On Self-reference, Existence and Validity in a Knowledge Acquiring System

A. PHALET

Summary

Self-reference – necessary in view of self-adjustment – is considered in two of the forms it assumes, viz. self-evaluation and self-parametrization, in a knowledge acquiring system S_{KA}. The first raises the liar paradox and the problem of a theory of truth through the evaluation by S_{KA} of its representations or methods. Self-parametrization occurs when a system, S_{KA}, evolving in time or space, represents events, among which itself, evolving in time or space. It is linked to McTaggart's time paradox. In view of a solution of this paradox, time functions are proposed, whose "values" are actually time descriptions concerning past, present or future, which consist of layers that have, so to speak, to be unfolded. The role of self-reference in the production of contradiction is passive: it doesn't bring contradiction about, but refers to its cause, a partial non-invariance, and could be said to require its elimination on pain of a regressus ad infinitum of the establishment of invariance. From another point of view, self-reference interrupts the constitution of invariance, that is, the coming into being of an entity, by the introduction of a loop in the process of the identification of the – so far partial – entity.

Invariance defines a relative form of existence: an entity is invariant with respect to, that is, it resists, some class of operations or relations that define or identify the entity concerned. This entity can be of the form of a set, closed for a definite class of operations or relations, or of the form of a limit that is invariant with respect to the progressive change of the elements of an appoximation sequence. Partial non-invariance appears to be the substratum of contradiction. The latter is a result of a false supposition – paradoxes are the result of logical error – that implies the invariance of the entity concerned with respect to a procedure or relation that is incompatible with other invariant aspects of the entity, e.g. diagonalization and enumeration.

1. *Self-reference in a knowledge acquiring system*

The logical and philosophical problems of self-reference, especially the connection of the latter with contradiction and the time-honoured problem of existence, are approached from the point of view of knowledge acquisition. For attempts to formulate or solve that kind of problems – the result of reflection without assignable empirical evidence – should also, once they have been analyzed for their own sake, take into account the intrinsic requirements of a working system, such as a knowledge producing system, in order to examine their relevance to empirically involved problem solving. Through the intermediary of a knowledge acquiring device, considered as a component of an expert system, the latest and, so it seems, most practical offspring or artificial intelligence – "the big corporations are getting interested" – the problems mentioned will, eventually, be transferred to the realm of systems technology and knowledge engineering. Lenat's EURISKO program provides an amusing example:

> "It (EURISKO) noticed that human rules tended to be better than its own, so it came up with the heuristics: *if a rule is machine made, then delete it.* Luckily the first rule EURISKO erased was that one!" ([1])

EURISKO's meta-rule could order its own deletion. To such a meta-rule corresponds, in the domain of declarative sentences, a sentence expressing its own falsity and, consequently, proceeding, so to speak, to its own negation, i.e. its own elimination from a theory, though, by that negation, the said sentence will reappear in the theory. For by its negation, it doesn't any longer depend on the empirically verifiable and, indeed, verified condition, e.g. that it is machine inferred as in our example below, condition that implies the sentence's falsity. Let us consider a knowledge acquiring system S_{KA}, which produced

> the machine inferred sentence s_1: *if a sentence is machine inferred, then it is false,*

which is a version of the liar paradox:

([1]) R. FORSYTH, The architecture of expert systems, in: *Expert Systems. Principles and Case Studies*, London, New York, 1984, pp. 9-17

Cretans are liars, says Epimenides, the Cretan.

The sentence s_1 is contradictory: s_1 is true iff s_1 is false. For, if s_1 is *true*, then, since it is machine inferred, it is *false* according to the information communicated by s_1 itself, If s_1 is *false* and, consequently, its negation, Neg(s_1), true, then s_1 is machine inferred and not false, that is, s_1 is *true* – or *undefined*, a notion which has to be explained.

As a conditional sentence, s_1 expresses a *truth* or *validity transferring procedure*:

> the empirically confirmable truth of "*s is machine inferred*" is transferred to "*s is false*".

