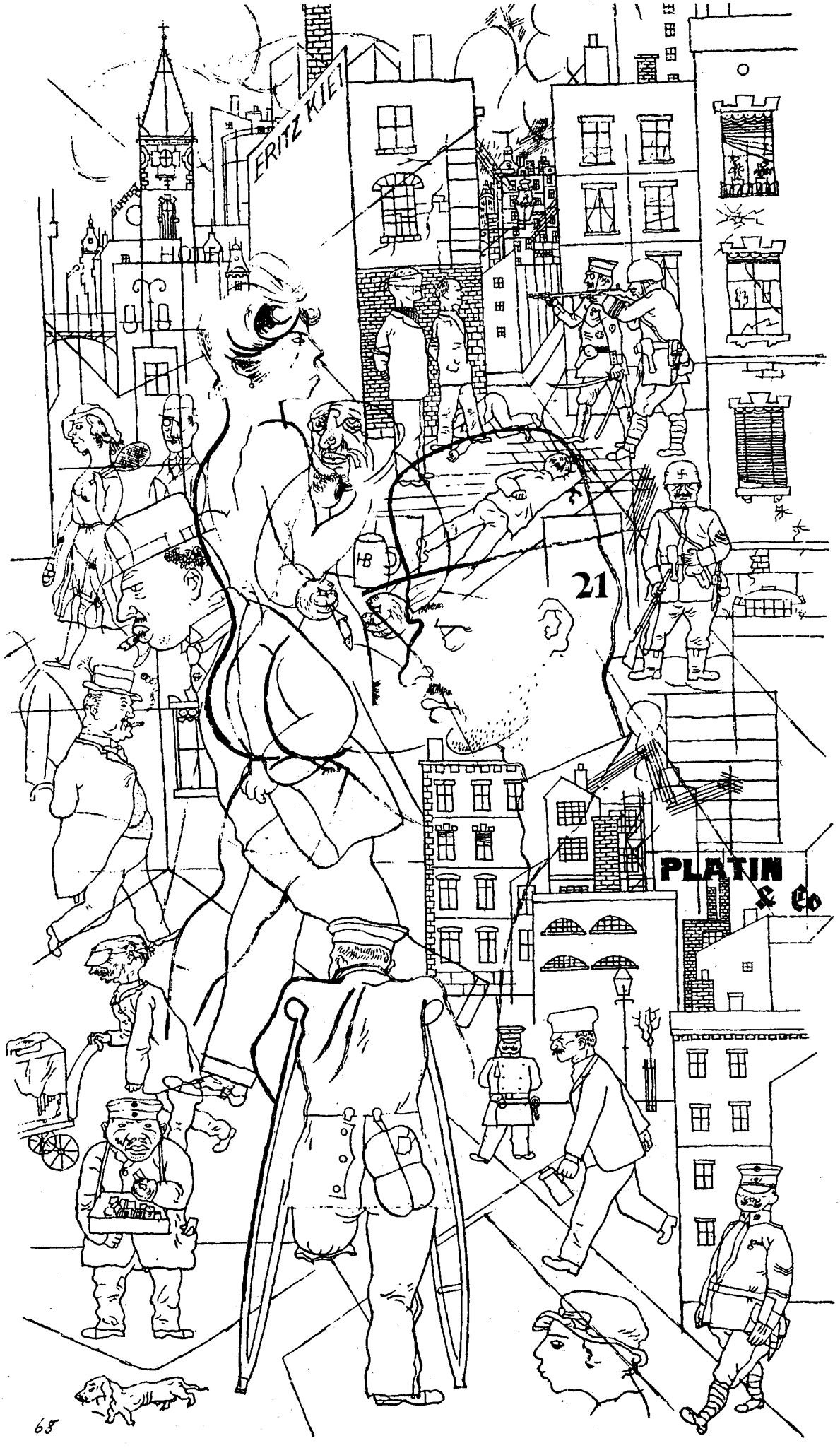
MODEL

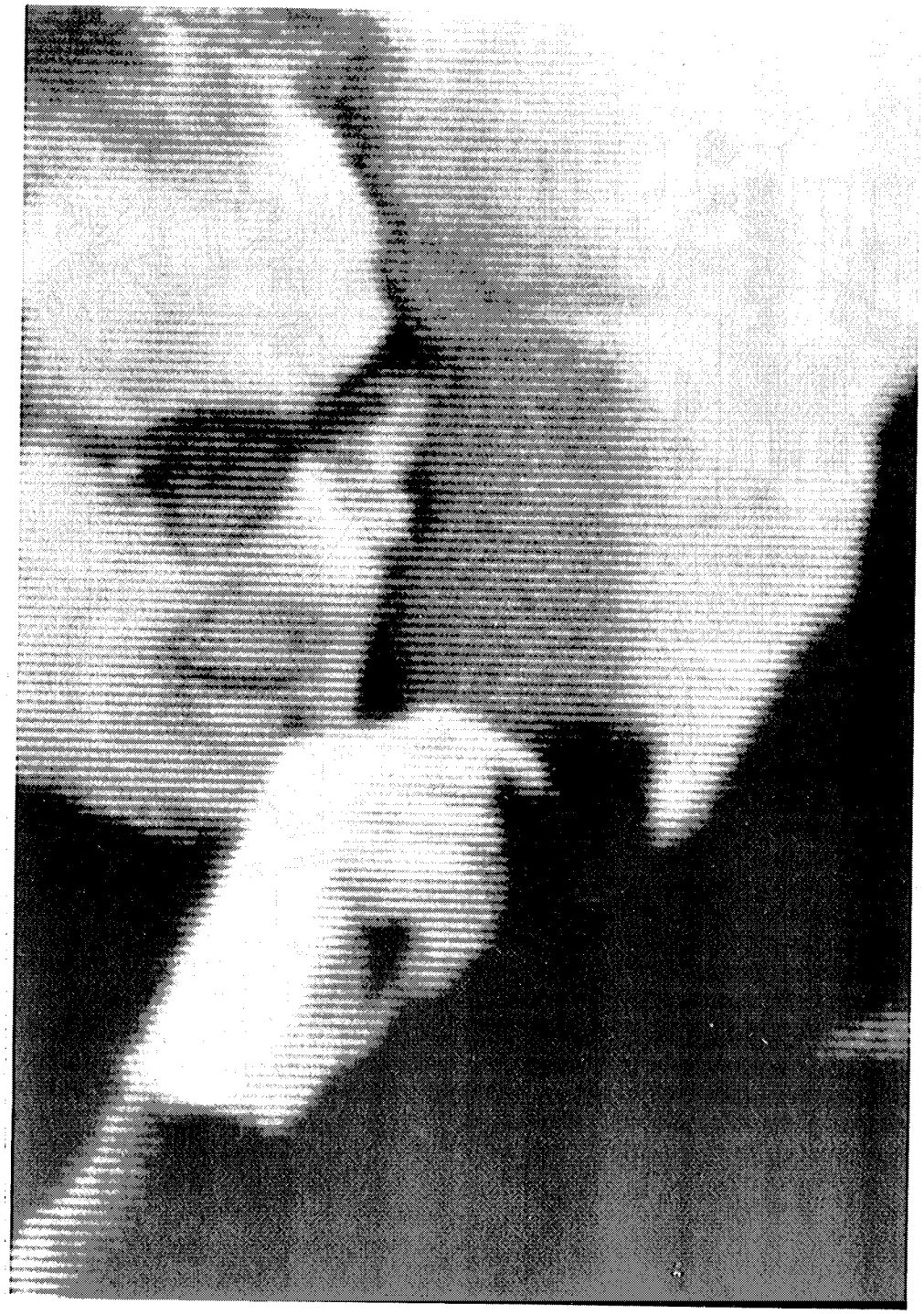
Model is a triadic concept consisting of that which is to be modelled, that which is to perform the modelling, and the context dependent assignment operator making the correspondency.. the model-filter. What this definition captures is the plinth of MODEL which is that a model is not a representation devoid of extension but a dynamic structure with a pulsation similar to that of the sea's edge constantly fitting and refitting to the land creating a shoreline. Said another way, the model is an interaction between an organizing system and the environment it is embedded in via context dependent transmitals of information. This allows models to predict and reinterpret the world. [J.K.]



MODEL

We see that persons express to one another and are impressed by one another on the basis of imaginative activity albeit for concerted activity, we fill out the imagined portion of interaction in similar ways. When we manipulate the time and space of social relations, we demonstrate our ability to imagine as the 'other' does. For example, when I exclude another by pushing my hand, etc., away from my body, the other sees the movement as coming toward him. The movement of the body and its parts are opposite in direction in the fields of perception of the actor and the viewer. That we understand the meaning or intent of gestures verbal as well as non-verbal, only underscores the imaginative activity of social life. We cut space with our arms, legs, eyes and words, circumscribing a place around ourselves in which we include or exclude others and social objects. The observer sees that we act out the sociological insight: personal identity is something within but beyond the skin, created between the subject and some other significant person or object. Time and space are the quasi-physical markers around which our memories and imaginations take shape. The observer is a modeler. If he enters the place created by another, they make a model which is social, its only elements: imaginings, memories, time and space. [P.C.]


