

1

Aspects of Machine Intelligence

Introduction by Gordon Pask

The current status of mindlike computer programs is summarized, at a philosophical rather than technical level, in the following short but authoritative papers: Minsky (1968), Simon (1966), Turing (1969). Whoever wishes to delve into this subject in greater depth may read the books where these papers are published in their entirety, augmenting them, to obtain comprehensive background, by Ernst and Newell (1969); Ashby (1960); Cohen (1966); Fogel, Owens, and Walsh (1966); Von Foerster and Zopf (1962); Uttley (1959); Von Foerster et al. (1968); McCulloch (1965); Oestreicher and Moore (1968); Amarel (1969); Rose (1970); Minsky and Papert (1969); Feigenbaum and Feldman (1963); Banerji (1969); and Garvin (1970). It is also worth perusing all volumes of the journal *Artificial Intelligence*.

Henceforward, it is assumed either that the reader knows the *kind* of symbolic operations performed by computer programs and other artifacts, that he will study the matter at leisure, or that he will take these operations for granted. With this supposition in mind I shall give a personal and possibly idiosyncratic view of the conditions under which *artificially intelligent* is a properly used term and offer an interpretation of these conditions with respect to *use of the architecture machine*. Apart from the pictograms or ikons developed in the text, the only special symbols used are the special brackets \langle and \rangle which enclose *ordered* collections of objects; the equality sign $=$; and \triangleq , which is read as "*defined as equal to*."

Overview

The contention is as follows: Intelligence is a property that is ascribed by an *external observer* to a *conversation* between *participants* if, and

only if, their dialogue manifests *understanding*. Each italicized word in this sentence requires careful attention. To give the flavor of the argument, *understanding* will be defined both in terms of the processes that give rise to such an interchange; roughly, understanding of a *topic* \triangleq (defined as equal to) a *relation* implies the existence of a *concept* \triangleq a *procedure* (for bringing about or satisfying the relation) and a *memory* \triangleq a *reproduction* of this *procedure*, together with a self-replicating organization, of which topic, concept, and memory are a part.

This point of view emerged in the late 1950s and has been reported, chiefly in connection with experimental data, in a series of publications. (See Pask, 1959, 1960, 1962, 1963, 1965, 1966, 1968, 1969a, 1969b, 1970a, 1970b, 1972a, 1972b; Pask and Feldman, 1966; Pask and Lewis, 1968; Pask and Scott, 1971). It resembles Von Foerster's theory of finite functional systems (1970b; see also Von Foerster, 1970a). It grew concurrently as part of a school of thought encouraged by McCulloch and owing a great deal to his concept "redundancy of potential command" (1965). Various formulations are possible. The present argument is most easily referred to Leofgren's (1968, 1972) mathematical model; an alternative formulation is given in Barralt-Torrijos and Ciaraviglio (1971). In this paper, mathematics is put aside in favor of ikons that *do*, however, have a deep logical connotation and are *not* simply loose visual analogies.

Insofar as *intelligence* is a property adduced by an external observer, the conversation has a great deal in common with the gamelike situation underlying Turing's Test (1963) (for intelligence in a somewhat different sense). But Turing's game and my conversation are not identical, and the interested reader may profitably compare the two and, in some respects, contrast them.

Aphorisms and Arguments in Support of the Definition

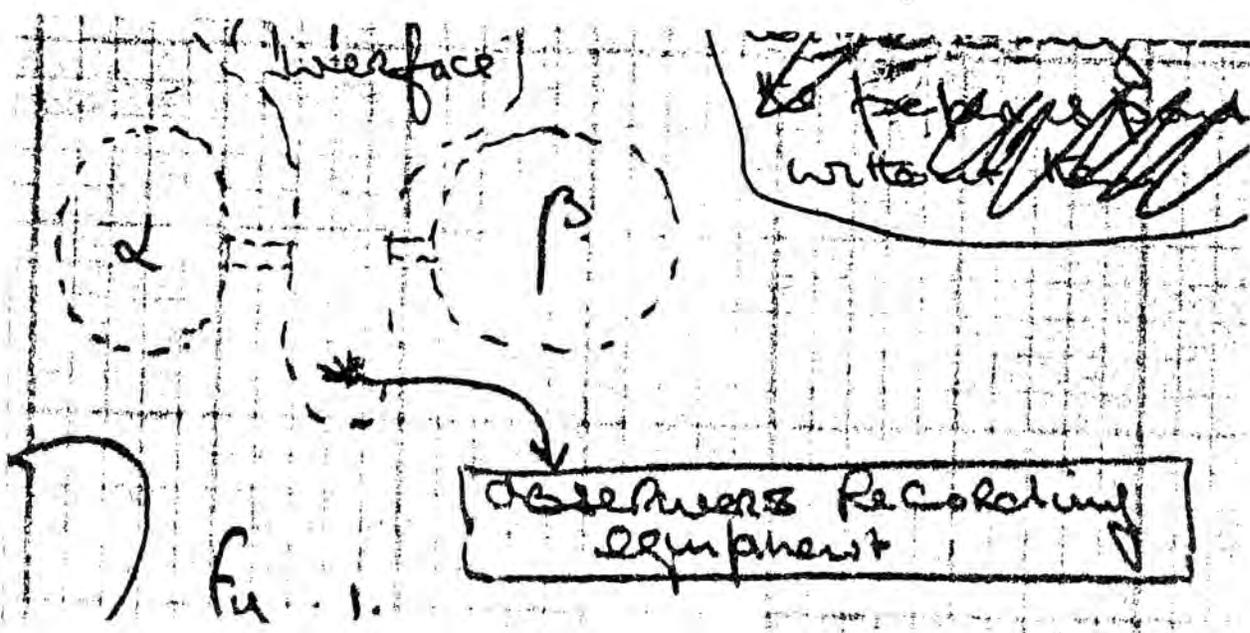
1. An external observer speaks in a metalanguage (L^*) used to discuss theories, describe experiments, and prescribe designs for equipment. The metalanguage is a natural language, very often scientific English.

2. The observer can distinguish stable entities of various kinds. Two kinds are of special importance: “mechanical individuals” or *M Individuals* and “psychological individuals” or *P Individuals*. In both cases, the stability is due to the same root cause—self-replication. But this fact is frequently suppressed in the case of *M Individuals*, since the replication process (being biological or due to the operation of natural laws) does not intrude into the phenomena under scrutiny.¹

2.1. An *M Individual* is distinguished by the familiar methods of classical physics and behaviorism. For example, a man is such a thing; so is an animal; so is a unique machine. It has a spatio-temporal location and is usually juxtaposed with another *M Individual* called its *environment*.