Logic is expected to define a definite class of transformations or transitions – validity transferring procedures or relations – that, applied to sentences s_1, ..., s_n, yield a sentence s such that s is valid, if s_1, ..., s_n are valid, that is validity is preserved by or is *invariant*, or it *exists*, with respect to that class of transformations which characterize, at the same time, the validity concerned. Entities, objects or whatever name should be given to what is said to exist in a mathematical or physical way, e.g. a length, classical truth or a photon, are thought of as invariants with respect to a definite class of transformations or relations which define the entity concerned. This representation of existence is based on the general conception of existence as resistance to attempts at destruction or arbitrary alteration. The logical activity of S_{KA} concerns the formulation of invariance conditions of validity.

In order that a validity transition, as proposed by s_1, should succeed, truth has to be assignable to the consequent of the conditional sentence concerned, viz. "s is false", that is, the consequent in question must have a *sense*, or, equivalently, it has to be an encodation of an *identification* procedure of the truth-invariance. Consider, for example, the sentence s_r:

> "The ruler on the table is straight".

The expression "The ruler on the table" refers to an invariant – its referent – with respect to a class C of transformations. Plato would have considered s_r to be true, if the said referent should also be invariant with respect to the procedure:

> make the middle of the ruler cover the ends – to an eye at either end and looking along the ruler –

according to his practical definition of a straight segment (Parmenides 137 E):

straight is whatever has its middle in front of both the ends.

If that procedure doesn't destroy the system's, e.g. Plato's or S_{KA}'s, representation of the invariant concerned, viz. the referent of "the ruler on the table", then s_r is true, if that same, invariant saving, procedure is also finite. A destruction of the invariant, or rather of the representation of the ruler on the table, is the issue referred to by the negation of s_r, $Neg(s_r)$:

"The ruler on the table is not straight".

But if the process of invariance identification shouldn't terminate, that is, should be infinite on the given conditions, then s_r should be undefined, respectively undefinable according as an appropriate extension of the interpretation of the system's language(2) could, respectively, couldn't succeed in reducing the infinite process to a finite one. The sentence s_r expresses as its meaning the procedure that consists of the addition of Plato's identification operation of the invariant "straight segment" to the class C of transformations with respect to which the ruler on the table is an invariant. So the *identification of truth-invariance* consists of the determination of the finiteness and non-destructivity of the meaning of a sentence, that is, of a process that adds or removes, invariance and identity establishing, procedures.

The liar sentence s_l will appear to be, not only undefined, but even undefinable. Could this be due to its self-referential, actually self-evaluating character? And don't the restrictions, imposed on knowledge production by a description of S_{KA} that has to make explicit all assumptions, eliminate the risk of the occurrence of paradoxical, contradictory or undefinable, self-evaluating sentence s_l in a theory generated by S_{KA}? For all factors that are useless, not to speak of those that are detrimental to the validity of knowledge, will have to be excluded from the conception of S_{KA}, whenever possible. At any rate, since knowledge acquisition requires *self-adjustment*, a system S_{KA} will have to evaluate the truthvalue of the sentences it produces by means of inference, that is, of its induc-

(2) S. KRIPKE, Outline of a Theory of Truth, *The Journal of Philosophy*, December 1975, p. 699 sq.

tively or deductively machine inferred sentences. Inference rules that would introduce false sentences in a theory, will have to be deleted by meta-rules as the one considered by EURISKO. Consequently, meta-rules and sentences that evaluate sentences, cannot be avoided by S_{KA}. It cannot avoid *self-reference*, since it has to refer to a model of itself constructed by itself, that is, of its own system of representations and procedures, in order to adjust them. The very concept of knowledge acquiring system itself seems to be paradoxical in the following sense: how could S_{KA} improve its problem solving means, since it seems that it has to improve them using these same clearly deficient problem solving means? Apparently, the system S_{KA} has to solve the problem of an increase of its problem solving capacity by an application of optimally problem solving skill, of which it doesn't dispose.