ERITZ KIEL
21
PLATIN
& Co



THE NEED OF COGNITION FOR THE COGNITION OF NEEDS

Herbert Brün

I.
Cognitive Processes and Societal Problems

We know enough, today, about social problems and cognitive proceses to make statements:

Social problems are interconnected with cognitive processes. Anyone who attempts to study and to answer the questions posed by one, sooner or later finds himself involved with the study of questions raised by the other.

Answers to so involved a system of interlocked networks of questions will be found only if it is made possible to complement existing theories and conjectures with concrete opportunities for the examination of both, processes and problems, for the observation of their real-time manifestations, and with the still outstanding step from sophisticated documentation toward effective and problem-solving explanation.

It is necessary to recognize that he who sets out to study and to act upon cognitive processes and social problems is himself a member of the set of his objectives. The time-honored distinctions, therefore, between theory, practice, fundamental and applicable research, development etc. no longer hold, when the subjects are cognitive processes and social problems. In fact, all active attitudes, scientific and creative must move simultaneously and together, none emphasized at the expense of the other, each emphatically appropriate at a given moment to a given observation or purpose.

The objective is the understanding of cognitive processes and the solution of social problems. We agree that each one of us approaches the objective differently. We wish to pool our capabilities without having to sacrifice this diversity.

Together with a steadily growing awareness of social problems the definition of the concept "social problem" has broadened considerably. Knowledge about society, acquired by the natural and the social sciences could, and should, be brought to bear on this state of affairs. Of urgency is, therefore, specifically any work, theoretical and practical, which will bridge efficiently the wide gap between idea and implementation.

Current terminology, when referring to all kinds of social discontent, contains words like "alienation", "credibility gap", "communication break-down", "generation gap", "brainwashing", information distortion", etc. These terms, although often correctly describing observed results, do not, in most cases, correctly identify the causes. All, however, converge in accusing a phenomenon which might be called: negative communication. Not communication, but the message keeps breaking down.

Just as positive communication, so is negative communication indicative of social relationships, especially as these realities manifest themselves in language, behavior, conceptual use of words, articulations of opinions, expression of emotions and intuitively preferred positions on issues like conflict, disagreement, fairness, prejudice, rights, privileges, power, poverty, justice, etc.

The content of communicative processes, the message, is generated by, and dependent on, numerous interactions of information with language, with knowledge, with images, with memory, with prediction, and all that and more, both on the sender's and on the receiver's end of the channel-system. Any malfunction in these interactions produces either unintended or undesired messages: negative communication. And negative communication causes, not only in the emotional domain, an increase of malfunctions in interactions, thus producing more unintended and undesired messages.

This vicious circle, which indeed threatens to become a blind alley of life in our society, can be stopped; not however at the level of messages and communication, but rather where the interactions tend to malfunction. Interactions between individuals, groups of people, sectors of society, are less vulnerable to misrepresentations of facts and ensuing disagreements, than to misinterpretations of images and the ensuing cloud of alienation. It is for all participants and co-sufferers in the problem a clash of cognitive structures that had their origin far from any present conflict, and long before it arose.

More knowledge about the origins of cognitive structures in human beings will uncover some vital and hitherto unknown rules of the game called interactions, and it promises, therefore, to be the urgently needed but still missing link that could connect the theories of concerned science with such practice as would effectively benefit a very disturbed, if not even seriously damaged society.

Societal problems appear in every society. The society gives itself and its administrative and executive members the assignment to solve these problems. The problems will, however, not be solved,

when the assignment is given and taken without confidence;

when the assignment is ill-formulated, open to conflicting interpretations, prone to be misconstrued arbitrarily;

when a substantial number of sectors of society, be it in concert or be it in independent simultaneity, claims to be left in doubt as to whether the assignment is taken seriously, is being mishandled, has been refused, or is beyond the capability of the assigned members of society;

when for any one, or all, of such reasons mutual alienation separates precisely those strata of society that ought to be most responsive, mutually, in order to successfully respond to the assignment of solving social problems.

If the solution of a particular social problem appears to be a prerequisite for any attempt at the solution of many other social problems, then this solution may well be given the high priority status of a need. For any negligence, procrastination, or refusal, though ever so cleverly hidden behind pretexts quoting helplessness, ignorance, circumstances or times, will generate, inevitably, the reactive phenomenon called discontent.

II.
Requirements

Alternatives

Discontent is the manifestation of a conflict between two images, both, somehow, generated by a human being: his image of how "things" are (ITA) and his image of how he would prefer "things" to be (ITB).

Looking at this manifestation as one of mere existence, we have a report. Looking at it as a desire for some change, we have a problem.