2.1.1. The term *environment* is specifically reserved for entities that can be described or prescribed in the manner of *M Individuals*: that is, in terms of *states* and *state transitions* (whether in the sense of automation theory or the very different sense of physical states) where *state* \triangleq the conjoint values of all descriptive attributes, and *state transition* \triangleq an operator carrying one class of states into another.

2.1.2. In the L^* description of a typical experiment, pairs of *M Individuals* *A* and *B*—one, perhaps, an environment—are coupled (Figure 1) via an *interface*. Apart from this interaction, they are isolated.



2.1.3. It is crucial to the argument that *all* observations occur at such a spatio-temporally localized interface; the observer's measuring and recording equipment is, in the last resort, bound to it. But the interface is neutral regarding the type of interaction, if any, that takes place across it.

In Figure 1, which introduces the notation for distinguishing *M Individuals*, α may be a user of the architecture machine regarded as a biological unit and β the architecture machine regarded as a chunk of metal and semiconductor material. But α may also be a rat and β its experimental environment.

2.2. A *P Individual* is distinguished as a self-replicating and (usually) evolving *organization*. It is respectably and precisely defined in terms of an object language L and a relational domain R described in L by a description $D(R)$ with respect to which it *is* self-replicating. Here, self-replication is intended in the abstract sense of the theory of reproductive automata, as originally conceived by von Neumann (1968) and as recently developed by Loefgren (1972).

2.2.1. Though, in general, the domain may be allowed to grow systematically under the control of the given *P Individual*, we confine our attention to cases in which R is fixed. Under these circumstances, it is possible to specify domains with the property that if a given *P Individual* is viable (that is, is able to reproduce) on occasion n , then it is also viable at any later occasion $n + r$ (r finite) for R_i in R .²

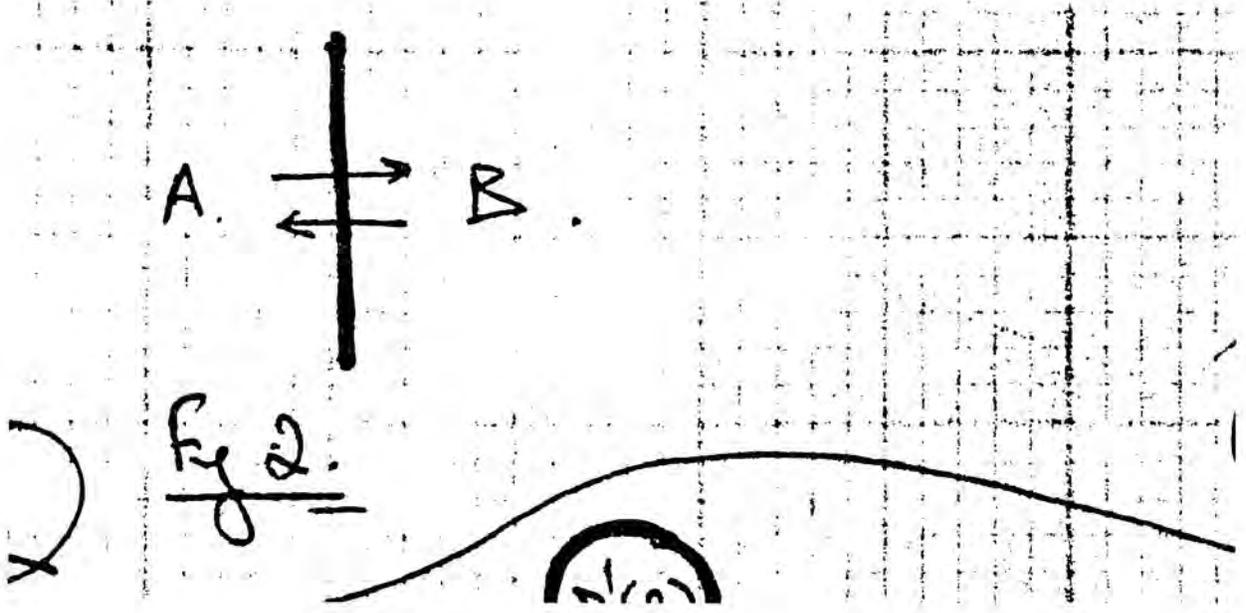
2.2.2. It is assumed that a *P Individual* is active or that any conversation in which it is a participant does in fact proceed, that is, for each occasion, some topic relation R (a part of R or all of it) is actually ostended for

discussion. Rather complicated but not esoteric conditions are imposed, in the full theory, to guarantee that this is so.

2.2.3. Typical *P Individuals* are people regarded as personalities—characters (in plays) executed by *any* actors, the performance of stable roles in society, the organization of coherent groups, factions, governments, cultures, and persistent ideas. A vertical cleft notation I is employed to discriminate *P Individuals* labeled *A* and *B*, as in Figure 2.

2.3. A *conversation* is taken to be the minimal situation for a meaningful psychological or, a fortiori, mechanical-psychological experiment. It consists of an activity involving at least one *P Individual A* and generating an *L* dialogue. On each occasion *n*, when the interaction is focused on a topic *R_i* of *R*, this interaction gives rise to a further *P Individual* called a *sprout* (growing point), which can be dissected into a portion *S_A* and a portion *S_B* with certain well-defined technical properties; namely, on occasion *n*, *S_A*, *S_B* are productive systems in respect of a surrounding *R_i* using the terms *productive* and *surrounding* in Loefgren's sense (1932) and at least one of them, *S_A*, (and possibly both) is reproductive both in the surrounding $\langle S_A, R_i \rangle$ in the surrounding afforded by *A* (of which *S_A* is an externally delineated subsystem).³

2.4. The circularity inherent in this specification is quite deliberate. *P Individuals* are recognized by the existence of conversations, and the conversation itself is, on a *given* occasion, a further *P Individual* (the *sprout*). Hence, the form of the dialogue in a conversation is determined as an *L explanation* or *L modeling* operation, which is precisely the *reproduction of the sprout*.⁴



2.5. Conversely, a certain (to be described) complex of explanation cycles is the L image of a reproductive cycle, and these L explanations are split, by the dissection that yields S_a and S_b , into questions asked by A of B (or vice versa), which are answered in explanations given by B to A (or vice versa).