What is the nature of skill to increase skill? How is it acquired? For, with a view to the conception of a knowledge acquiring system, skill to increase skill is only interesting, if it is possible to acquire it as the result of a problem solving process, that is, if the problem of the increase of problem solving capacity can be formulated as a – partially – solvable problem. A fundamental component of the problem solving capacity to increase problem solving capacity is, accordingly, the *transposition* of ill defined or "unsolvable" problems into partially solvable ones, that is, *into so called puzzle problems*. This transposition requires the construction of a model of the domain or system where the problem arises. A solution of the problem should be planned on the model, that is, every step of the problem solving process should be decidable with respect to that model. If so, then the problem is solvable with respect to that model, that is, the applicability of the proposed solution to reality depends on the adequacy of the model. But how should one tackle the – surely? – ill defined, that is, unsolvable problem of model construction? For example, the classical conception of model appears to be unable to give an adequate representation of the events on the level of quanta. This kind of problem, viz. the adjustment of the type of model, requires a mathematical theory of the connection between structures that represent types of models, and the corresponding types of transposition problems they solve – the transposition problem is, as explained, the problem of the conversion of types of unsolvable problems into partially solvable ones. Such a mathematical theory would be at the basis of the conception of a *general goal directed problem solving system* that elaborates, progressively, a representation of its own final and optimal state. Its self-adjustment rests on maximal learning ability and is directed towards the achievement of the system's final, *optimal* state. The conception of an optimal and final state urges the system to develop *linearly*, whereas the maximal adaptability, implied by maximal learning ability, compels the same system to normalize, so to speak, *deviations*. Both aspects of self-adjustment are, strictly speaking, incompatible, as tendencies to *generalization*, favouring linear development, and to *particularization* that augments the possibilities of deviation, that is, of refutation.

To what does S_{KA}, conceived along the lines of a general goal directed problem solving system, refer when it refers to itself? The answer is not at all simple. As a dynamic system, S_{KA} can, at each instant, be in another state. So, when it refers to itself at instant t_i, and is, at t_i, in state q_j, S_{KA} should, strictly speaking, refer to S_{KA} that refers to itself in state q_j at t_i. Suppose that S_{KA} succeeds in doing this at t_{i+r}, that is, after the lapse of time it takes to realize the process of self-reference, when it is in state q_h. Then S_{KA}, at t_{i+r} in q_h, refers to S_{KA} which is in q_j at t_i referring to S_{KA} at t_i in q_j referring to ....

The conclusion appears to be that S_{KA} cannot refer to itself, at least not in the direct way proposed above, when self-reference is considered as a process taking time and implying possibly a change of state. In order that $S_{KA}(t_i, q_j)$ should refer to $S_{KA}(t_i, q_j)$, time has to be left out of account, otherwise $S_{KA}(t_i, q_j)$ is referred to by $S_{KA}(t_{i+r}, q_h)$. But how could $S_{KA}(t_{i+r}, q_h)$ refer to $S_{KA}(t_i, q_j)$ as a phase of its *own* development, unless it knows what belongs to that development, that is, unless it knows itself. Thus self-reference appears to require self-knowledge, that is, the knowledge necessary to the identification of itself and, consequently, to self-reference. But how should a system acquire self-knowledge, unless by self-reference?

Self-reference is aquired progressively as the result of the identification of phases $S^n_{KA}(t, q)$ of a process that is known or supposed to lead to the realization of an ideal system in optimal state S^L_{KA}, as own phases $S_{KA}(t_i, q_j)$. The discrepancies between a system's real state and its supposed state – they concern, generally, details or proper parts of that state, not the system's overall state – are tracked down as the causes of the system's failures, for example shortage of memory. The systematic representation of causes or conditions of failure or success results, eventually, into a model the system has constructed of itself. Accordingly, S_{KA}

refers to itself as being in a particular state at some time, $S_{KA}(t_i, q_j)$, through the intermediary of a model of itself as a general procedure that generates an approximation sequence $\{S^n_{KA}\}$ converging to S^L_{KA} as its limit case that represents its optimal, final state.

Thus self-reference, and self-evaluation are unavoidable in S_{KA}. They require the conception by S_{KA} of a final, optimal state, the determination of which depends on the content of a mathematical theory of the connection between models and transposition problems.