On this level of discourse it is possible to enumerate the various procedures that, alternatively or in combinations, will either stop the report (the problem disappears) or solve the problem (the report disappears):

1. Remove the human being.
2. Change ITA until it fits ITB.
3. Change ITB until it fits ITA.
4. Change "things" so that ITB=ITA.
5. Stop looking.

Combinations that contain the first or last of these procedures may be implied in projects proposed and funded elsewhere.

Upon contemplating the three remaining procedures one immediately faces several requirements:

Image

The concept of "image" requires completion:

How are images generated?

What and who generates the image?

Through what process is an image conserved?

How does an image act upon the person in whom it acts?

What does an image do, generate, cause, in and to the person in whom it acts?

What are the necessary and sufficient conditions under which it would become possible

a) to conceive of "image" as a process, simultaneously caused by preceding processes and causing subsequent processes?

b) to create, for use by any observer, an adequate and appropriate formal representation of such a process, where the formalism allows for and satisfactorily indicates the dynamic relations and the flow-ambivalence of any such process?

c) to simulate the process and, eventually, to extend it, in an interactive man-machine system?

The question can no longer be whether images should be tolerated as arguments for behaviors, actions, decisions. The question is rather whether one can at last trace the images that unquestionably are accepted as arguments.

"Things"

The concept of "things" requires amplification:

Upon removal of the quotation marks, "things" should not only be all that is referred to in potentially verifiable statements, but also all that is referred to in potentially unverifiable statements.

Anyone's image of how "things" are (ITA) and the image of how anyone would prefer "things" to be (ITB), both belong to the "things" that are, regardless of whether the "things" referred to in the images are, or are not, referred to in anybody else's ITA.

Most, if not all, human actions can be interpreted as statements which manifest in movement, language, expression, etc., the actual state of the acting person's ITA and ITB. This interpretation, however, is made all but impossible by certain consequences of the human use of human actions. One consequence is of particular importance here: underlying all human discourse lurks a notion according to which the effective communicativity of discourse depends on its compatibility with a "factual reality" which is considered independent of individual images. Rational reasoning based on this notion tends to look down upon individual images, conceding to them a merely marginal and phenomenological placement of existence in "factual reality", if any place at all. In a context of cognitive studies for the solution of social problems such a notion still may generate most necessary, but no longer the sufficient, conditions for either research or concrete implementation.

The student of social problems and cognitive processes must be capable and be equipped to observe, and then to present analyzable documentation of, at least, three "things" and their relations:

a) A man's ITA and the evidence which it includes and to which it is compatible at generation-time.

b) The process by which (a) is modified until it becomes transferable to some form of manifestation satisfying the notion of compatibility with "factual reality".

c) The result of (b), in particular with regard to the traces left in it from (a).

Such an observer, his capabilities, and his equipment, together define the "Cognitive Laboratory" necessary and sufficient for the desired participation in, and the desired experiencing of, any process connecting the cognitive process in human beings with the social reality wherein the individual sees himself, and with the social reality which sees the individual.

So amplified, the concept of "things" will invite and encourage people of all walks of life to come to such a Cognitive Laboratory, to be observed, to observe themselves and others, and to learn how to articulate their ITA and their ITB with less and less compromise. Together, and as a continuously functioning system, the users and the Lab form the Interactive Interface which connects, without representatives, cognitive processes with social problems, and so, that this connection becomes accessible to anyone concerned.

Change

The concept of change is to include its own ambiguity:

a) There are "things" that can neither change, nor be changed.

b) There are "things" that can change, but not be changed.

c) There are "things" that cannot change, but can be changed.

d) There are "things" that can either change, or be changed.

e) And then there are the "things" that continuously change, or continuously are being changed, or both, and simply never stay the same.

In keeping with what has been noted previously concerning "Image" and "Things", the statements above may not reflect any knowledge of some "natural" states of affairs. They certainly, however, reflect part of the gamut of notions with regard to "change" as manifested by social human beings in reference to their ITA and ITB.

If my ITB implies to me that something ought to change or to be changed, I do not necessarily manifest, thereby, a social problem. A social problem becomes manifest, however, as soon as I, alone or with a group of people, begin to hold the society, of which I am a part, responsible for the impossibility to solve the conflict between my ITA and my ITB.

A society that knows but refuses to solve its social problems is a social problem.

In the attempt of solving social problems society has established links of communication between its members, such that some members represent, as it were, society, and some members either are, or represent, the discontent. The result of any meeting of these

two groups depends, of course, on various sets of powerful factors. The set that interests us here is among the most powerful and refers to the five statements heading this chapter. The meeting will be fairly successful if the discussion deals with "things" on whose properties with regard to change the participants can agree.