2.6. The reproductive cycles of P Individuals (the sprout included) are due to procedures executed in *some* processors; it is apposite to concentrate on the architecture machine *qua* processor and the *user's brain*. But it should be emphasized that a P Individual has no necessary spatio-temporal location, and procedures that constitute P Individuals may be executed in several M Individuals just as an M Individual may execute several P Individuals. In ordinary conversations many-to-many correspondences are ubiquitous. Stable *concepts* are frequently shared, and *memories* (which may be legitimate P Individuals) are distributed throughout society.

2.7. Conversations occur autonomously and are discovered or noted by accident. Most of these conversations take place in natural language; in the limiting case, $L^* = L$. Hence, with certain exceptions like autogenous committee meetings and tribal rituals that perform a regulatory function, an observer is hard pressed to maintain the impartial poise of an external observer. Since it is important that he should do so in adjudicating the conversation as "intelligent" or "not intelligent," he needs to maintain a firm distinction between L^* and L .

3. The following remarks are thus confined to conversations *brought* into existence by an external observer who contrives some type of *contract* with any stable entity capable of understanding enough of L^* to agree to the contract and capable of interpreting L (of which the full semiotic is described in L^*). The nature of the entity that is party to the contract with the observer is, at this stage, left open.

3.1. In general, contracts are made with human beings or groups of them; in general, the observer speaks to (glances at, projects his voice toward) a human being or group in the sense of an *M Individual*; but at the same time, he negotiates the contract with a sentient creature, that is, the man or group regarded as a *P Individual* larger than the participant A .

3.2. The contract has the following clauses:

a. That the contracting entity will, henceforward, speak only in an object language L (in other words, the vocabulary of L will be used, and its syntax will be respected). Commonly, L is a mechanical language that does not involve verbal utterance.

b. That L will be interpreted with respect to a domain R , described as $D(R)$ (this is the semantic of L ; it contains topic relations germane, for example, to architecture, geometry, and mechanics).

c. That the contracting entity will play a role, designated A . This is the pragmatic aspect of L or A 's *intention* (for example, to be a *designer*, or, in selecting one R_i in R , to carry out a particular design). In particular, " A seeks a goal" means either " A aims to bring about R_i ," or " A learns to bring about R_i ," for some topic relation R_i in R .

d. That A will converse in L with a further entity B , that is, on each occasion n , A will aim for some goal; hence, some L expressions are used in an imperative or interrogative mode to pose and solve problems.

e. That the observer, for his part, will choose an L that is rich enough to accommodate the required questionings, commandings, answerings, etc.

f. That the observer will furnish a participant B (for example, the *heuristic* in the architecture machine) so devised that it will be *possible* for the other participant to realize the agreed-upon intention of playing the role of A .

4. In order to satisfy clause (6) of Section 3.2, an external observer must have an unambiguous representation of A . Because of that condition—because he wants to distinguish between a *concept* \triangleq a goal-directed or problem-solving procedure \triangleq the reproduction of a relation, such as R , and a memory \triangleq the reproduction of a concept, because he wants to judge the conversation “intelligent” or “not intelligent”—an observer finds it convenient to avoid dilemmas of self-reference: for example, the notion of a program that “writes itself” or a procedure that “questions itself” or even the operational evocation of a self-reproducing system (so that the *sprout* of a conversation, which is a *P Individual*, can be represented as a productive pair, S_a, S_b). One expedient adopted for this purpose is to stratify L , that is, to specify $L = L', L^\circ$ where expressions in L° refer to the bringing about of relations R , (the solution of problems, the achievement of goals), and expressions in L' refer to the construction or learning to formulate and achieve goals or learning to solve problems.

5. The distinction between levels of discourse in the object language L', L° , is symbolized by a horizontal cleft —.

5.1. Moreover, once imposed, the stratification engenders two descriptions of R , namely, $D(R) = \langle D'(R), D^\circ(R) \rangle$.

5.2. $D'(R)$ is a grammarlike structure indicating *what may be known or learned*.

5.3. $D^\circ(R)$ is grammarlike structure indicating *what may be done* (either by physical operations, to make a tangible model for some R_i in R), or by intellectual operations, to model R_i as an explanation—literally, of how to solve problems under R_i .

6. On making the distinction I and the distinction —, the observer declares the tableau of Figure 3 the conversational *skeleton*. This skeleton L and R are all described in L^* .

7. To lay foundations for the representation required to satisfy clause 6 of Section 3.2 and, simultaneously, to exhibit levels L', L° , in L as levels of control, the spaces in the skeleton are filled by boxes (Figure 4) representing classes of goal-directed or problem-solving procedures, *Proc i* being a procedure that *brings about* \triangleq *reproduces* a topic relation R_i .