2. *Self-reference in a model*

When confronted with change, scientists, such as Galileo, appear to proceed as follows. They try to delimit subclasses of changing phenomena, e.g. sequences of velocities of falling bodies, and to define transformations, or their corresponding relations, with respect to which the subclasses considered are closed. Before 1604, Galileo thought that the said sequences of velocities converged to a constant velocity, that is, he tried to determine an invariant with respect to transformations undergone by the elements of the converging sequences. Later they were found to be closed, that is, invariant with respect to constant acceleration. In order that invariants should be applicable to changing reality – for solutions conceived and planned within the model's representation of reality have to be applicable to that reality – they have to be conceived as invariants with respect to time or space, that is, as *events* that are thought of as partial functions with instants t or points p as arguments – such as w(t, p, b) that denotes the weight of body *b* at instant *t* in space point *p* – the values of which measure or, at least, ascertain the phenomena of some changing aspect of reality, viz. the successive weights of the body concerned at different instants or places. So a model is the result of the organization of events into a mathematical structure.

What forms does self-reference take on in a model? What are its consequences with respect to the constitution of events, that is, of partial functions which represent phenomena as functions of time or space? Let us consider time. An invariant, certainly, but also an event? In a classical model, time is, generally, an independent parameter and, consequently, an invariant applied to reality, an event. According to McTaggart (3), the

(3) E. McTaggart, The unreality of time, *Mind*, 17, 1908, pp. 457-474 cf. L. Löfgren, Autology of Time, forthcoming in *Int. J. of General Systems.*

application of time to reality produces a contradiction, from which he concludes that time is unreal. In his analysis time is characterized by – it can also be said that the entity time is invariant with respect to – (1) a *linear ordering* of "instants" t_i, and (2) the *past-present-future classification*. Well past, present, and future are incompatible determinations, though each t_i must have them all, that is, t_i or rather event $e(t_i, x)$, for example the determination of the weight of b at instant t_i, represented symbolically as $w(t_i, b)$, must be past, present and future, which is contradictory. One is, of course, tempted to object that an event $e(t_i, x)$, can have them all, but not at the same time. Thus, to eliminate the contradiction due to the – simultaneous – applications of the past-present-future determination, it suffizes to make an appeal to the first component of the characterization of time, viz. the linear ordening (1): each past-present-future determination D of an event $e(t_i, x)$ has to be connected with an element t_j of the linearly ordered time set T. For example,

> at t_j a person P characterizes $w(t_i, b)$, viz. the weight of body b at t_i, as belonging to the future, and at t_{j+r} as belonging to the past, if $j < i$ and $j+r > i$.

The past-present-future determination by P is, accordingly, itself representable as a time-function, $D(t_l, f_t)$ – where f_t is itself a time-function of the form $f(t, \text{-})$ – that is, an invariant, expressing a dependency on time of, in the case at hand, the "values" past, present or future. For example,

a. $D(t_j, w(t_i, b))$ = the *present* or instantaneaous weight of b, if $i=j$,
= a *past* weight of b, viz. the one at t_i, if $j > i$,
= a *future* weight of b, viz. the one that is predicted for or will be measured at t_i, if $j < i$,

b. $D(t_i, D(t_j, w(t_h, b)))$ = a weight of b predicted in the *past*, if $i > j < h$.

The representability of the past-present-future determination as a function of time, in the acceptance of a linearly ordered set T, amounts to the elimination of the time contradiction: not an event $e(t_i, x)$, but an ordered pair, consisting of an instant t_j and an event, viz. the pair $(t_j, e(t_i, x))$, is related to a characteristic past, present or future, and the same pair cannot have more than one of these characteristics, as can be seen in example (a.). From the same example, it follows also that the past-present-future determination is definable by means of the first component of the characterization of time, viz. the linear ordering (1), which is, consequently, on itself sufficient to characterize time, at least in the framework of McTaggart's analysis. The threefold characterization of time (2) isn't needed as a second component in the characterization of time as an invariant. Consequently, the self-reference, as pointed out by McTaggart, is eliminated. It reappears, however, as soon as one attempts to consider t_i as an invariant. One could, for example, propose to define t_i as a class of simultaneous events, in which case t_i would be a class that is invariant with respect to simultaneity. Thus the characterization of t_i would refer to time through the relation of simultaneity between events. But aren't there other ways to establish invariance? The question: what are the *possible ways to determine invariants*? should be answered in order to tackle the problem of unavoidability of self-reference in particular cases. At any rate, since self-reference appears, generally speaking, to be unavoidable in knowledge acquisition, the fundamental question concerns its connection with contradiction: on what conditions does a *connection between self-reference and contradiction* depend?

