Linguistics Research Institute
Report No. 41

PUTTING CONCEPTS IN THE CONCEPTUAL SCHEMA

Lowell S. Schneider
1985

Linguistics Research Institute
Report No. 41

PUTTING CONCEPTS IN THE CONCEPTUAL SCHEMA

Lowell S. Schneider
1985

Abstract*

This paper the addresses the general problem of the conceptual schema in database systems. It concludes that a fundamentally new approach to data models that treats concepts as concepts rather than as something else is clearly required. The State of Affairs System is evaluated in this regard, and found to be indicated for several of the acknowledged problems in data modeling.

Additional copies of this report may be obtained from:

Linguistic Research Institute (Publications)

P.O. Box 1294

Boulder, Colorado 80306

*This research was supported by the United States Air Force, Rome Air Development Center, Post Doctoral Program, Contract Number F30602-81-C-0205, Griffiss AFB, NY.

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

The objectives of the research reported herein were:

(1) to evaluate data modelling in light of the State of Affairs System; and

(2) to evaluate data representations in light of the Data Independent Accessing Model (DIAM I).

Due to the limitations of time and budget, only the first was achieved.

The evaluation of data modelling wa performed in two phases. The first was to reformulate the traditional problems of data modelling as explicated by Kent [1]. The reformulation was accomplished by stating these problems in terms of the State of Affairs System explicated by Ossorio [2]. The results of this phase were originally reported in a viewgraph presentation midway during the research. The presentation exposed two significant continuations (not necessarily exclusive) worth pursuing:

(1) developing a new data modelling technology based directly on the State of Affairs System itself; and

(2) using the State of Affairs System as a Precaution Paradigm in the development of existing data models.

Of these, (1) was pursued to the point of demonstrating that such a model could, in principle, solve virtually all of the fundamental problems exposed by Kent. This demonstration is the main body of Chapters 2 - 12. It argues, in general, that data modelling as presently formulated fails fundamentally in not explicitly recognizing "concepts" in the "conceptual schema." Specifically, it points out that existing data models could be vastly improved by:

A. replacing the concept of "definition" with "paradigms";

B. replacing the concept of "time" with "ex post facto formulation";

C. replacing the concept of "unique identifiers" with identity co-ordination";

D. making a clear distinction between "elements", "individuals", and "eligibilities";

E. providing for unlimited part-whole relationships (invoking "limiting cases" only when necessary to avoid indefinite recursion);

F. relying to a much greater degree on what the user already knows and, therefore, does not have to be

"known" by the model. (E.g., a rule processor can correctly process a term such as "has attribute x" as both an antecedent and a consequent without having to know anything about "has" and "x" since it can rely on the user's competence to distinguish among legitimate and nonsensical interpretations of he term. The system need only know about the syntax - which, in this case, is implication.)

A methodological approach to making these improvements is not implicated. The reason for this is that making such improvements is not a case of mere refinement. It involves a fundamental change in the foundation-level theory on which these models have been based. Existing models, without exception, are founded on "mathematical" logic while the approach recommended is founded on "behavioral" logic. As such, the recommendation is simply a more detailed exposition of the position taken by the late Dr. Michael E. Senko in that: "...we always have the choice of making a data processing system behave like a mathematical formalism; or making it behave the way we would like it to. ...in the latter case, there usually exists a mathematics into which the system can be mapped or, if not, we can invent one! [3]"

The second continuation exposed in the presentation was a critical analysis of existing data modelling technology. It is this that constitutes the main body of Chapters 13 - 18. It is

written, in fact, using the presentation itself as the outline. The motivation for this approach is that the presentation parallels the Kent text which, in turn, systematically addresses the limitations of present data models. It proceeds from a fundamental discussion of data and the issues traditionally addressed by General Semantics. It then examines the practice of Data Processing as the context in which such fundamental issues have a place. Finally, it reconsiders the basic assumptions in data modelling technology that have arisen from this context: first, in terms of the inevitable paradoxes these assumptions guarantee; second, by redescribing them as a set of restrictions on the State of Affairs Model which makes no such assumptions). The net result is to show that in virtually all of the cases exposed by Kent, the referenced assumptions aren't assumptions at all, but merely self-imposed constraints that confine the behavior of data models tot he well-understood subset of relational mathematics; i.e., first-order logical theory. Moreover, since this is not a mathematical treatise, but rather a discussion of these constraints in terms of ordinary, everyday real-world behaviors and social practices, it soon becomes apparent that from this standpoint, the constraints are an ad-hoc collection of "fingers in the dike" of complexity, as opposed to the consistent system perceived by mathematicians. It therefore serves two purposes simultaneously: (1) a redescription of relational theory in terms of behavior, and (2) a redescription of State of Affairs in terms of computer science. If it achieves

the first, it has satisfied the goal of precaution paradigm. If it achieves the second, it will hopefully persuade more than a few technologists that behavior is, at least, on a par with mathematics in is reference to computer science.

Finally, as a precautionary note to the reader, familiarity with the referenced texts by Kent and by Ossorio is presumed although a "first" reading can be accomplished without this.

2. THE PROBLEM WITH CONCEPTUAL SCHEMAS

Of late, a number of criticisms have been directed at the state-of-the-art in data modelling. These range from highly technical issues such as the failure of normalization [4] to more sweeping examples of what data models fail to do [5], and extending [6][7], and improving [8] existing approaches to data modelling. If the critics have failed, it is primarily because they offer no fundamentally different alternatives. Hence the research has taken on a distinctly lateral quality in that the fundamental approach remains the same. In this report we level what is perhaps the most severe criticism of all. Then we present a fundamentally new alternative without the failings we criticize.

We recognize that it is virtually impossible to define or reify a concept and any attempt to do so for the concept "concept" would be doubly absurd. But we can talk about concepts in terms of the way they're used, why they're needed, and the consequences of having them. And we can do this in a straight-forward, non-mysterious way. After all, we as persons successfully use concepts in every behavior in which we engage, so we are intimately familiar with the phenomenon. Furthermore, as computer scientists rather than psychologists we can afford to be rather parsimonious so long as the discussion serves our purposes. But engaging in the discussion is unavoidable. We absolute-

ly have to provide at least some insight to the subject of concepts before we can consider how to schematize them.

To introduce a formalism where no such formalism exists merely succeeds in changing the subject. Yet we, as data modelers, have persisted in trying to reduce the original notion of "conceptual schema" to something else: relations, entity sets, predicate logic, set theory, ad infinitum. If the terminology "conceptual schema"; i.e., a schema of concepts, was not accidental then what we have done is to lose the subject entirely. There is nothing even remotely like a concept to which we can reduce the concept "concept" so why continue to try? In this report we don't. We recognize that, to ultimately achieve a true conceptual schema, we must start by dealing with concepts as concepts and not as something else.

3. DISTINCTIONS

Concepts enable persons to distinguish one thing from another and, ultimately, to act on those distinctions. An excellent paradigm is Kent's use of Goguen's continuum "between some given chair and table, constructed by letting the chair back shrink while its seat expands and flattens, and its legs become higher." Clearly, without the concept "chair" and the concept "table" it is not merely that we couldn't tell which it was: it would be literally impossible to ask the question that way. We might legitimately ask what shape the wood was (provided we had the concept "shape" and "wood") but whatever it was, it couldn't possibly be a chair or a table unless we already had those concepts.

If Kent's use of the example failed, it was primarily because he didn't pursue it to the point of acting on the distinction. in that fuzzy middle ground there isn't any ambiguity as to what it is because "chair" and "table" are behavioral concepts, not physical phenomenon. Any person encountering it will act on his distinction by attempting to treat it as a chair (by sitting on it) or a table (by putting things on it). And if he succeeds at one or the other, he will take it to be the case that that's what it is. If he is uncertain and successfully treats it as both or fails to treat it as either, he may conclude that it's a combination chair/table, a useless piece of furniture, an art object or

any number of things but he certainly won't deny that it exists. If another person takes it to be a different case then this is simply an exemplification of Kent's point that "Things exist in the database because they exist in people's minds independent of any physical existence. Therefore, we very much have to deal with the fact that (concepts) may exist differently in different people's minds."

4. JUDGEMENT

In contrast, it is fortunately the case that by and large, despite minor differences, persons do share the same concepts. It is by virtue of this that we can engage in social practices or enterprises (the things we keep insisting the data base is supposed to model). It is because we do share the same concept of chair and table that we are able to argue or make judgments or disagree about Goguen's continuum. And in general, persons in a social practice do disagree and engage in negotiations to resolve their differences. When the social practice is a highly specialized enterprise such as intelligence, there are merely more and better refined concepts being shared which consequently form the basis for more and better refined judgments and negotiations. In turn, this is the principle reason for having databases at all. If we knew we would never disagree, ergo never had to make judgments, then why would we bother to collect information to support those judgments? But this makes it apparent that databases are as much a social as a technical phenomenon because it is embedded in the same enterprise as its users and it too must share the same concepts to make any useful contribution. While this may make the data base system appear to be somewhat of another "person" in the enterprise, it is definitely not intended as an argument for artificial intelligence as a future requirement. To the contrary, the database is already "intelligent" and does, in fact, behave in ways that a person does.