7.1. The *superscripts* signify levels.

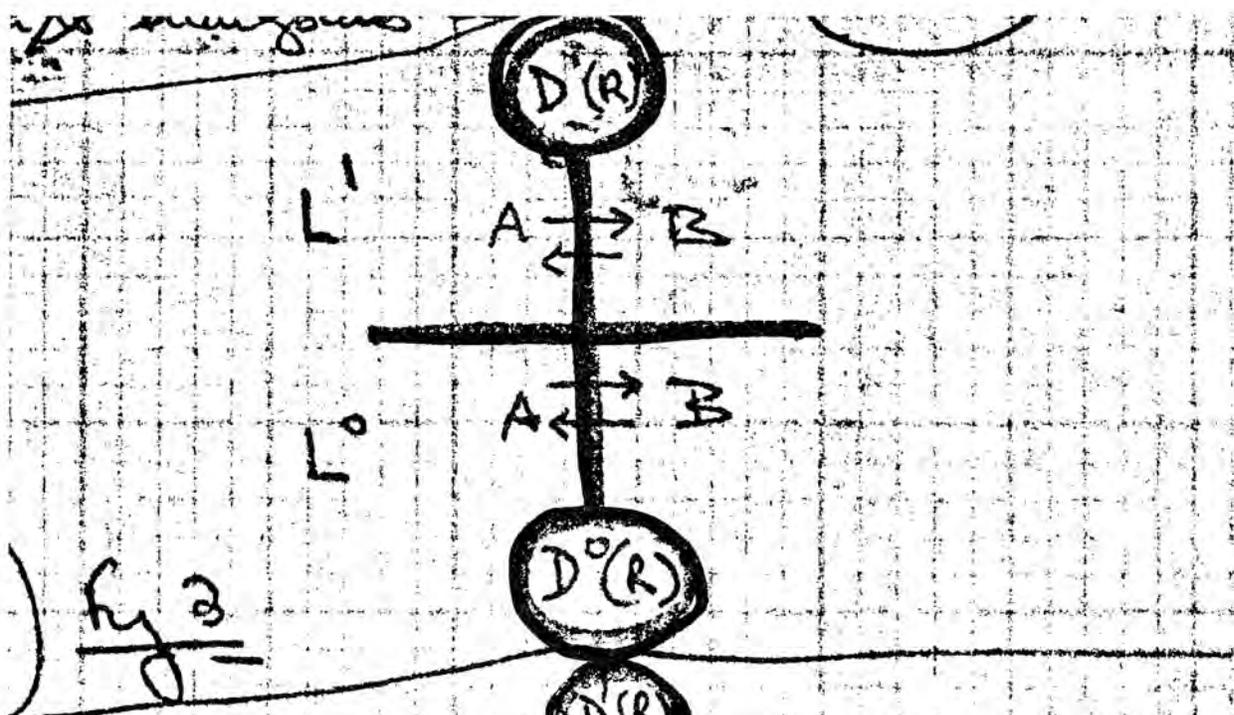


fig 3

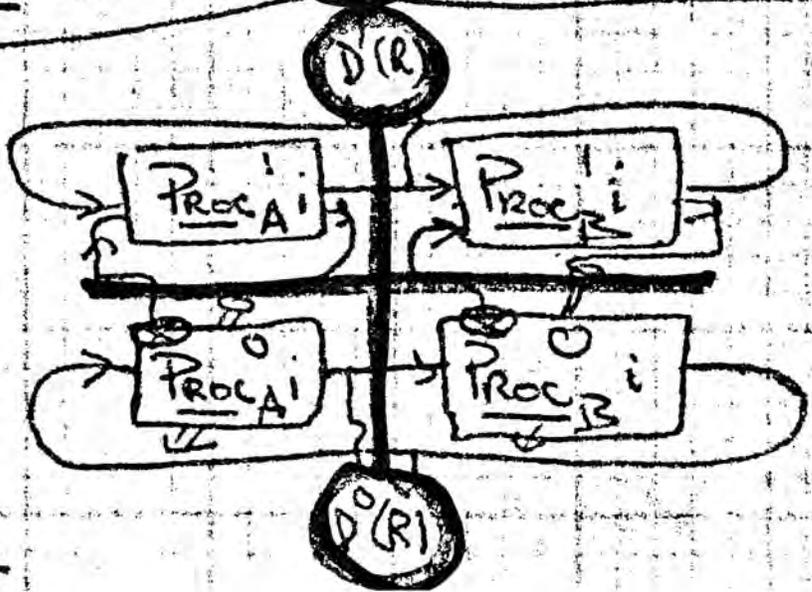


fig 4

7.2. ♀ means “operates upon according to a hypothesis,” and ⊗ means “gives a description (in the language appropriate to the level where the line terminates), which may or may not confirm the hypothesis.”

7.3. Thus a complete circuit on one side of I, starting at ⊗, passing through — to a *Proc*, and returning by way of — and ♀ on the original *Proc* is a *causal* coupling, or, equivalently, it permits *reproduction* of the original *Proc*.

7.4. The unadorned, horizontal connections have a different meaning: they are *inferential* couplings, which, limiting cases apart, entail the notion of choice.

7.5. Hence, any complete circle (such as the line emanating from *Proc_s i* to *Proc_s i* and terminating on *Proc_s i*) may be called a deductive chain.⁵

7.6. Finally, the lines to and from $D'(R)$ and $D^o(R)$ indicate whatever is referenced by the inference, that is, whatever R in R is ostended by the participants A and B on occasion n .

7.7. Call this ikon (Figure 4) the conversational *paradigm*.

7.8. If one ikon is created by filling the spaces in Figure 3, then (obeying the proper rules) the process can be iterated laterally to yield a further *paradigm*, for example, the ikon in Figure 5. The motivation for doing so is noted in Section 2.1.1 ≙ to represent as much of mind as desired.

7.9. Parsimony alone dictates as few inscriptions as possible.

7.10. Figure 4 sufficiently represents the sprout of a conversation if R_i is ostended on occasion n (a *P Individual* $\langle S_a, S_b, R_i, n = \langle \langle Proc_a i \rangle, Proc_a i \rangle, \langle Proc_b i, Proc_b i \rangle, n \rangle$, where n itself may be a vector) and the full requirement for understanding is satisfied if the form is iterated to the left until A is also a *P Individual*, even if devoid of S_b (a similar construction being possible, but not mandatory, for S_a and B).

7.11. To condense the notation, these iterated systems called *repertoires* of procedures (at level L' and L^o , available to A and B) are designated.