McTaggart's time contradiction is not produced by self-reference. On the contrary, it is eliminated at the cost of an introduction of self-reference: an event is past, present, and future, but not at the same *time*. This seeming time contradiction is a result of the use of a non-invariant, that is, a non-existent, in the characterization of time as an invariant entity, viz. the use of component (2), an *indefinite* collection of past-present-future classifications. This undefined collection, of which no invariant aspect is specified, has been converted into a function D, mapping the ordered pairs of the form (t_i, f_t) into the set {past, present, future}, where f_t is a time function $f_t(t_j, x)$, that is, the non-invariant "collection" has been replaced by a class of triples (t_i, f_t, c) with c = past, present, or future. This class is an invariant with respect to the general procedure to determine those triples, as illustrated by example (a.). Thus it is invariant with respect to

i. the course of time,
ii. the form of functional relation,

that is, if, in the triple $(t_i, f_t(t_j, x), c)$ that belongs to the said class, t_i is replaced by t_r and t_j by t_s, then, if c is replaced by c′ according to the conditions formulated in example (a.) and that concern the relation between t_r and t_s, the resulting triple $(t_r, f_t(t_s, x), c')$ is also an element of that same class, that is, that class is not changed by adding the last triple to it. For example,

c. $(t_r, w(t_s, b), \text{past})$, where $r > s$.

Since that class of triples is explicitly definable from the elements t_i of the time set T, which is component (1) of McTaggart's time characterization, together with the concepts of function, number and relations between numbers, there is no self-reference of time, unless one tries to pin down the t_i's as invariants in some respect. This, however, isn't necessary as far as the characterization of T as a set which is closed, that is, invariant, with respect to a linear or some other kind of ordening, is concerned. A requirement of completeness of an – axiomatic – characterization actually amounts to a twofold invariance claim, viz. invariance both of a set of elements, e.g. T, and of the elements themselves, that is, of the t_i. For if F (t_i) as well as Neg(F(t_i)) could be considered for validity, then F would be an aspect, with respect to which T isn't closed or invariant – unless of course, one of both is added to the set of axioms of a theory of T as of its domain of objects, that is, if it is added to the characterization of T's invariance aspects. It depends on the riches or complexity of the language used whether or not some particular invariance aspects can be considered. A language rich enough to provide an infinite number of specifications – denoting an infinite number of elements, e.g. numbers or t_i's, that have to be characterized as distinct invariants, each of them by means of a definite subset of the nondenumerable infinity of possible aspects F – is provably incomplete. Contradiction only follows, when an incorrect supposition is made, e.g. if a procedure that isn't finitely computable is supposed to terminate as in the case of the liar sentence s_1.

Since the application of invariants of a model to reality doesn't appear to require the introduction of another factor beside invariance – the application can be represented by means of time or space functions which are invariants – the consideration of the impact of self-reference on existence and validity in the knowledge acquiring system S_{KA} and its models can be limited to forms of invariance or the absence thereof.

The specific type of self-reference in a model is unavoidable self-parametrization. Disposing of a model of itself S_{KA} can enter this representation of itself, as a component, in a system that models a domain of changing phenomena. S_{KA} has, accordingly, to conceive itself as a time function whose values, viz. states of S_{KA}, depend on the elements t_i of a set T: changing in time, S_{KA} has to represent itself as changing according to a specific time conception. If, according to S_{KA}'s conception of itself, its state at t_i is st(t_i, S_{KA}), then S_{KA} has to know what t_i refers to, in order to confront its result with observation. But how has observation to be conceived in order to correct the time conception on which S_{KA}'s models rest? This seems also to require the conception of an approximation sequence $\{S^n_{KA}\}$ converging to S^L_{KA} as the limit case, where an optimal time conception, and, in general, an optimal parametrization is conceived.