The automated intelligence system is a more complete "person" by virtue of having a schema of concepts that the enterprise shares. That it acquired these in a direct, programmed way instead of by some mysterious learning mechanism we don't yet understand does not lessen its contribution. For the purposes at hand it makes no difference.

5. THE PARAMETER PRINCIPLE

Inherent in both distinction and judgement is the fact that one very important function of concepts is to establish boundaries on the ways in which something can change. We could change Gougen's chair by lowering the height of its back because chairs have backs, and the backs are measurable in terms of their heights. Admittedly, there are bounds on the value of that height parameter: it can't possibly be negative; it probably can't be zero (for then it wouldn't have a back and we would call it a stool); and anything over six feet, while possible, is also a little absurd (for a chair; not necessarily for a throne). But more important are the things you can't possibly change about it:

a) you can't change its fuel economy (it doesn't use fuel);

b) you can't change its f-stop (it's not a lens); and

c) you can't change its truth value (it's not a proposition).

Ergo, you can't change a chair into the color red any more than you can change the number 17 into a banana. They are different concepts entirely.

Yet, it is amazing how much effort we expend on attempting such feats and dealing with the paradoxical situations that result. Appealing again to one of Kent's classic examples:

> "Sometimes two distinct entities are eventually determined to be the same one, perhaps after we have accumulated substantial amounts of information about each. At the beginning of a mystery, we need to think of the murderer and the butler as two distinct entities, collecting information about each separately. After we discover that 'the butler did it', have we established that they are 'the same entity'? Shall we require the modelling system to collapse their two representations into one (in order to maintain 1:1 correspondence between records in the database and entities in reality)? I don't know of any modelling system which can cope with that adequately." [9]

The reason Kent doesn't know of such a modelling system that can cope with this situation is that, the way he described it, the situation itself is not possible in reality. There is nothing you can do to change the concept 'butler' into the concept 'murderer'. Thus, the question is not "which do you

sacrifice (when you find out that the butler did it) to maintain 1:1 correspondence with the entities?" Rather, it is "what other concepts such as the process "murder", the object "person", the object "butler", the fact that "the butler did it" and the event that "we discovered that the butler did it" do you have to add to make the whole state of affairs even a possibility in reality?"

6. EX POST FACTO FORMULATION

When we attempt to violate the Parameter Principle (as just described in Section 4, Parameter Principle) we begin to see the real consequences of having concepts at all, let alone trying to schematize them. The consequences of creating conceptual schemas "after the fact" as we typically do in database management, lead to confusion as articulated by Ossorio [10].

> "The most dramatic examples of what happens when we attempt to violate the Parameter Principle typically occur in the form of origin questions. 'Where did persons come from? Where did language come from? Where did the world come from?' What we're really asking in these cases is the general formulation: 'What was it, X, that changed into Y?' And it is this kind of formulation that inevitably leads to the 'missing link'. There is nothing remotely like language that wasn't already language that could have possibly changed into language (the Parameter Principle). If we are to avoid this inevitable paradox then something has to give. Fortunately, something can give and it's best illustrated through a series of progressively more provocative examples starting with something rather tame on the order of Kent's 'murder and the butler' problem.

"(a) We're sitting in the stadium at 1:30 in the afternoon; the referee blows the whistle; and the ball is kicked into the end-zone. I ask 'What was that?' to which you reply 'That was the first play of the game.' Being in a philosophical mood I say 'Wait a minute; there hasn't been a game yet and if there hasn't been a game, nothing could be the first play of it. At 5 p.m. when the final gun sounds there will have been a game but for now there isn't a game of which that could have been the first play.' Knowing that anything you say at this point will just get you in trouble, you just smile and wait. At 5 p.m. the final gun sounds and you say 'See, I told you it was the first play of the game.'

"The point is that at 5:00 it became the case that at 1:30 what happened was the first play of the game: even though at 1:30, when it did happen, it wasn't already the case. This is not verbal sleight-of- hand. What holds historically for the first play of the football game also holds categorically for any aspect of any game.

"(b) I show you a chess set of which all the pieces are carved out of onyx. I pick up one of the pawns and ask

you what it is to which you obviously reply 'A pawn.' I now tell you that I know for a fact that the set was carved 5000 years ago - which is interesting in as much as the game of chess was only invented about 3000 years ago. In light of that I pose two questions:

(1) 'Was it a pawn when it was carved?'

(2) 'When it was carved, was it a pawn?'

Thinking I'm about to play another philosophical game, you shrug your shoulders and find someone else to talk to.

"The point here is that it's not a philosophical game. As we look upon it is a pawn and has been a pawn from way back. But at the time it was carved it couldn't possibly have been a pawn because the game of chess is the only place that the concept 'pawn' has a place and chess didn't exist then. It could have been a lot of other things (e.g., an unusually shaped piece of onyx) but it couldn't possibly have been a pawn. But when chess was invented 2000 years later it became a pawn and has been a pawn ever since.

"(c) You come back after a while and I ask you 'When it was carved, was it a piece of onyx?' Feeling safe you respond affirmatively. Now I tell you another interesting fact - namely that the onyx has been carbon dated as 20 million years old. In light of this I ask 'Was it a piece of onyx 20 million years ago?' You leave again without responding. "The point again is that it could have been a lot of things 20 million years ago but without the conceptual system (geology or mineralogy). When that conceptual system was created by people, it became the case that the carved object and a lot of other pieces of material around the world were onyx. What holds for the pawn holds for onyx. There is a difference, however, because chess and geology are different sorts of conceptual systems and we play different "games" with them. When geology was invented it not only became the case that the carved object was onyx, it became the case that the carved object already was onyx and had been for a long time. Further, since onyx is a type of material we could, in accordance with the Parameter Principle, imagine or invent some other type of material that changed into onyx.

"What holds for onyx holds for everything else in the world, and for the world itself. This may seem to lead

to the conclusion that there wasn't a world before there were people, but that is incorrect, just as it would be incorrect to say that the football game did not begin at 1:30. The correct conclusion is that there wasn't, and couldn't possibly have been, a world before there were people...before there were people.

"What has given way is the explanation of history as a simple temporal succession of events in which persons and concepts were merely accidents (after all, that explanation itself is a concept created by persons). Instead, we have embedded that in a relativistic view in which physical reality (whatever that is) and conceptual reality are very much interdependent."

If it's not already obvious, there is a moral in all this of enormous import to anyone engaged in creating databases, data models and conceptual schemas. The successful creation and systematizing of concepts literally changes the real world. And that consequence is one that needs to be taken very seriously when we presume to tell someone what entities and relationships his enterprise is "nothing more than."

7. IDENTITY COORDINATION

There is another class of questions that bring about paradoxical situations if we persist in ignoring the existence of conceptual reality. Relying again on Kent's examples: "Is Walnut Street in Boulder, CO the same as Walnut Street in Denver, CO?" "Is the Boulder Turnpike, which is also Highway 36, which becomes 28th Street when you get into Boulder, which again becomes Highway 36 (but not the Boulder Turnpike) when you leave Boulder on your way to Lyons, one street, two, three, etc., and where exactly do the changes occur?" What we're really asking in these cases is the general formulation:

> "What is it that two different descriptions of the same thing are descriptions of?"

The paradox is that the answer is an infinite regression. If we ever do discover "what is being described" it, too, is a description and we have to ask:

> "What is it that three different descriptions of the same thing are descriptions of?"

The problem has its heritage in mathematics manifested as "canonical forms" and "deductive guarantees". For the purposes

of intelligence analysis, this is best illustrated by a paraphrase of one of Ossorio's Wil and Gil dialogues [11]:

Wil: The Zylons are preparing to attack the Empire.

Gil: Now wait a minute. You can't just say that - these questions have to be settled factually.

Wil: How would you do that?

Gil: I'd collect information "I".

Wil: So if the question is "Q" - namely, 'Are the Zylons preparing to attack the Empire?' then information "I" will give us the answer. Am I right?

Gil: Yes.

Wil: Now you wait a minute. That's another question just like the first one; namely 'Q1' - "will the information 'I' give us the answer to the question 'Q'? Either you've just violated your own principle or you're going to have to collect information 'I1' to answer question 'Q1'. But if you take that approach, you'll have another question 'Q2' for which you'll have to collect information 'I2' in order to answer, and so on, ad infinitum.

Gil: Quit your philosophical nit-picking. You know what I mean, don't you?