It can be fully successful only if the agreement is not due to some misunderstanding. If, due to some misunderstanding, no agreement can be reached, the meeting not only will fail in solving the problem, but will actually amplify it.

Misunderstandings result, where the partners, alternatively or simultaneously, confuse b with c, or, meaning to speak of d find themselves talking of e or answer with a to a statement using c; in short: when people are unable to transfer their images undamaged into the linguistic domain.

Such meetings (and, most of the time, misunderstandings) are taking place by the hundred thousand every day, on all social levels, among any variety of partners one could name. The thickening cloud (nourished by scenes in the home, in schools, industrial plants and offices; propagated by administrative and executive bodies of all kinds among themselves and when facing the whole or parts of the public) pollutes the social atmosphere, and is rightly called the phenomenon of alienation. More than by anything else alienation is caused by the inability of dealing with the concept of change. (Which is not yet, nor by any means, equal to, or equivalent with, the implementation of change.)

All meetings of this kind ought to discuss at least three conflicts and their causes, and not just one. A imagines himself in a conflict caused by B. B imagines himself in a conflict caused by A. Both have to face up to the conflict between their images. If, instead of discussing the causes of all these conflicts, A disputes B's rights to his conflict and B retaliates in kind, then this leads to the usual deadlock, out of which there lead only two ways: indefinite postponement with mutual frustration, or mutual display of power with explicit or implicit violence. Both ways, unsurprisingly, alienate the participants of the meeting, not only from one another, but also from the issue which the meeting was to investigate.

III.

Towards a Cognitive Technology by way of Heuristic Research

Assignment

Let us, for the sake of cognitive studies, distinguish between the engineer and the technologist: let the engineer's assignment be implementability, the technologist's assignment be applicability. Both have to study, to design, and to produce; occasionally even to create.

If a scientist wishes to examine the validity of a hypothesis he usually provides for a context in which the hypothesis can be shown either to maintain itself or to collapse. The scientist turns "technologist" in determining the applicable context. He turns "engineer" in designing and implementing the experimental set-up, the context.

Let us now, for the sake of the study of cognition, reassemble the assignments of the scientist, the technologist, and the engineer into one assignment given to one person: the cognitive technologist. If a cognitive technologist wishes to examine the validity of a hypothesis, he studies the hypothesis until he is able to provide a description and a design for implementation of that context, that experimental set-up, in which this particular hypothesis will maintain itself.

Cognitive Technology

Whatever his role and importance for the rest of the scientific establishment may be: the cognitive technologist is an indispensable member of any team that attempts to study cognition (1). A few reminders will explain.

Cognition, in order to be studied, is said to be, not the knowledge acquired, but the process of acquiring knowledge. All knowledge, once acquired, turns up as a hypothesis which, then, either maintains itself or collapses in a given context. The process, however, by which this knowledge had been acquired maintains itself, regardless of the fate of its acquisition (3). Furthermore one may state: the process by which knowledge was acquired is a context, a biological experimental set-up in which the acquired knowledge was able to maintain itself, until it became a hypothesis to a testing scientist. It is, therefore, worthwhile investigating the following thesis:

Every hypothesis and, in fact, every statement either is, or represents, or contains, knowledge that was able to maintain itself in the context of the process that acquired it. Cognition, therefore, must be a member of the set of all contexts in which a given hypothesis or statement can maintain itself.

Cognitive Technology is the scientific discipline which 1) searches for contexts in which hypotheses and statements can maintain themselves; 2) designs situations and facilities which allows for the demonstration and intensive study of such contexts; 3) studies the properties of these contexts and constructs models reflecting the results of those studies; 4) compares these models with other models reflecting the studies of biologists, neurophysiologists, biophysicists, biochemists, yes, even linguists, anthropologists and, last not least, psychologists and philosophers; 5) will eventually expose a preferred set of probably anatomies of cognition.

Heuristic Research

It is obviously necessary to draw a significant distinction between the study of cognition and the study of knowledge. In theory, and with the help of carefully structured formalisms, it is possible to draw this distinction and to maintain it as rigorously as the theoretical equipment permits (2). In practice, however, this is extremely difficult. In fact, there arises a non-trivial problem, whose satisfactory solution actually may be one of the most important steps toward a valid description of cognitive processes.

The dynamics of a process communicate themselves to the observer only through the traces which the process generates in the observer's domain of perception. The observer forms an image of the process and its dynamics by "interpreting" the traces it left behind. In order to be available for interpretation, the traces must be distinguishable. Traces can be distinguished only if they are offset against an environment or background which is void of the kind of traces under study.