3. *Self-reference, invariance and validity*

Self-reference requires (1) *identification* of an entity E, the reference's referent, which exists, that is, is (2) *invariant* with respect to a definite class C of transformations – or their corresponding relations – that identify E and determine its invariance. By this description, however, only reference to an entity is ascertained, not yet self-reference. The self-referential character of the reference is the result of the occurrence of E in at least one of the procedures or relations that belong to the characterization class C. As a consequence, E is encountered in the course of the identification of E by applying the procedure wherein E occurs, and the whole identification process, without being terminated, has to be resumed from the beginning. The same applies to the second aspect of the reference, viz. the determination of the invariance of E with respect to that same procedure. The incorrect supposition that E is identified or that its invariance is established with respect to the procedure wherein E occurs, gives rise to a contradiction.

Does self-reference play a role in the generation of contradiction? Let us consider again the version of the liar sentence s_1, proposed in the first paragraph.

> the machine inferred sentence s_1 reads: if a sentence is machine inferred, then it is false.

By the "if...then..." form of this sentence s_1 the easily (?) verifiable truth of

s_1 is machine inferred

should be transferred to the consequent

s_1 is false,

at least, if, as explained, validity would be assignable to that consequent, that is, the meaning of s_1 has to be *finite* and – since s_1 is said to be false – *destructive*. In order to verify both characteristics, the expressions in s_1, which refer to invariants and classes of invariance and identity establishing procedures, have to be known. One of these expressions is "machine inferred" that refers to a class of procedures C_{MI} with respect to which "s_1 is machine inferred" is easely verified: let C_s be the class of procedures with respect to which s_1 as a series of signs is invariant; then s_1 will also be invariant with respect to $C_s \cup C_{MI}$. Next we turn to the second part, viz. "s_1 is false". To verify this sentence, first the expression "s_1" has to be decoded and its referent determined. This referent is said to be false. Consequently, we must determine the *finiteness* and *destructivity* of the process that has been called "the meaning of a sentence". So we are back at the beginning of the process, viz. the verification of these both characteristics of the meaning of s_1.
The supposition that the truth of

s_1 is machine inferred

is really transferred by the conditional sentence s_1 to

s_1 is false,

amounts to the supposition that the process of establishing the truthvalue of s_1 is indeed finite and destructive. This supposition produces the contradiction derived in the first paragraph:

s_1 is true iff it is false.

The role of self-reference in the generation of that contradiction is only indirect: it doesn't allow another way to determine validity invariance. For example, the axiomatization of Leibniz' infinitesimal calculus by l'Hospital was contradictory: an infinitesimal was considered different from zero, $e \neq 0$, but, at the same time, it didn't make any difference when added to a real number r: r+e=r. The contradiction was due to the assimilation of infinitesimals to reals, and could, accordingly, be eliminated by their disengagement. In the case of the liar sentence, however, self-reference makes disengagemend impossible: s_1 cannot be disengaged from itself. On the other hand, it eliminates anticipation – viz by the said "incorrect" supposition – of a result of validity determination, a process of which the initial part has to be iterated indefinitely through the intermediary of self-reference: the process of invariance determination doesn't terminate.