Wil: Gee, I wonder what information we'd need to answer that question.

The point is, of course, the paradox that information cannot possibly be so complete that we have a deductive guarantee for what we do. In that regard we have to accept it as being incomplete. Thus, the mathematically inspired "unique" identifier is a theoretical impossibility. This, in turn, leads to the impossibility of that transcendental "thing" that all the descriptions are descriptions of - it can't be uniquely identified.

What we're left with is the realization that identification can only be accomplished, not proven. Persons describe (identify) things differently because they have different perspectives. The road engineer would have a completely different way of referring to Kent's streets than would someone giving directions to his house. Yet there is no difficulty in giving the directions to your house to a road engineer. How is that possible in the light of the fact that there is not canonical "street" to which we can both refer? It's possible because we, as persons, don't need to refer to anything - we don't depend on a "theory of reference".

Instead we use our concept of the street in question rather than the physical street and then perform the equivalent of a rotation and translation to ascertain that we're both talking about the same thing. (If we have done this correctly it will show up later in the conversation; e.g., "Oh, that's what you meant"). In other words, we have coordinated the reference of the concepts involved rather than uniquely identifying some "thing" as their 'referent'.

8. CATEGORIZATION

Reality doesn't divide into well-defined sets nor does it behave according to universal principles. To use Wittgenstein's words:

"The world divides into facts, not things." [12]

But in the current practice of data modelling, we are constantly struggling with categories, functional dependencies, domains, properties, etc. And this is largely a result of our field's preoccupation with mathematics. Current data models depend on establishing such things a priori because you can't define a function before you define its range; and you can't define an n-ary relationship before you define the participating domains. Thus, as Kent points out, I can own a pencil, and you can own a car, a corporation can own an estate, and these are all perfectly natural paradigms of ownership. But you need three mathematical relationships to describe them because the domains are different. Generalization [13] helps for a while, but it eventually has to fail. You can say that you, me, and corporations are subsets of the domain "owners" and that pencils, cars and estates are subsets of the domain "property". But then you encounter an estate (a property) owning a corporation (an owner). In summary, you simply can't pre-empirically categorize what there is in someone's enterprise. Rather than attempt to struggle any further with these or any of the preceding problems, perhaps it

is time to fulfill the promise made at the beginning of the discussion of conceptual schemas. And categorization is as good a place as any to start.

> "You can describe everything in reality as either objects, processes, events, or states of affairs. It doesn't make much difference which you use - anyone of them will suffice. [14]"

Choosing "states of affairs" for the moment (shortly we will propose that you don't have to make a choice) what we have to deal with is:

(1) some elements (domains if you like);

(2) some individuals (historical particulars);

(3) some eligibilities as to which individuals can participate as which elements.

If you want categorization, you can have it by describing that state of affairs in which the categorization occurs. In fact, you can describe several such states of affairs each of which applies on different occasions. This is empirical categorization that you do when it serves a purpose. It is not pre-empirical policy nor a universal principle. Kent proposes an approach that

takes a step in this direction and he will probably succeed eventually. But there is an approach that already has succeeded. In the following sections we will present it as the precursor of the conceptual schema technology necessary to support a "new generation" database environment.

9. THE STATE OF AFFAIRS SYSTEM

The articulation of the concept of "reality" is accomplished by reference to the four basic reality concepts, namely, "object," "process," "event," and "state of affairs," and their further development.

By way of preliminary examination, we may note that these are not invented technical terms. Rather, they are already straight-forwardly concepts of reality or the real world. A primary and paradigmatic use of these concepts is as the categories of "what there is." Thus, for example, one of the principal ways of formulating the claim that Z's are real is to say that they are a certain kind of object (e.g., a mental object, a mathematical object, an invisible physical object) or a certain kind of process (e.g., a mental process, a submicroscopic process, a learning process), etc.

Also, and by no means unrelated, the four reality concepts are observation concepts - we observe exemplars of each kind. To observe something on a given occasion is (at least) to find out something about it without on that occasion having to find out something else first (observation contrasts with inference). For example, we observe an object when we see an automobile, smell a fish, hear a bell, touch a person, or taste an apple. We observe a process when we hear the automobile coming down the

road, feel the water turning warm, hear the music rising to a climatic pitch, or see the infant bouncing in his crib working himself into a rage. We observe an event when we hear the motor stop, feel the wire snap, or see the flash in the sky. We observe a state of affairs when we hear the singer is off-key, feel that the coat is threadbare, taste the difference between brand X and brand Y, or see that he is overjoyed or that they didn't understand, that the brass instrument is faulty, that the respiration rate has increased, etc.

What we observe is the real world. The fact that some exemplars of each of the four kinds of concept are observable provides one entree to the logical relations among these concepts. For without those relationships our observations would be as unrelated as the number 17, the color orange, and the Day of Judgement; and the very concept of "observation" would be lacking. The fact that our separate observations can be formulated as observations of a single world; i.e., the real world, requires that there be logical relationships among the concepts in terms of which our observations are made and our world described.

The choice of basic reality concepts is by no means arbitrary as we will show shortly. But it is first necessary to introduce their logical relationships as a basis for the discussion. These are expressed as a set of transition rules:

T1) A state of affairs is a totality of related objects and/or processes and/or events and/or states of affairs.

T2) A process (or object or event or state of affairs) is a state of affairs which is a constituent of some other state of affairs.

T3) An object is a state of affairs having other, related objects as immediate constituents. (An object divides into related, smaller objects).

T4) A process is a sequential change from one state of affairs to another.

T5) A process is a state of affairs having other, related processes as immediate constituents. (A process divides into related, sequential or parallel, smaller processes.)

T6) An event is a direct change from one state of affairs to another.

T7) An event is a state of affairs having two states of affairs (i.e., "before" and "after") as constituents.

T8) That a given state of affairs has a given relationship to a second state of affairs is a state of affairs.

T9) That a given object, process, event, or state of affairs is of a given kind is a state of affairs.

T10) That an object or process begins is an event and that it ends is a different event.

T11) That an object or process occurs (begins and ends) is a state of affairs having three states of affairs ("before", "during" and "after") as constituents.

In addition to these, because of their inherent (and necessary) recursiveness, it is necessary to introduce four limiting cases which can be invoked to stop the unlimited decomposition or composition permitted by the rules:

LC1) The state of affairs which includes all other states of affairs (i.e., "the real world").

LC2) An object that has no constituents, hence is an ultimate particle.

LC3) A process that has no constituents, hence no beginning

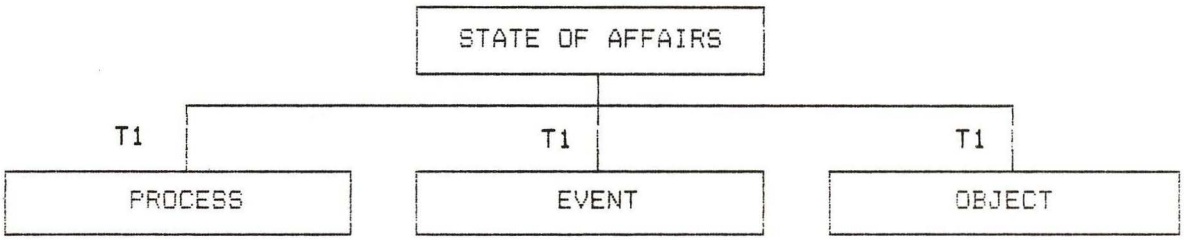
that is distinct from its end; i.e., the effective equivalent of an event.

LC4) An event that has no constituents; hence the effective equivalent of an object during a period during which the object undergoes no change (e.g., molecules at absolute zero) hence also a timeless state of affairs.