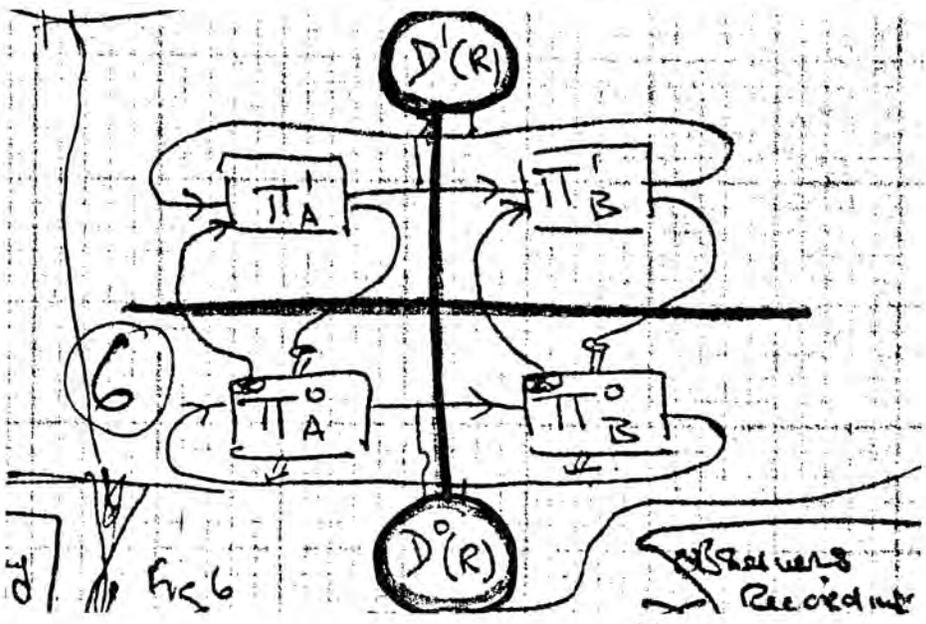
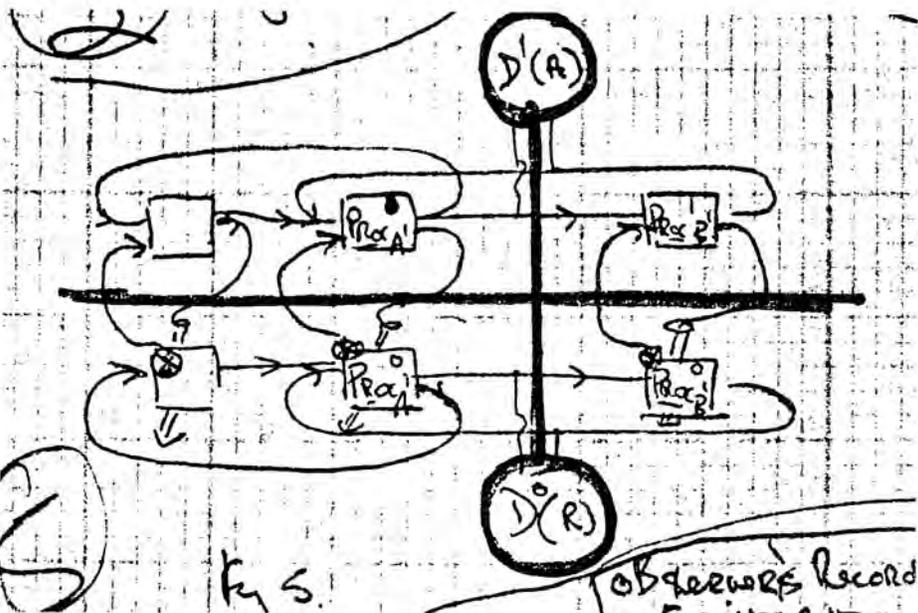
7.12. Repertoires are constrained by the rule that any such configuration contains a *sprout* on any occasion n (Figure 6).

8. The L dialogue across I implied by the existence of a *sprout* (specifically, by the ikon of Figure 4) is as follows:

8.1.

a. B can ask A to explain R and obtain an answer that *before* the end of occasion n matches *some* explanation B could have given in reply to the same question asked by A and, furthermore, A could have asked the question.

b. B can ask A to explain how he *knows* or is currently *learning* to explain R , and obtain an answer that *before the end of occasion* n matches *some* explanation B could have given in reply to the *same* question asked by A and, furthermore, A could have asked it.



c. Since the closure condition is in force (Section 2.1.3), the possible explanations in (a) above are described in $D^{\circ}(R)$.

d. Again because of the closure condition (Section 2.1.3), the possible explanations in (b) above are described in $D'(R)$.

8.2. Conversely, the joint holding of conditions (a), (b), (c), and (d) implies the sprout of a conversation, hence, a *P Individual*.

8.3. Likewise, this joint condition implies an *understanding* of R_i by A in which (a) is the L expression of a *concept* of $R_i \triangleq Proc_i \triangleq$ the reproduction of R_i , and (b) is the L expression of a memory of R_i ($Proc' i \triangleq$ the reproduction of $Proc^{\circ}i$).

8.4. If these conditions are not all satisfied *until* the end of occasion n (recall from Section 2.2.2. that the series of occasions is *assumed*), then the ikon represents an evolutionary process called *learning* the concept ($Proc^{\circ}i$) of R_i .

8.5. To obtain the general case, the entire argument is applied to the ikon in Figure 6.

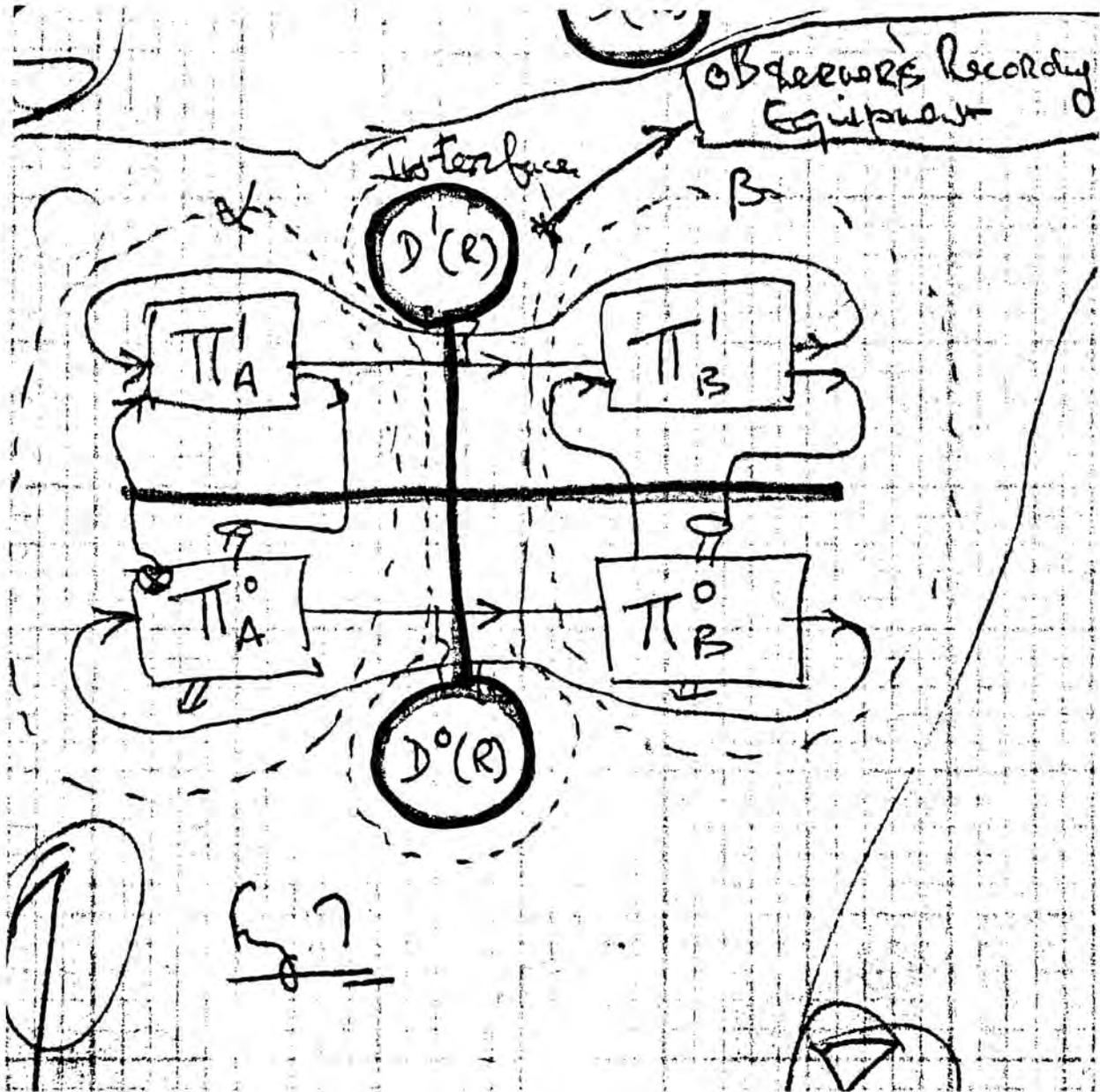
8.6. That such systems exist can be demonstrated in the abstract; that the understanding they image can be appreciated by participants is a matter of experience.