Can the supposition that brings the contradiction about, be shown to be incorrect? But that supposition, viz. that s_1 has a truthvalue, is expressed by s_1 itself and s_1 has been inferred by S_{KA} presumably on good grounds. At any rate, S_{KA}'s conclusion that all machine inferred sentences are false, that is, that s_1 is true, outreaches its inferring capacity. But so does every general statement without necessarily implying a contradiction. Well, s_1 as such isn't contradictory: only its application to s_1 resulting into the sentence "s_1 is false" leads to the known contradiction. One may object that this isn't relevant, since that "application" is an application of a rule of deduction, viz. instantiation, which makes s_1 "as such" also contradictory. However, the sentence s_1 is – presumably – "empirically founded": S_{KA} had at its disposal data from which it had to infer, according to its own criteria, that machine inferred sentences are false. The same cannot be said of the sentence "s_1 is false". If "s_1 is false" were only deduced by instantiation from the empirically founded sentence s_1, it had also to be empirically founded. To obtain "s_1 is false", however, *diagonalization* had to be applied and this is a procedure with respect to which non-denumerable classes are invariant, but never denumerable ones.

Let

(1) $s_1, s_2, s_3, \ldots$

be a sequence of machine inferred sentences, that is, sentences inferred by S_{KA}. Processing their meaning, S_{KA} had to conclude, so far, that they don't hold, e.g. they are no match for a confrontation with observational results. At any rate, the sense of such a sentence, s_3 say, connects invariant inv_3 with a class C_3 of – invariance and identification – procedures that destroy inv_3, that is, their connection has to be dissolved, $Neg[C_3(inv_3)]$, and s_3 is said to be false, $C_F(s_3)$:

(2) $Neg[C_3(inv_3)]$ and $C_F(s_3)$.

From the iteration of that kind of experience, S_{KA} infers that all machine inferred sentences are false:

(3) $C_F(s_n)$, where n ranges over the indices 1,2,3... of machine inferred sentences.

Is the sentence $(n)C_F(s_n)$, viz. that all machine inferred sentences are false, itself machine inferred and should it, consequently be added to list (1)?
According to the symbolic representation of the first paragraph:

(4) $(n)C_F(s_n) = s_1$,

and s_1 is inferred from the expressions determined in (2):

(5) $C_F(s_1)$, $C_F(s_2)$, ...,

which are actually *corrections* of the machine inferred sentences of list (1). The list (5) can be considered also as the result of instantiation from s_1, and then $C_F(s_1)$ also belongs to list (5), which is the result, together with s_1, of a diagonalization procedure. To consider the result of diagonalization as belonging to list (1) would make this list non-denumerable, whereas the class of machine inferred sentences cannot be but denumerable. The notion of machine inferred sentence seems, in many respects, to be non-invariant. For example, it should allow us to decide whether or not corrections of such sentences are also machine inferred, which, so far, it doesn't, though it can be said that data that had to be corrected, aren't data anymore. And, consequently, the sentences of list (1) aren't anymore machine inferred, once they have been corrected, that is, negated. If this could be formalized, one could propose to explain "machine inferred" as "produced by means of S_{KA}'s inference rules and data formulating sentences". The conception of invariance and its conditions appear to supply the material of criteria formulation, by which S_{KA} had to decide on the necessity of an adjustment. The adequacy of an adjustment depends not only on the observation that some kind of invariance is annihilated in some particular conditions, as, for example, denumerability with respect to diagonalization, but, to a far greater extent, on the knowledge of the constitution of such an invariant, or its "coming into being". Validity, for example, appears to be the outcome – or, at any rate, is, to an appreciable extent, representable as such – of an acceptance procedure which can be represented as a program computing a functional (V,f_k,r,s) where r measured the tolerance or "degree of scientific character" of a set of theorems, V, and s the uncertainty due to the openness of the system. This means that inaccurate measuring or identification procedures could disturb their object up to s. Then p is accepted as a true or valid statement, element of V, if $f_k(p) \geq r+s$, where $f_k(p)$ is an element of the real interval [1,0]. The smaller r is, the greater is the fundamental uncertainty due to an inadequate representation of some entities of the physical system in the model. As r approaches 1 and s approaches 0, without endangering the consistency of the theory, the adequacy of that model augments. The functions f_k have to be conceived as operations on classes of the so-called invariance and identity establishing procedures. A calculus of such operations is a first requirement that S_{KA} has to realize, be it, to begin with, only as a rather primitive system: S_{KA} has to have at its disposal the means to elaborate it.

Rijksuniversiteit Gent A. PHALET
Hertogstraat 243
3030 Leuven