10. AN IRREDUCIBLE THEORY

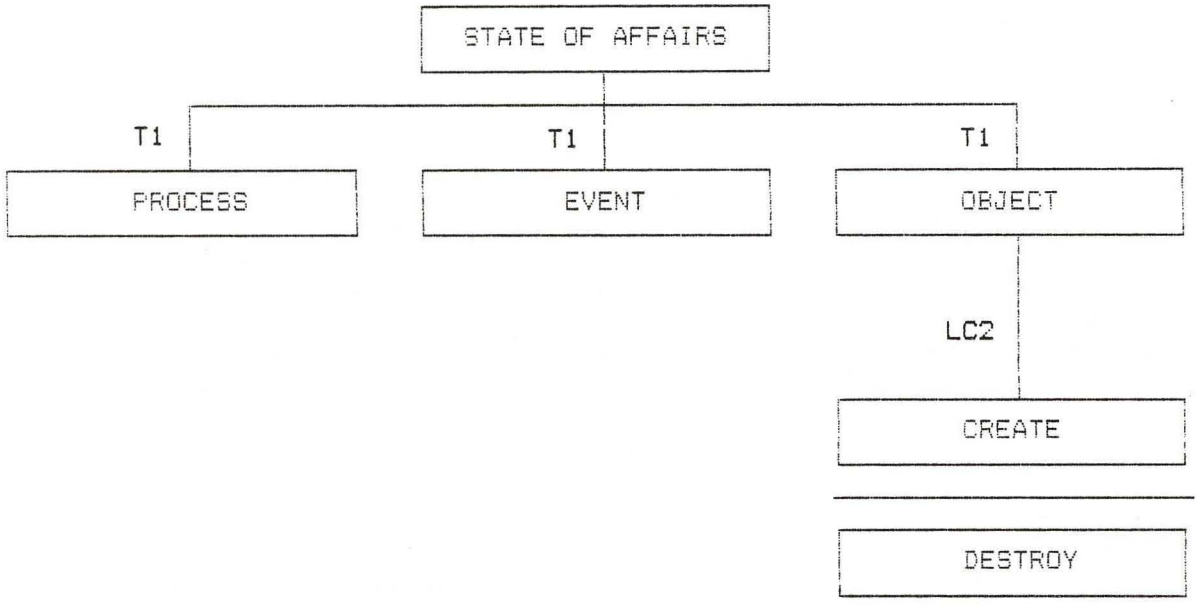
By virtue of the four basic reality concepts and the transition rules, we have a totally non-reductive scheme as we will now illustrate. We begin with a "reality line" (analogous, heuristically, with a real number line). The choice of axis is important only in that the endpoints are exclusive and the line is exhaustive. In the case of State of Affairs (SA) the line is from process (pure change) to object (changelessness):

We add, in addition, some arbitrary point in between for clarity (in fact, it is precisely the midpoint, but that will be established subsequently) which, allows us to depict Transition Rule number one.

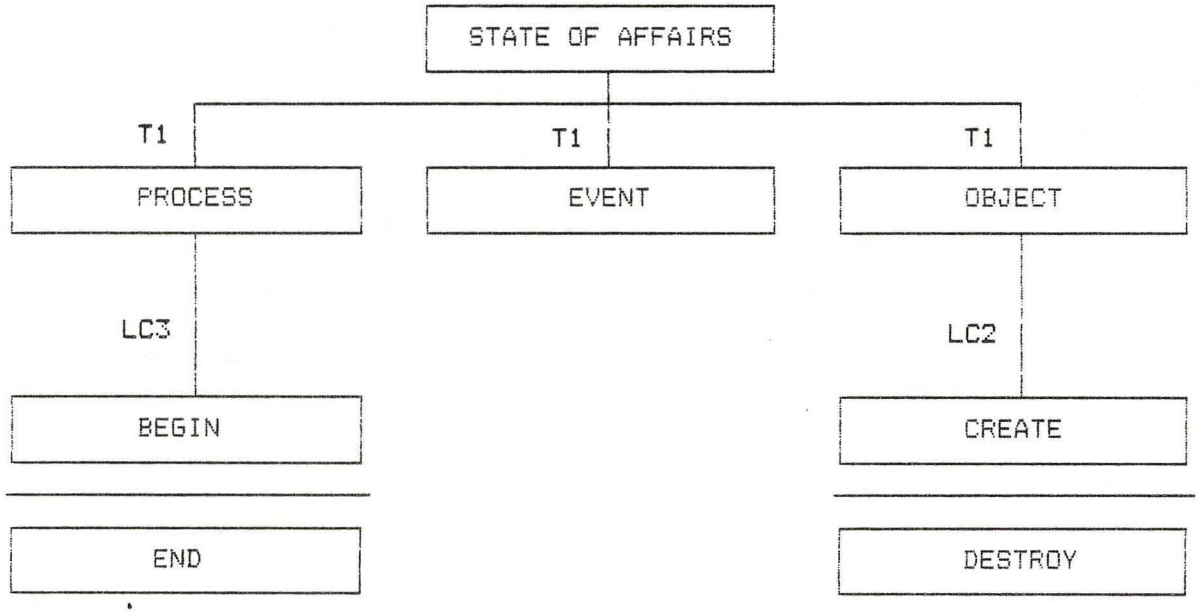


The T1 relation from SA to SA is assumed implicitly: it is reflexive and serves the purpose of acknowledging "depth of field;" i.e., not everything is in focus simultaneously. Proceeding, via LC2, we can decompose an object until it's so

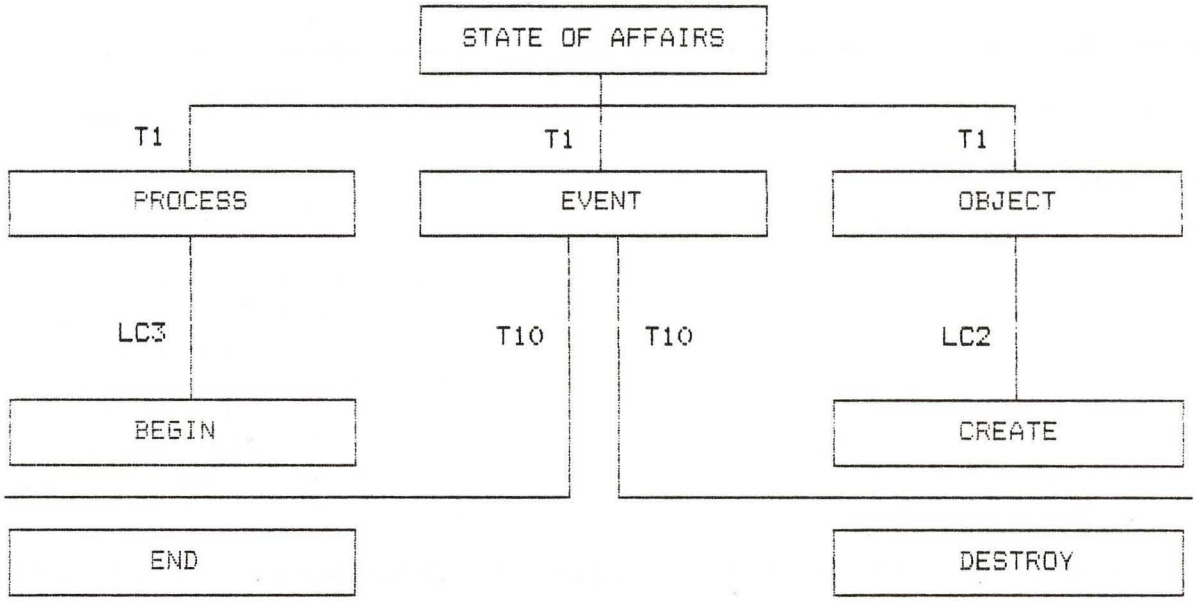
elementary that all you can do is create or destroy it: it has no constituents to rearrange.



We can do the same thing for a process by decomposing it via LC3 into something so elementary that all you can do is begin it or end it: it has no stages.

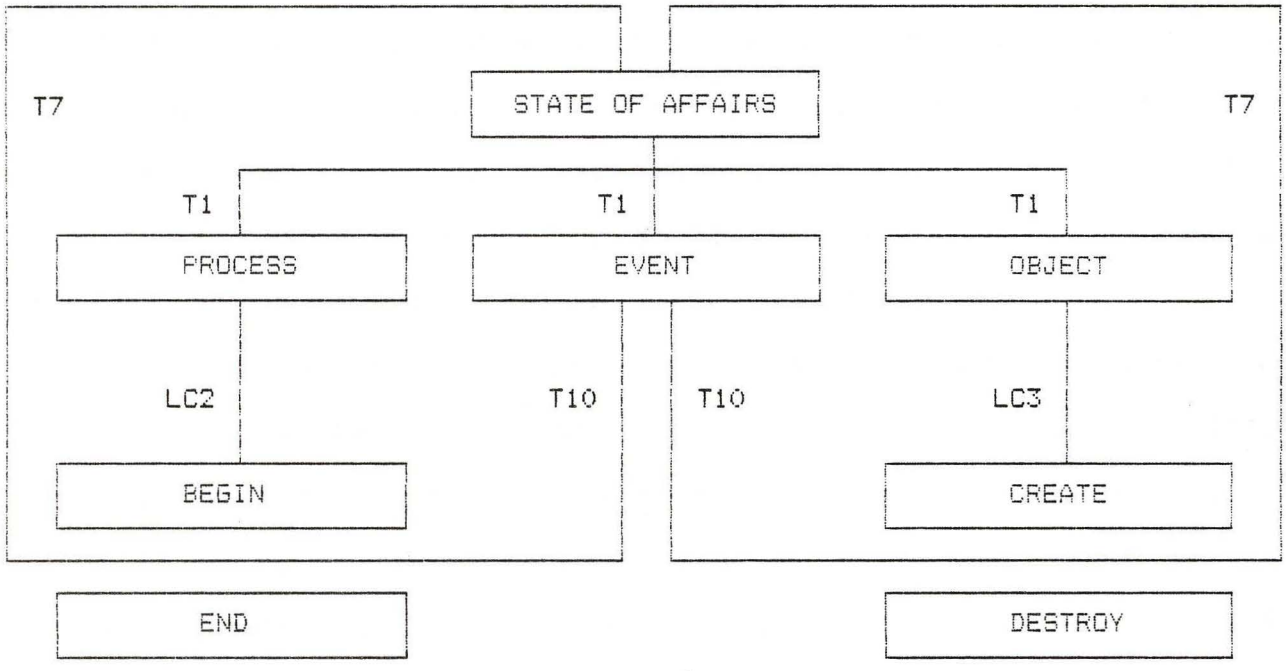


But, by T10, both decompositions are, in fact, events which now places event, as a reality concept, precisely at the midpoint of the temporal axis.