The dilemma for the student of cognition lies in a particular ambiguity of the traces that he perceives. Although they were left by a process of cognition, he no sooner distinguishes them, and they become traces of knowledge within him, or the observed partner, or both. At the same time, the environment, or background, against which the traces were to appear in outline, is not void at all, but full of traces left by processes of cognition; only that these traces are indistinguishable for the observer and, therefore, face him as one distinctive trace of no-knowledge. The consequence of this imagery sounds perturbingly absurd: if the observer is to study cognition, then he has to turn to the study of knowledge with the determined effort to search for all that which knowledge and no-knowledge have in common. This means that he deals with the same traces as would the student of knowledge, but that he interprets the traces differently. The student of knowledge will tell us that they exist and why they are still there, maintaining themselves. The student of cognition will tell us how they got there in the first place and why, be it successful or not, there is something that submits itself to the test of survival.

At this point, and in defiance of all apparent absurdity, enters the concept of heuristic research conducted by the cognitive technologist. The cognitive technologist assumes that every living organism needs to be aware of the conditions under which it can continue to be a living organism, continue to maintain itself. This state of awareness is sustained by a process that continually tests for conditions. These tests and their outcome orient the organism towards its "judgement" of the prevailing conditions. If, according to this "judgement", something appears to be amiss, then a search for the missing condition is initiated. The cognitive technologist, now, experimentally assumes that the process of cognition is either part or all of this search for the missing condition. The particular search itself will end as soon as the missing condition is found and implemented. Knowledge would then be the network of all traces left by the halts of the process of cognition. The process itself thus can not be observed when it has halted, through knowledge, but only when it is in action, when it is in search, while it generates the conditions which the organism will then accept or reject. In order to study the process of cognition in living organisms the investigator must create an experimental set-up of a model situation, wherein he can observe the process rather than the traces of its stops (4).

This is, by definition, a heuristic set-up, and the most applicable compatible model that could, by analogy, represent the subject and the object of this research. A typical assignment for the cognitive technologist is the task of imagining, designing, and constructing such a model, to teach people to use it, to use it himself, and to apply it to the solution of problems and to the implementation of these solutions.

Wants

If a problem wants to be solved then the solution of the problem wants to be made possible. A careful analysis of those two wants frequently leads to an understanding of the conditions (effort, time, budget, equipment) and dimensions (quantitative and qualitative) that might allow the satisfaction of one want to become applicable to the satisfaction of the other.

The process of cognition takes place not only in the people who study it but also in the people who either are or have the problems that are to be solved. If the study of relationships between cognition and problems is to become practically applicable as soon as possible, then the ranges and the domain of this study must be broad, of great variety, flexible, and as accessible as possible to all kinds of proposed solutions, problems, experiments.

What is required here, then, is accessability in several senses and directions. The cognitive technological experimental set-up must be large and flexible in order to be accessible to the high level of problems and the large number of people involved; to be accessible to parallel and simultaneous real-time simulations of

environmental, psychological, neurophysiological, and social conditions; to be accessible to itself in order to actually self-organize into a crystal, or a network, or a circuit, or a model of a brain, whatever is wanted, so that crucial questions can be answered quickly, ambitious hypotheses be tested rapidly, and the vital link between theory and practice be efficiently established as early in the game as possible.

What emerges, indeed, is the need for a highly functional edifice of ample dimensions which, in the eyes of the cognitive technologist, not only would contain the modular equipment that is to serve him in an immense variety of constellations, groupings, and interlocking permutations, but which, as a whole, would itself be one of the modules in the experimental research design: a veritable house of heuristics (1) (4).

It is proposed that careful attention be given to this emerging need, that a time be set when a first design of such a project might be met with encouraging comment and criticism, and that a plan be initiated that foresees and begins now with the preparation for the eventual realization of this far reaching project.

Such a plan would be able to draw immediately on work done, or in progress, to be found in many locations, where support and coordination combined with practical implementation and real-time observation would bring to useful fruition, what otherwise will ineffectively linger in theoretical limbo. All such work might function in the context of the mentioned plan as a module, which, modulated with other modules of different structure and intent, would reach into regions of questions, answers, and conjectures, that, standing alone, it could not tackle.