9. But for the L dialogue satisfying (a), (b), (c), and (d) to be unambiguously recorded and adjudicated by an external observer calls for the further requirement, specified in Figure 1, that the cleft I shall

coincide with a spatio-temporally localized *interface* to which the observer's measuring equipment is attached; in other words, that Figure 1 is superimposed upon Figure 6 (say) so that the *interface* is in register with I and engulfs some physical representation of $D(R) = \langle D'(R), D^o(R) \rangle$; A is in register with α , and B with β (Figure 7). If, under these circumstances, an observer *says* (in L^*) there is an *understanding*—that is, (a), (b), (c), and (d) are satisfied—then he deems the conversation *intelligent*.

Notice, however, that the form of interaction across the interface engendered by this construction is highly specific; it is L dialogue and could *not*, for example, represent the reactive interchange between a (laboratory) rat and its environment (whereas, in Figure 1 taken alone, it *could* do so).

10. An environment, in the strict sense reserved for this word in Section 2.1.1, can be added to the picture (Figure 8). It consists in a box U with the characteristics of a state and state transition system, as described in Section 2.1. The descriptors X_A are those properties *apparent to A* that tally with L^o predicates; its descriptors X_B are the properties *apparent to B*; its state is altered by the operations Y_A , that A may prescribe and describe in L^o (as m -tuples of values of L^o predicates), and the operations Y_B are those that B may prescribe. Hence, the environmental state is a function of two classes of variables, indexing the operator classes Y_A and Y_B . Its state on occasion n is relevant if it *instantiates* the relation R_i ostended at n . The members of X_A are those relations subordinate to R_i for which A has *memories* and which it *treats as properties*; a similar comment applies to X_B and B . A special interface V is used to localize transactions of this *causal* type.



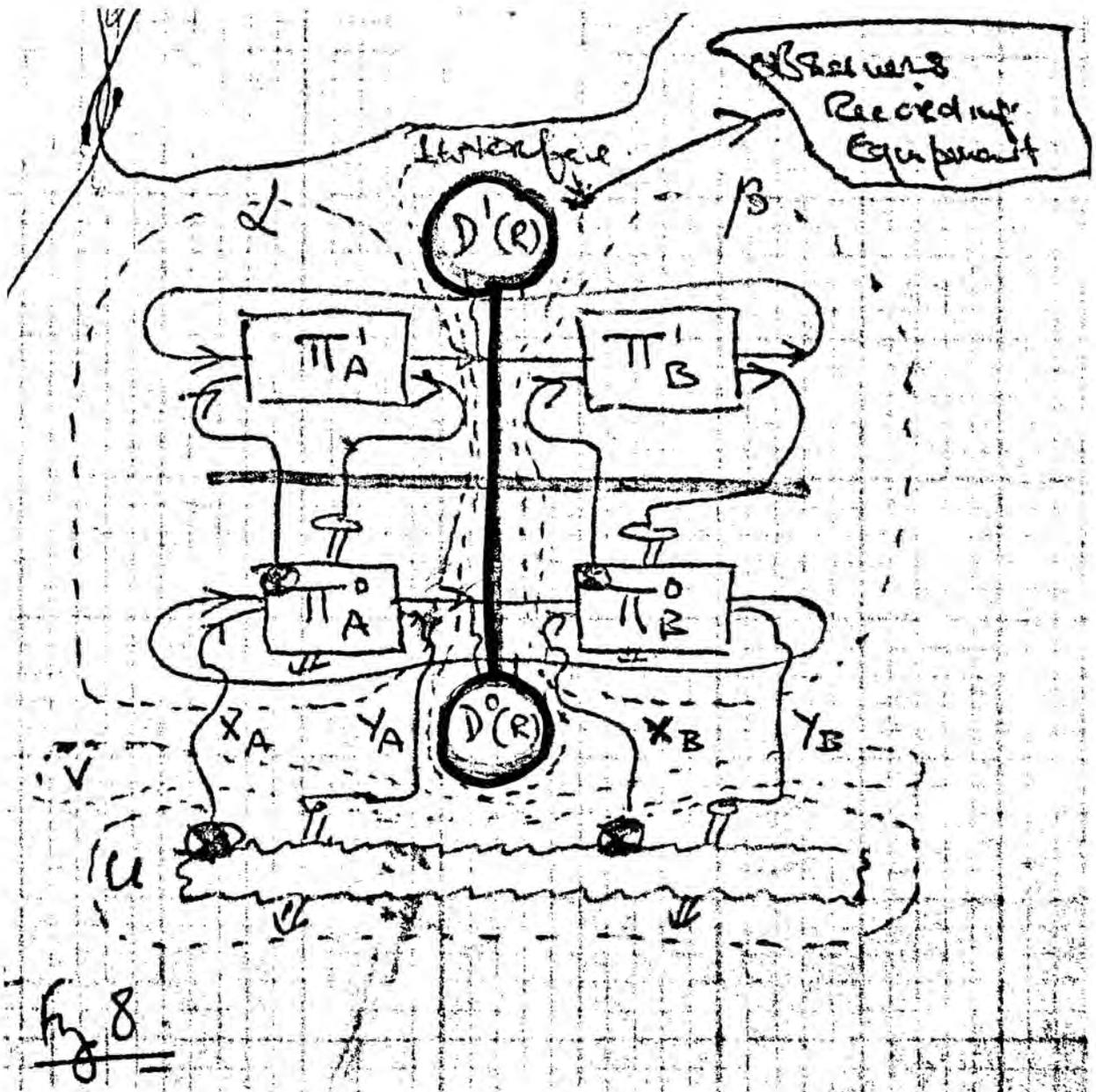


Fig 8

Naive forms of behaviorism are solely concerned with observing *causal* transactions across V and are thus not very informative. In particular, no conversation occurs by virtue of these transactions.