The only way to continue the decomposition from this point is via LC4; i.e., stopping time so that no event can occur (e.g., lowering the temperature of the world to absolute zero). Thus, we obtain a world (real or imaginary) in which all objects that exist remain existent and all processes that are occurring remain in occurrence and no new objects can be created and no new processes can begin. And, moreover, this is the entire world from which we started. Therefore, we have a state of affairs which includes all possible states of affairs (in the world from which we started), which is the articulation of T7 at LC1; i.e., it cannot be decomposed any further while, concurrently,

it is also the whole world so there is nothing else of which it could be a part.



The preceding argument of the non-reductiveness of SA is not merely to make a mathematical point. It is the essential property for any possible model of reality. For if a model is even partially reductive, then there is at least one decompositional path out of it. This introduces the paradox that if you take that path, and the model is of total reality, then you are henceforth modelling non-reality! Thus we have no choice but to accept that fact than any reductive model can only model a subset of reality because there has to be something left as the referent of the model when you succeed in the reduction. By contrast, SA is a possible model of total reality precisely because it is totally non-reductive.

11. THE SA REPRESENTATION FORMATS

Corresponding to each of the four basic reality concepts in SA is a format for representing concepts of each type which, in SA terminology, are the Basic Process Unit (BPU); the Basic Object Unit (BOU), the Basic Event Unit (BEU), and the State of Affairs Unit (SAU). The reason for different units is that the first part of each represents the observational aspect of the concept (the way it was described by the observer: i.e., a process has STAGES, an object has COMPONENTS, etc.) The second part of each format represents the State of Affairs aspect of the concept since this is the means of convertibility via the Transition Rules (note that all the rules operate by first transforming into a State of Affairs and then into one of the other observational units).

For brevity, we introduce only a single format which encompasses all four by referring to a constituent and then classifying it as to process, object, event or state of affairs. Furthermore, instead of illustrating an empty format and attempting to explain it in generality, we choose instead to introduce it via example - and the example we use (as you might well guess) is Kent's "murderer and the butler" problem. The example begins with a structured formulation (one might say program) of the situation in terms of the Transition Rules. It then proceeds to define, using the Representation Format, two of the primary constituents.

As such, the example is "stubbed-off" quite severely but sufficient to demonstrate the power of the system (in actual use, SA formulations are always stubbed-off to some degree - you can't define everything nor do you need to because to do so is to assume that the user has no competence whatsoever; i.e., is the equivalent of an infant).

STATE OF AFFAIRS "PROGRAM"

```
SA:  "DID THE BUTLER DO IT"
   begin
      Process: "MURDER"
      Object:  "BUTLER"
      Process: "INVESTIGATION"
         begin
            Event: "FINDING THE BODY"
               begin
                  Before:
                  After:
                     begin
                        SA:  "SOMEONE DID IT"
                           begin
                              SA: "SUSPECT"
                                 begin
                                    Object: "PERSON"
                                    Process: "MURDER" (element:  "PERPETRATOR")

                                 end; {"SUSPECT"}
                           end; {someone did it}
                     end; {after}
               end; {finding the body}
            Event:  "SOLVING THE CASE"
               begin
                  Before:
                  After:
                     begin
                        SA:  "THE BUTLER DID IT"
                           begin
                              SA: "SAME AS"
                                 begin
                                    Object: "JAMES"
                                    Object: "PERSON" (element:  "NAME")
                                    Object: "BUTLER" (element:  "EMPLOYEEE")
                                    Process: "MURDER" (element: "PERPETRATOR")
                                 end; {"SAME AS"}
                           end; {the butler did it}
                     end; {after}
               end; {solving the case}
         end; {investigation}
   end; {did the butler do it}
```

STATE OF AFFAIRS DATA STRUCTURE "MURDER"

```
Process "MURDER" is
   Paradigm "PERSONAL PREMEDITATED" consists of
      begin
         SA: "MOTIVATION" with options
            begin
               SA: "ANGER"
               PROCESS: "FINANCIAL GAIN"
            end;
         Process: "PLANNING MURDER"
         Process: "KILLING"
      end;
   with element
      begin
         Object: "PERPETRATOR"
         Object: "VICTIM"
         Object: "WEAPON"
         SA    : "MOTIVATION"
      end;
   with individual
      begin
         Object: "GARDENER"
         Object: "BUTLER"
         Object: "GUEST i"
         Object: "HOST"
         Object: "GUN"
         Object: "STOKER"
         SA    : "INSURANCE POLICY"
         SA    : "WILL"
         SA    : "ANGRY"
      end;
   with eligibility
      begin
        "PERPETRATOR":  ("GARDENER","BUTLER", "GUEST i")
        "VICTIM":  ("GARDENER", "BUTLER", "GUEST i", "HOST")
        "WEAPON":  ("GUN", "STOKER")
        "MOTIVATION":  ("ANGRY", "INSURANCE POLICY", "WILL")
      end;
   contingent upon
      begin
         (PERPETRATOR,MOTIVE) SA: "HAS"
         (PERPETRATOR,WEAPON) SA: "ACCESS"
         (PERPETRATOR,WEAPON) SA: "COMPETENCE"
         (PERPETRATOR,VICTIM) SA: "NOT"
      end;
end; {murder}
```

STATE OF AFFAIRS DATA STRUCTURE "BUTLER"

```
Object "BUTLER" is
   Paradigm "PERSONAL LIVE-IN" consists of
      begin
         Object:  "EMPLOYER"
         Object:  "EMPLOYEE"
         Object:  "RESIDENCE"
         SA:      "LIVES AT"
         Process: "STOKE FIRE"
         SA:      "DOES PROCESS"
         Object:  "STOKER"
         Object:  "SERVES MEAL"
         SA:      "COMPETENCE"
         SA:      "ACCESS"
      end;
   with elements using consists of
   with individual
      begin
         Object : "JAMES"
         Object : "MAXWELL"
         Object : "VICTOR"
         Object : "SIR JOHN"
         Object : "BRADFORD HOUSE"
         Process: "SERVING STEAK"
         SA     : "IN THE SAME ROOM"
         SA     : "HAS DONE IT BEFORE"
         Object : "FIRE IRON"
      end;
   with eligibility
      begin
         "EMPLOYER": "SIR JOHN"
         "EMPLOYEE": ("JAMES", "VICTOR", "MAXWELL")
         "RESIDENCE": "BRADFORD HOUSE"
         "STOKER": "FIRE IRON"
         "ACCESS": "IN THE SAME ROOM"
         "COMPETENCE": "HAS DONE IT BEFORE"
         "SERVES MEAL": "SERVING STEAK"
      end;
   contingent upon
      begin
         (RESIDENCE,EMPLOYEE,EMPLOYER) SA: "LIVES AT"
         (EMPLOYEE,STOKES FIRE) SA: "DOES PROCESS"
         (STOKES FIRE,SERVE MEAL) SA: "AFTER"
         (STOKES FIRE,STOKER) SA: "ACCESS"
         (SERVES MEAL,EMPLOYEE,EMPLOYEE) SA: "ACCESS"
         (EMPLOYEE,STOKES FIRE) SA: "COMPETENCE"
         (EMPLOYEE,SERVES MEAL) SA: "COMPETENCE"
      end;
end; {butler}
```

By way of examination of the preceding formulation, a number of contrasts between SA and present data models should be apparent:

(a) PARADIGMS - SA allows for the fact that the occurrence of something on one occasion may have nothing whatsoever in common with the occurrence of that thing on another occasion except that it was an occurrence of the same thing. Note in the example that the "Sherlock Holmes" version of MURDER would have virtually no constituents, elements, etc., in common with, say, a conspiratorial political assassination except that they are both occurrences of murder.

(b) ELEMENTS VS INDIVIDUALS - SA makes a clear distinction between the ingredients necessary for something to occur (in any of its varieties) and the individuals (historical particulars) eligible to be those ingredients on a given occasion. In the example, note that the butler is not eligible to be the weapon in this version, thus ruling out that he might be a Karate expert and actually used a part of himself (e.g., his foot) to perpetrate the crime.