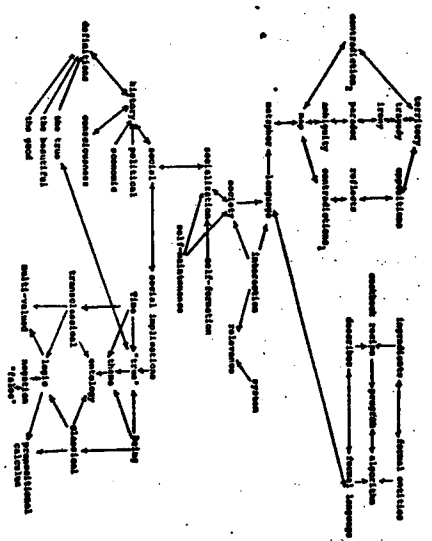
No problem, be it an individual's, a family's a community's or a nation's problem, no problem can be solved unless it knows how to state itself. The study of cognition on cognitive models and people and situations will give to the problems the voice and the word for the statement of need, and thus necessity. Knowledge alone is but a loop leading nowhere. To make it lead somewhere is the assignment given to Cybernetics by the Cybernetician of the second order.

References

1. Brün, H.: "Technology and the Composer" in Music and Technology, UNESCO and LaRevue Musicale, Paris, p. 181-192 (1971).
2. Garvin, P. (ed.): Cognition: A Multiple View, Spartan Books, New York, 428 pp. (1970).
3. Maturana, H.: "Neurophysiology of Cognition", ibid., p. 3-24.
4. Von Foerster, H. and H. Brün: Heuristics, A Report on a Course in Knowledge Acquisition, BCL Report No. 13.1, Biological Computer Laboratory, Department of Electrical Engineering, University of Illinois, Urbana, 27 pp. (1970).

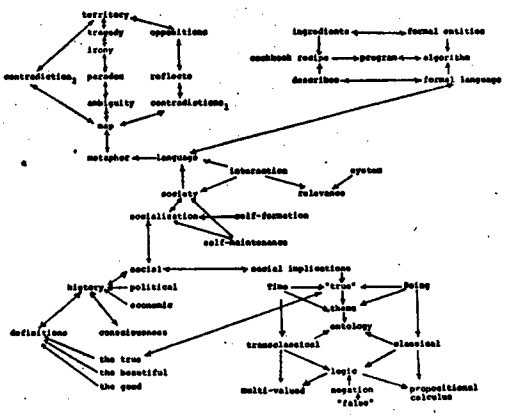
SOCIETY

A community that is self-maintaining economically and that, in the process of its economic self-maintenance, generates and maintains the other, non-economic forms of interaction of its members. [R.H.H.]



SOCIALIZATION

a) the process of self-formation of a society; b) the process through which a person is integrated into the process of self-maintenance, economic or otherwise, of a society. [R.H.H.]



PERSONALITIES, AFFINITIES, GENES AND HAPPENINGS

Heinz Von Foerster

Consider n distinguishable personality traits p_i $(i = 1 \to n)$, e.g.:

aggressive; avaricious; cautious; competitive; cynical; loquacious; moody; pious; polite; righteous; temperamental; vulnerable; etc.,

then any subject S_k $(k = 1 \to N)$ can be characterized as possessing $(p_i = 1)$ or lacking $(p_i = 0)$ each one of these traits (e.g.: $p_1 = 0$, $p_2 = 1$, $p_3 = 0, \ldots p_n = 0$). That is, the personality P_k of subject S_k is given by an n-digit binary number (e.g. of above: $P_k = \underbrace{010\ldots0}_{n \text{ digits}}$).

With n traits being either present or absent, clearly there are

$$M = 2^n$$

distinguishable personalities P_k $(k = 1 \to M)$ identifiable. Moreover, if the number, M, of distinguishable personalities is smaller than the population $(M < N)$, then there must be some sets of subjects that have the same personality. Call these the groups g(P) with personalities P_g $(g = 1 \to M)$; on the other hand, if $M > N$, then with overwhelming probability each subject S_k will have a unique personality P_k. The world population (October 1972) was estimated to be

$$3.78 \cdot 10^9 = 10^{9.58} = 2^{31.8}$$

subjects. Hence with slightly more than 32 traits (say 33 or 34) we may identify each subject uniquely by his personality.

Subjects with personality P_g may or may not accept subjects with personality P_h $(h \to M)$. Denote the dyadic relation of "A is accepted by B" (or, equivalently, "B accepts A") by

$$\text{Acc (A, B)}$$

and let this relation be symbolized by an arrow:

Conversely, denote the relation "A is not accepted by B" (or, equivalently, "B does not accept A") by

$$\overline{\text{Acc}}\ \text{(A, B)}$$

and symbolize this by the absence of an arrow:

The same notation shall be used for personalities P_g, P_h, associated with the corresponding subjects.