11. The joint requirement that a conversation (see clause (6) of Section 3.2) exists and its cleft is in register with an interface is satisfied when A and B are conscious human beings, one of whom is a skilled interviewer (B , correlated with B).

11.1. Moreover, the same is true if the interviewer's capabilities are truncated by adherence to a heuristic (thus deleting the right lateral extension of B that generally represents B 's mind).

11.2. I have shown, by constructing a rather elaborate machine with liberal facilities for graphic representation of $D'(R)$ and $D^o(R)$, together with arrangements to mark their constituents with tokens of aiming, access, working on, ostension, and exploration that B , in this minimal but adequate sense, can be the heuristic embodied in an electro-mechanical artifact. Using CASTE, the acronym for this equipment, it has been possible to investigate roles for different P Individuals (notably, $A = Student$, $B = Teacher$, and $A = Respondent$, $B = Interrogator$) and to plot, in considerable detail, the development of conversations and of the evolutionary component, which is regarded as *learning*.

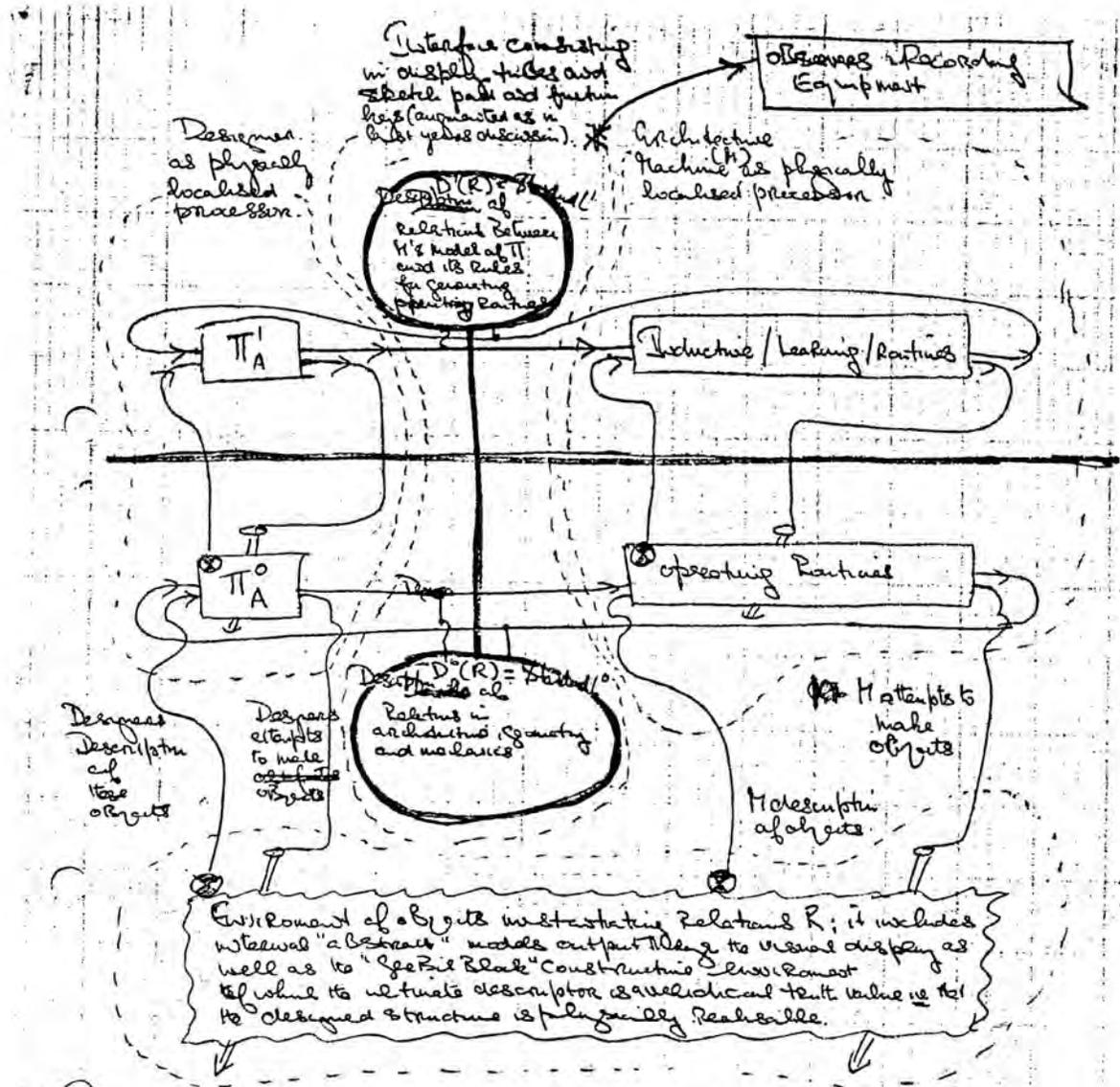
11.3. Further, the closure condition can be relaxed so that a conversational domain may grow as the discourse proceeds, though not in an unlimited fashion.

11.4. With some minor augmentation, judged feasible after technical discussions with Negrofonte's group, the Architecture Machine could, like CASTE, act with respect to *P Individuals* playing roles such as *Designer* and *Codesigner*. Our experience with the tutorial mode of CASTE suggests that this application would be well worthwhile. The outline interpretation for the Architecture Machine is shown in Figure 9.

11.5. In either case, the resulting conversation is deemed "intelligent" by an external observer since the conditions for understanding are secured by the regulatory *B* heuristic, which makes it possible for *A* to keep the contract he intends to keep (clause 6 of Section 3.2) as well as to maintain on the interface.