(c) CONSTITUENCIES - SA allows even more complex relationships to occur among the constituents. For example, we specified that the perpetrator could not be

the victim, thereby ruling out the possibility of suicide. Another paradigm might not have this constraint. This and the preceding point are both clear cases of empirical categorization as opposed to pre- empirical policy.

(d) UNLIMITED COMPOSITION/DECOMPOSITION - SA does not require the specification of ultimates pre-empirically (i.e., determination of the atomic attributes). In our highly stubbed-off example the process of stubbing-off a formulation makes those constituents the ultimates but only on a pro-tem basis. "James" is an atomic attribute in our example because there is nothing to prevent us from eventually adding such a description in which case "James" would not be an atomic attribute and, instead, were serve to coordinate our description with a more complete description.

(e) COMPETENCE - SA achieves much of its representational power by relying on the competence of its users. As said previously, it's not just that you can't afford to define everything: it's that you don't need to. The SA "AFTER" was not defined further and there is no need to do so. Any six year old child knows what "after" implies and will use the term correctly without the computer having to provide him a definition

of the concept. Moreover, any user community already shares the concept "after" and there is no need for the database to have a representation of it. It's the complex, fuzzy and difficult concepts of which analysts may, in fact, have different perspectives where the power of a representational system like SA should be applied. That it takes a lot of information (and, consequently, time) to record these should not be a concern. If analysts spend time already determining how they will interpret data under varying circumstances, then SA merely provides a place for recording those determinations so they don't have to reconstruct everything from the basic facts each time the circumstance occurs. This seems a far more productive way to spend time as opposed to designing algorithms whose real purpose, when viewed from the SA perspective, is to construct SA descriptions when they're needed. With SA, the descriptions are already there and subject to continual refinement. In such an environment the algorithms somehow seem much less important.

12. "INFORMATION PROCESSING" APPROACH

To generalize upon the last point, SA attempts to maximize the use of structures as opposed to processes. And this is, perhaps, its principle virtue. Structures and processes are, of course, interchangeable as a structure can always be described as the process that computes it and a process can be described as its initial and terminal structures. And this interchangeability probably accounts for much of the progress in database theory over the last two decades in terms of increasing emphasis on "what" data should be retrieved as opposed to "how" to retrieve it. By contrast, today's "expert" or "knowledge-based" systems have progressed (or, perhaps, remained) in the other direction. Rule-based systems typically specify a complicated chain of inference through which elementary facts can be "processed" to determine an implication. No doubt, the "structural" and "information processing" approaches are logically equivalent. But from a pragmatic, that is to say, behavioral perspective, the differences are vast, as the following example will attempt to portray.

"Take a basketball. It's leather, inflated with air, and spherical. It bounces true; i.e., its angle of incidence and reflection are always the same. Now, take the same basketball, deflate it, soak it in a bathtub full of water, and reshape and restitch the leather so that when it's reinflated it has the

shape of a football. For all practical purposes it now bounces randomly (ask any football player who's tried to recover a fumble). There are at least two ways to explain the now erratic way in which the ball behaves.

"One way is the 'information processing' explanation. In this explanation, the ball started with the initial algorithm that prescribed the bouncing behavior of a sphere. Then, as the ball was deformed by soaking and restitching, each change was recorded in the ball's 'mind' as a sequential change to that algorithm. Now, when the ball is called upon to bounce, it computes its old departure angle (that of a sphere) and then sequentially process the effect of each distortion on the newly computed angle, and finally arrives at the correct angle of departure and goes in that direction.

"But there is another explanation, too. The ball now bounces like a football because it is a football. Whatever it was, or whatever happened to it in the bathtub, is merely a museum piece, the knowledge of which has no bearing whatsoever on how it behaves. [15]"

The difference between the two is far more than just a way of looking at it. If we persist in the "informational processing" approach, it is literally impossible for a person to get up from his chair, let alone walk around the room. The necessary

computations would exceed the capacity of all the computers in the world working in parallel; yet we maintain that one brain could do them in a fraction of a second. The difference simply can't be that great.

13. "WHAT" IS DATA

It is customary for a report to begin with a definition of its subject matter. Given the way we toss about such terms as data, information, knowledge, fact, etc., the need to do so is self-evident. But it is also an exacting and more or less impossible task as this exercise will show. Consider the following definitions (emphasis added) from Webster's [16]:

> DATA applies to a real or assumed fact from which conclusions can be inferred.
>
> INFORMATION applies to facts gathered by observation but does not necessarily connote validity.
>
> KNOWLEDGE applies to a body of facts gathered by observation and to the ideas inferred from these facts.
>
> FACT applies to the state of things as they are; reality; a thing that has actually happened.

Applying the definition of "definition" (the convention in this paper is use quotes to signify the real-world concept for which the term is merely a locution) and employing our familiar logical forms, we can arrive at, at least, two conclusions that may not be so obvious.

(1) While "fact" is the atomic unit or building block of the three preceding definitions, it is only so with respect to that context. The "state of things as they are" invites everything that is presently the case; i.e., the real world. Hence, "fact" is simultaneously a Limiting Case II (LC2; recall the Limiting Case definitions) ultimate particle for the definitions; and an LC1 State of Affairs that includes itself. (Recall the recursion argument to see that we can't even start without that property.) This phenomenon will be dealt with at length when we get tot he section of Relations. For now it is simply worth noting that while it is safe to say that a database is a collection of facts, it is not safe to ignore that a fact may also be a collection of databases.

(2) For the first three definitions, the "statement of what a thing is" clearly includes (indeed, is dominated by) what is appropriate to do with it. For example, the basis for distinguishing between "data" and "information" is whether or not it is appropriate to draw conclusions from it. Directly from its definition, such behavior is appropriate for "data" while, indirectly, this is not the case for "information" since no conclusion can be inferred from an invalid premise if we adhere to formal logical principles. Similarly (and even more so), it is not appropriate to do whatever one does with "knowledge" unless that "knowledge" was the product of some sort of inference

beyond mere observation (that this, again, makes "getting to first base" seem impossible will be dealt with shortly). For now, let me just note that to use "definition" in this regard requires that it not be merely "a statement of what a thing is" but, in addition, it must be someone's statement, i.e., the essential difference between data and information is the status assigned to the fact in question. Moreover, as there can be no rigorous (inductive or deductive) schema for ascertaining this in advance (the use of any such schema would, by definition, elevate the facts status to "knowledge"), the status is, ipso facto, subjective and thus personal, and thus a function of that persons assign to the functioning, accuracy, etc. of the device that determines the status of the data that the device provides.)

To summarize this double-edged exercise, we need simply point out that: in the first place, data (or information, etc) is not the haven of objectivity and empiricism that mathematicians may have perceived; and secondly, the mathematician's notion of "definition" is of very limited utility in this habitat (more on this when we get to Names). For now, we want simply to be able to define terms without imposing unnecessary limits.

13.1. Traditional Definitional Deficiencies

To understand the motivation for a larger context in which the traditional (mathematical) concept of definition can be

embedded, we need to point out what the traditional concept loses in the first place.

> "A definition of X is a set of necessary and coefficient conditions, Y and Z, for distinguishing between cases of "X" and cases of "~X" (in which of course, X does not appear in an essential way)."

In this definition, the first and foremost thing we lose is "x"! For if Y's and Z's are always necessary and sufficient conditions for "X" then all cases of "X" reduce to cases of Y's and Z's and we have to ask if it's legitimate to simply sacrifice "X's" altogether as superfluous (and, ultimately, "Y's" and "Z's" ad infinitum or until a limiting case is reached.) The second casualty is X. For when a set of necessary and sufficient conditions for "X" exist, it is always a second way of defining "X" and, in normal behavior, is for the sake of the listener who can't already distinguish "X". By far the most common definition of "X" is simply X else common dialogue would be virtual impossibility. The third loss, and perhaps the most important, are all the "Xi's" for which the set of Yi and Zi doesn't exist or isn't practically denumerable. (Try to define an ordinary wooden pencil without inadvertently admitting mechanical pencils or pens.)

Finally (this is a special case of the first), we've conceded any direct access to the Y's and Z's, so that we can make judgments about "X". If we disagree about "X" our only recourse is to change the subject to Y or Z. But Y and Z are immutable at that level of access since other "Xi's" are also reducible to these. (Hence the proverbial "Y if including but not limited to... ...hereinafter referred to as..." that plagues much of the technical and all of the legal literature.) Moreover, as Y and Z are most often status assignments rather than logical conditions, disagreement is the rule rather than the exception.