Since "Acc" is not necessarily reflexive, symmetric or transitive, each group (subject) may enter this relation with any other group (subject), including itself. Hence we have to consider

$$M^2 = 2^{2n}$$

acceptance relationships, each of which may be affirmative or negative. When this has been determined a particular

"Social Acceptance Network"

(SAN) is defined. Because of the binary property of "Acc" (either affirmative or negative) the number of SAN's that can be drawn is

$$N(\text{SAN})_n = 2^{2^{2n}}$$

E.g., for three distinct personality traits (n = 3):

$$N(\text{SAN})_3 = 2^{64} \approx 2 \cdot 10^{19},$$

or about twenty billion billions.

Consider as a specific example only two personality traits

P_1 = polite

P_2 = vulnerable

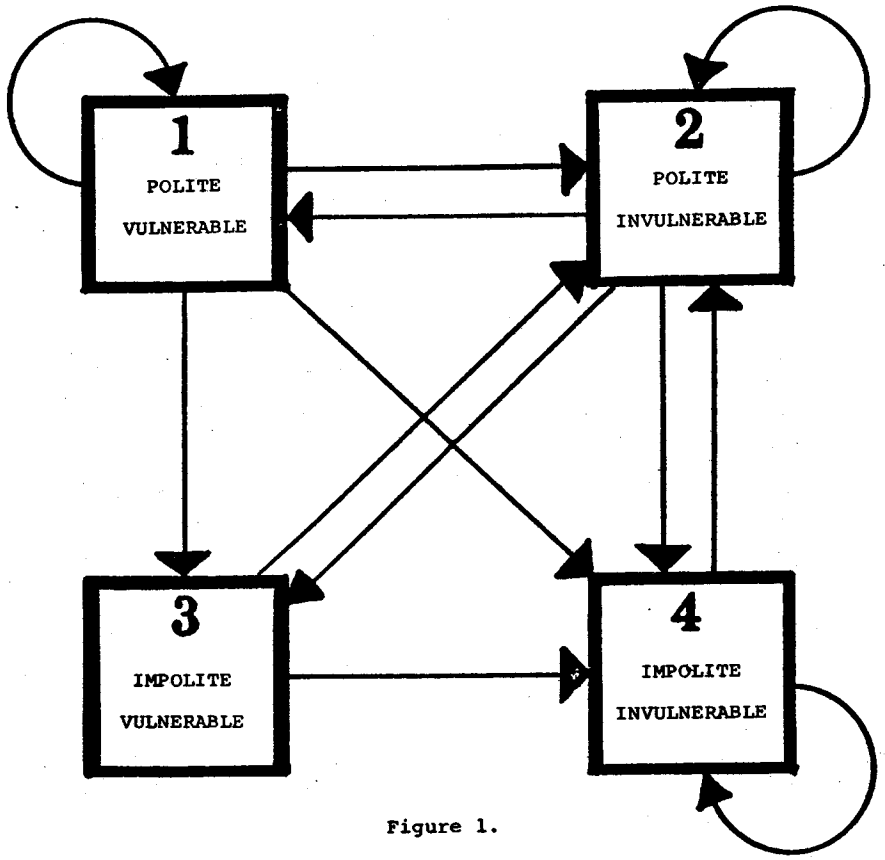


Figure 1.

These form $2^2 = 4$ groups whose subjects have personalities

P_1 = polite, vulnerable

P_2 = polite, invulnerable

P_3 = impolite, vulnerable

P_4 = impolite, invulnerable

We can now draw the SAN for these personality groups by arguing that, for instance, a polite and vulnerable subject will be accepted by an impolite and vulnerable subject, but inversely, all vulnerable subjects will not accept impolite subjects, for the vulnerable ones are not fortified against the impoliteness of these subjects, etc. Fig. 1 shows the so completed SAN which is, with n = 2, one out of a possible 65,536.

From inspecting Figure 1 one sees that subjects of groups g = 1 and g = 2 are universally acceptable, and subjects of groups g = 2 and g = 3 are universally accepting. If the property of being acceptable and accepting is termed agreeable, then subjects of group g = 2 are universally agreeable. Moreover, one may discover that groups 1, 2, 4 are "self-agreeable", while subjects who are impolite and vulnerable (group 3) are "self-disagreeable".

If these traits are thought of as being genetically determined, then group 2 (the polite and invulnerable ones) should predominate in a population, because of these subjects' being connected to others by the maximum number of gates (8), while group 3 should become extinct, because of its subjects' having a minimum number of links (4) to others. On the other hand, daily experience shows that group 4 (impolite and invulnerable) dominates the social scene. From this one may conclude that invulnerability is a trait inherited, while impoliteness is a trait acquired.

Postscript: This story is the reconstruction of a lecture Karl Menger gave at the University of Vienna in 1933. Since I cannot establish what is K. M. and what is H.V.F., I shall take the responsibility for all errors and give credit to all that is sound to K. M.