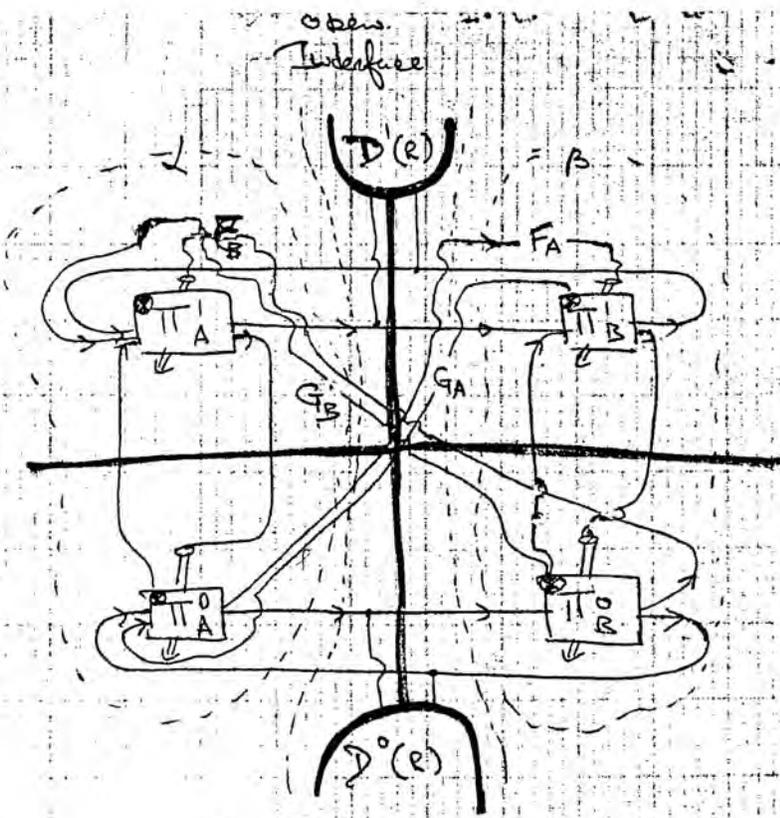
11.6. Said differently, the price paid for observation is that the external observer takes the conversation as his own environment in exactly the sense (Section 10) that the *P Individual* in Figure 8 takes *U* as its environment. The observer's description (analogous to but *not* at all *identical* with *L* expressions involving X_A, X_B) is an L^* description of *L* dialogue about *R*. This is what he records. To secure impartiality, he establishes a contract, which could be symbolized by *constant-valued* parametric arrows (analogous to but not identical with Y_A, Y_B) penetrating the uppermost process boxes adjacent to the cleft. To regulate the dialogue so that its conditions are satisfied on the interface (Section 11.5), he prescribes *B*, an interviewer or a machine, to act as his emissary, yet also as a participant.

12. Since one *M Individual* (*B* in Figures 6, 7, and 8) is a machine, the intelligence might be rated "partially artificial." The question of whether it is possible to achieve a "fully artificial" intelligence by making *A* (of Fig-



[Handwritten signature]

F. S.



10. The conditions to be satisfied as a prerequisite for electric and magnetic activity in a formal way be exhibited by the set of F_A, G_A or F_B, G_B are properly couplings. For electric (communicable usage) it is essential that both F_A, F_B and G_A, G_B are defined and realized.

10

ures 6, 7, or 8) out of metal is stated in Figure 10. The connections F_A , F_B , G_A , G_B , which allow A to take B as A 's *environment* and/or B to take A as B 's *environment*, are crucial to all manner of creativity and innovation; for, if these connections can be made, then a *P Individual* (the sprout of a conversation, at least) is an *observer* (Section 11.6) of itself. Once these connections are established, the closure condition is removed, the domain can expand (though not in an unlimited fashion), and, at the same moment, the stratification of L is lost, so that L may as well be L^* . If A and B stand for the brains of human beings, this trick is often played, and because of it, *P Individuals* are seldom fully correlated with *M Individuals*. I see no reason, in principle, why that trick should not be played with mechanisms, also. But, if it were, the mechanism would not be inanimate. Having this disposition, I prefer to avoid the qualifier "artificial" when speaking of intelligence.

1. In a coarse-grained account of the matter, a "natural law" is equivalent to a doctrine of "structural invariance." Considered in greater detail, it is possible to place natural laws in correspondence with regulatory principles that maintain and, as later, reproduce relations immanent in nature. This notion was mooted long ago (by Von Foerster, amongst others) and gives a nontrivial interpretation to causality, thus, for example, eliminating the confusion between *cause* and *enable*. The interested reader is referred to M. Bunge, *Scientific Research*, Vols. 1 and 2, (Springer Verlag, 1967) and requested to communicate with L. Perriera and L. Montiero (Dept. of Cybernetics, Brunel University or Centro de Estudos De Cibernetica, G.E.U.A. 53-9E Lisbon 5), who are systematically rewriting the principles of (near classical) physics in terms of feedback and regulator equations.

2. Throughout this paper it is assumed that the domain is of this type because heuristics exist for constructing such domains as relational structures with L^* description $D^*(R)$ and L descriptions $D(R) = \langle D'(R), D^o(R) \rangle$ as in Sections 5.1, 5.2, and 5.3. It should also be noted that $D^*(R)$ includes a set of *descriptors* for the graph or *entailment structure* expressing *what may be known* as well as the graph itself; thereby, for example, a real student can appreciate a topic relation in the context of others before he *knows* it or attempts to *learn* it. This class of knowable domains is much more restrictive than necessary. We have, for example, a CASTE-executed heuristic permitting evolution of the domain and can show that *this* is too restrictive. Though it can also be shown that there are limits upon knowable domains, or, at any rate, memorable domains, we have not yet done much empirical work to check that certain predictably immemorable relations are not, in fact, reconstructible.

3. Due to the special construction of the domain (Section 2.2.1 and its footnote and Sections 5.1, 5.2, and 5.3), R_i appearing in this expression covers all those relations needed by a given P Individual to learn R and thus to understand it. But, even with this construction, R might be learned in *many*, perhaps *infinitely many*, ways; that is, we are *not* characterizing domains as simple *hierarchies* of relations.

4. Though this statement is accurate, my theory includes several caveats and conditions. For example, the existence of a sprout on each occasion n ; that the conversational domain $D(R)$ is so organized that it is possible to consider more and less comprehensive relations, R_i ; and that the *sprout* selected on occasion n is a system that is reproductive and partitionable in a pair, S, S' , with respect to a surrounding that is the most comprehensive of the R .

5. Notice that this usage makes *induction* simply a higher level of deduction (for example, if the L^o grammar admits statistical inference, according to Bayne's rule).