To summarize, the best we can say of the traditional form of definition is that it provides a necessary and sufficient set of logical conditions for the correct use of X as a locution. And while this is certainly an essential tool for linguistics, it does little toward providing any insight to "X" beyond what we already knew.

13.2. Paradigm Case Formulation (PCF)

In contrast to the logical form which tells us, ex post facto, how to correctly use the term X, PCF is a behavioral form that, at least, gets started without inherent contradiction. In essence it is a codification of observation. "To observe something on a given occasions to find out something about it without, on that occasion, having to find out something else

about it first (observation contrasts with inference)" [17]. In other words, PCF deals directly with at least some case of "X" without:

(1) requiring Y's and Z's;

(2) requiring X; and

(3) requiring an account of other cases of "X".

If definition has a role in learning, PCF is at least a possible form of this phenomenon while the logical form clearly is not (by infinite regression [18]. PCF is simply:

I. State of paradigm case of "X" (i.e., an indubitable example of "X)

II. Incrementally induce transformations on the paradigm case that either:

A. yield other cases of "X"; or

B. yield cases of "~X".

III. (optional but useful) enumerate one or more X's as locutions for "X" (this can be quite arbitrary as in a computer implementation; or more or less natural language if that is appropriate).

Let's return to the original example of "data" and see the difference between the behavioral and logical forms [19].

13.2.1. Paradigm Case Formulation of Data (Version 1)

I. Data is a collection of facts that are descriptive of something that is identified.

II. T1. That a fact has a truth status (e.g., true, likely, doubtful) is a fact.

T2. That a fact has an interpretation is a fact.

T3. That a fact is interpreted is a fact.

T4. How (the way in which) a fact is interpreted is a fact.

T5. That something is identified is a fact.

T6. How (the way in which) something is identified is a fact.

T7. A fact can be regarded as a something.

T8. A description of something can be an identification.

T9. An identification of something can be a description.

T10. That something is observed is a fact.

T11. A something can be a fact, a description, and interpretation, an observation, an identification, or a something else (!).

T12. A something can be a collection of somethings.

T13. While permissible that "an X1 of an X2 of an X3 of an ...Xn is (or can be) a fact (or Xm" generates an infinite number of transformations, some of which appear above, we must put a halt to unbounded descents into these levels. (Note: Xi is a member of the set {description, interpretation, observation,identification}.

T14. The identification must be agreed upon.

T15. A collection of facts is a finite set of size greater than one.

T16. The logical conjunction of the elements of a collection of facts has the same truth value as the weakest truth value among its members.

T17. The logical disjunction (inclusive) of the elements of a collection of facts has the same truth value as the strongest truth value among its members.

T18. There may be a minimal collection of facts to consensually describe a given identified something.

T19. In principal there is no maximal collection of facts to consensually describe a given identified something but we must agree on one for a given identified something.

13.2.2. Paradigm Case Formulation of Data (Version 2)

I. Something that does (or could) produce a change in someone's knowledge.

II. T1. The something is information.

T2. The change is not a something.

T3. Someone's knowledge is not a something.

T4. Information can be regarded as a collection of facts under the same transformations as the previous PCF. There somehow it is easier to limit T13 (there); for after certain level of descent, there is no change in knowledge (except in the 4-foot round apple case!).

T5. We don't want (or do we?) to strengthen "could" to "must" or "does", but failure to do so leaves the someone riddled with problems (e.g., does the someone have to be of some particular background - linguistically, genetically,...).

T6. The change may be measured by the someone or by someone else, either subjectively or objectively.

T7. Perhaps information could be expanded to include other things besides a collection of facts; but we may not want to allow this.

Of course, this is in now way complete (nor could it ever be - the example of a 200 pound tomato reifies this point) in any mathematical sense. But it's sufficient (indeed, on might say more than sufficient) as a pragmatically workable access tot he subject. Furthermore, it transcends deficiencies inherent in the logical form:

1) It defines "X" directly without any essential references to Y's and Z's. The only essential references are to "X" (more precisely, other "X's") which while defying a cardinal rule of the logical form, is utterly natural in behavior (particularly, learning behavior in which it is only by reference to possible other "~X's" that a concept for distinguishing "X's" from "~X's can emerge: competence = knowledge + experience).

2) Any disagreement about "X" can be settled directly in the PCF of "X" by denoting which transformation, specifically, is at issue. Not only does this codify that there is disagreement, it does so in a way the particularizes the essential difference between what I take to be "X" and what you take to be "~X" and leaves the balance of the formulation completely intact. More so, such disagreements are codified entirely within the context of "X" and not in some other context such as Y or Z, such that other dependencies on Y's and Z's are not affected. Instead, PCF "settles" disagreements about "X" by producing "Xi's" and Xj's.

2') This also holds to agreements. If we conclude that "Xi" and "Xj" are the same, we do so in the context of a particular transformation that codifies why they are the same; i.e., what is the way in which they are the same (and, indirectly, for what purpose without affecting other purposes).

3) Self-evidently, PCF deals easily with an "X" for which no denumerable Y's and Z's exist since it doesn't depend on y's and Z's in the first place. (But this should not be taken as a claim to "Fuzzy Logic" which is an extension to Mathematical Logic to deal with imprecision. To the contrary, it is ordinary behavioral logic in which "fuzziness" is completely accounted for a priori.)

To conclude this section on the definition of data, it is useful to strip away the "double-edginess" and restate the point of the discussion more directly. THis entire report is a critical analysis of data modelling which, in practice, is more often referred to as data definition. If there were no alternative to the logical form of definition, the balance of the report would be mostly writable. And if we had no access to the subject of "data" itself (apart from how we process it in database systems) there would be little point in writing it anyway. It is of paramount importance in understanding the subsequent sections that the reader remembers that we are discussing "data" in

its everyday, commonplace form; not data as computer scientist have chosen to restrict it.

14. GENERAL SEMANTIC ISSUES

A frequently espoused lament in database circles is that we would not be in the mess we're in if we had started with semantics instead of relational algebra. Even Kent prologued his book with his Message to Mapmakers:

> "Highways are not painted red .... and there are no contour lines on the mountain." [20]

It's more or less customary to discuss Korzyski's General Semantics in the course of data modelling because, after all, data models are very much maps of real territory and the relationship of maps and territories is what General Semantics is supposed to be about. This report follows custom, but only to point out enough shortcomings to show why General Semantics won't form a sound basis for data modelling. It also discusses the other semantic issues raised by Kent such as existence, uniqueness, sameness and change.

14.1. Reality

Most expositions on data modelling or the database concept itself contain directly or indirectly something like the following as their starting point:

1) A database is a model of (some subset of) the real world;

2) Therefore (so that (1) remains true), events that occur in database systems (i.e., data processing are models of (some aspect of) a real-world event.

As an aspiration, it is difficult to find fault with this premise. Moreover, almost everyone agrees that it remains only an aspiration and that, for now, a weaker premise is more appropriate:

3) The real world is what data in databases is typically about;

4) Real-world events are what the events that occur in databases are typically motivated by (i.e., data processing is not arbitrary and capricious).

The issue at stake here is not faithful real-world modeling per se (this whole report is about that) but having admitted to a "fidelity gap", how do we deal with it systematically; preferably in a quantifiable way so we can measure the gap and thus the success of our attempts to reduce it through better modelling technology. Not surprisingly, the course of events thus far have been rather predictable and not arbitrarily different from those

that lead to the Universal Recursion (UR) assumption. Ignoring, for the moment, the merits of relational normalization it is instinctive to examine the circumstances that motivated the UR assumption; the value of the current debate over this validity; and the relevance of all this to the issue at hand.

Early in the study of synthetic normalization (building n-ary relations out of (binary) functional dependencies (FD's)) it became apparent that the third (and subsequent) normal forms were not unique for a given set of FD's. (There could exist several, all correct, shemata for the same database as defined by a set of FD's). Since they were different yet all correct, some way had to be established to show that they were, in fact, equivalent in behavior (else at least one had to be incorrect?).

The principle of the method pursued was a referential theory. It assumed (imagined) the existence of a Universal Relation comprised of a subset of the Cartesian product of all the domains in the database. For each normal schema, a chain of loss-less (reversible) projections and joins was constructed that would transform the UR into schema in question in such a way that the schema could be transformed back into the UR by reverse calculation. The equivalence proof then proceeded as follows. A UR in state Ui would be transformed into the equivalent schemata states Si, Ti. Then, equivalent operations would be performed on Si and Ti to obtain Sj and Tj. Finally, Sj and Tj would independ-

ently be transformed back into Uj' and Uj" with the expectation that Uj' = Uj" implies the equivalence of the S and T. I.e., S and T are two different descriptions of the same thing, namely U. And that's why it's a referential theory, it depends on the description of something else, U, to which both S and T are referring.

If two data models are different, the tendency is evaluate them by reference which, if it wee do-able, would yield the desired quantification. But, as defective descriptions of reality, the only possible referent is the real-world and our only access to that is by description. Hence we're back in the infinite regression game:

> "What is it that two different descriptions
> of the same thing are descriptions of?"

If we answer that question, the answer, too, will be a description so we then have to ask (ad infinitum):

> "What is it that three different descriptions
> of the same thing are descriptions of?"

If we don't answer the question, then we have to accept that only a non-referential theory is possible. I.e., we have to evaluate a data model by the way in which components (primitives) relate

to each other and then in terms of what sort of world (reality) could be obtained in a model constrained to those kinds of relationships. And the result, while qualifiable, doesn't lead to any direct, model to model, comparisons. Instead, it offers only the criterion, for any given model, that is useful to someone who can operate in that reality and useless to someone else who can't. It is even the case that supersetting (upward compatibility) is not a useful measure. The presence of an "additional" relationship (e.g., a generalization hierarchy in a semantic model) without adequate support for that relationship (distinguishing among taxonomy, status assignment and appraisal) may yield an reality in which a person cannot operate whereas eliminating that relationship would yield one in which he could. This is in direct contrast to purely syntactic systems. In a purely syntactic system you can safely ignore a provision (feature) of the syntax id you don't find it useful. E.g., in PASCAL you have access to both a "WHILE condition DO" and "REPEAT UNTIL condition" and since you can do (syntactically) anything with the first that you can with the second then you can ignore either one. But PASCAL also makes a conceptual distinction in that a REPEAT...UNTIL will always execute at least once even if the condition is false on the first iteration. Thus (unless you only read your own programs) you cannot safely ignore the distinction because others who write in PASCAL may have employed the distinction for their won purposes. I.e., whenever the producer and consumer of information are not necessarily the same

person, the consumer must always be aware of any conceptual distinctions permitted to the observer by virtue of the syntax, and furthermore, to know which distinction the observer employed. This has been well known in intelligence systems for a long time and is an important reason that such systems rigorously reduce the possible distinctions rather than (inadequately) attempt to enrich the possibilities.

14.2. General Semantics

As presently formulated, General Semantics, albeit a study of maps, also depends on a theory of reference and ultimately fails for precisely two reasons. The problem arises as follows:

1) while it is correct to say that "a map is not the territory" and that "a map can never completely represent the territory" and that "a map of a map is a map of a different order";

2) it is incorrect to assume that a map could be a map of anything other than another map.

The failure attributes to (2) above arise in connection with the inevitable incompleteness: i.e., a map can never completely represent another map. Thus if all maps are incomplete, we are

left with the question of "How can a map convey what it doesn't represent?" If the answer is by reference to the territory, this is just another map and we have to ask the question in terms of what the referenced map doesn't represent... ad infinitum. As with the general case of descriptions (of which maps are a special case or codification), we can't admit the (unanswerable) question as to:

"What is it that two different maps
of the same territory are maps of?

Since the answer is, of course, yet another map, we're back to an infinite regression. Thus, if General Semantics is to provide any advantage, it has to be exploited non-referentially. The approach to doing this is essentially that for descriptions. A map is useful to the extent that a person can accomplish an intended task (finding 34th Street, finding oil, and the consequential world it defines. And, like descriptions but, in this case, more obvious; more detail is not necessarily an advantage and can typically be counter-productive.

Having cited referentiality as a fundamentally erroneous assumption in maps, descriptions, data models, and any other contrivance for dealing with reality, all of which are useful, everyday tools for normal behavior, would seem to introduce a paradox. If

they are based on fundamentally wrong assumptions and we use them, then our behavior should be fundamentally wrong as well. But,of course, this isn't the case and there must be some alternative explanation. One possible explanation is that the assumptions ar ex post facto attempts to explain how we are able to use these things; and that, as explanations,they are not only wrong but also totally irrelevant in that we don't, as persons, in any way depend on the existence of an explanation of how we do things that we ordinarily do. This is a fundamental assumption of State of Affairs (SA). The following "reality" issues are considered in that context.

14.3. Existence

If there is any aspect of traditional approaches to data modelling where we, in fact, allow the map to control the territory, it is in deciding what exists. For example, in post-relational models we choose, a priori, some sets (classes) of entities and these are what the resulting model and its future extensions will be about (i.e., will record relationships among). The arbitrariness of these choices and the subsequent difficulty of redefining a relationship as an entity (or conversely) inevitably leads to a "forced" world view which everyone must share to use the system sensibly (arbitrary in that, as Levin pointed out [21], the only difference between an entity and a relationship is that we happened to notice the entity first).

Thus, if we happened to notice "TRANSACTIONS" early enough, they will be regarded as existing and will have to be given (usually complex, concatenation-style) names so we can refer to what exists.

By contrast, SA takes the approach that "independent" or "referential" existence apart from behavior is not only unnecessary but essentially paradoxical [22]:

1) What exists is whatever is presently the case (i.e., the real world); and

2) we can only determine this by what it is we presently take to be the case.

I.e., existence isn't a "pipeline to the truth" nor something we can deal with objectively. If we take it to be the case that they are attacking us, then an offensive exists and we will make observations about that offensive and ultimately draw conclusions (e.g., their likelihood of winning; how we should react; etc.). And we will do this independently of whether or not they really are attacking us (and conversely - we will make observations about a military exercise even though they may be attacking us).

Of course, SA provides no solution per se and, in fact, recognizes explicitly [23] that:

> "We require a concept of something more significant than our immediate thoughts and observations so that those thoughts and observations can be about that something."

What SA provides is the understanding that what we take to be the case is what we will act on, regardless of any objective truth. Hence, objective truth, if there is such a thing, is essentially an irrelevant concept.

14.4. Uniqueness and Sameness

A data model with its extension at any point in time is an unelliptical description of what there is. This is an absolute requirement in databases otherwise the concept of identification would be lacking. Post-relational models are an improvement over previous models in that they employ an arbitrary system catalog (surrogates) as opposed to unwieldy attributional concatenations but they still embody unique identity of what there is. There is no parallel to this in natural language. Any (every) statement is an ellipsis and there is inherently no limit on the length of the unelliptical form (other than everything that is presently the case; i.e., the real world). Recall the first

section of this report on "definition" and the impossibility of a set of necessary and sufficient set of conditions to identify something. I.e., there are no deductive guarantees and other than our mathematical heritage, there seem to be no sensible reasons to require any.

SA offers a much more useful and no less rigorous (mathematical) approach. Data (facts, statements, etc.) are merely an incomplete description of something for which a more complete description is always available.

T2: An object (or process, or event, or state of affairs) is a state of affairs that is a constituent of some other state of affairs.

T3: An object (or process, or event, or state of affairs) is a state of affairs having other states of affairs as immediate constituents.

There are, of course, the basic notions of composition and decomposition respectively and part-whole relationships in general. the approach, again, does not solve the issue of unelliptical forms, per se, but it puts the issue in perspective we can deal with in a natural and non-paradoxical way.

SA recognizes four limiting cases on unlimited composition and/or decomposition which are context-free; i.e., can be arbitrarily invoked at whatever context-specific level is necessary.

LC-I: The state of affairs that includes all other states of affairs (e.g., computer science, the world of chess, the real world, God);

LC-II: An object that has no constituents (e.g., a brick, a quark, an imaginary particle);

LC-III: A process that has no constituents, hence the equivalent of an event (e.g., a computer run, a battle, a synapse);

LC-IV: An event that has no constituents; i.e., no beginning that is different than its end (another day, the action of a molecule at absolute zero).

The invocation of these cases is, by contrast, completely situation-dependent. E.g., tactical intelligence would normally invoke the engagement at hand as LC-I while strategic intelligence might invoke a concert of engagements as well as political considerations. And firing a missile may be an LC-III event tactically, but an enormously complex process to a ballistics expert.

The summation of this in the context of what is the same and what is different is simply that "identity need only be accomplished, not proven."

14.5. Chance

The most profound failure of data modelling technology is in its ability to represent change. This is due, in no small part, to the aforementioned schematic depiction of relationships. Data models divide "what there is" into entities and relationships and these are defined a' priori; i.e., the relationships in which a given entity may participate are declared in the schema, and the only possibility for representing change is the boolean determination of whether or not two entities are, at the moment, participating in a predefined relationship based on attributional constraints. While this certainly qualifies as change, it is only that of the most simplistic sort; that is, the kind of change limited to the Parameter Principle directly and with no flexibility to "step back" (compose) or "move in" (decompose) to account for changes of a (much) more significant sort.

In SA, the very notion of change (in something) is the kind of relationships in which it can, at present, participate. The "Murder and the Butler" problem [24] is exemplary of this phenomenon, but so are any number of examples one might conjure up. Then years ago, I could not possibly have been related to

this writing in capacity of another. Arguably, this may have, in part, been due to attributional constraints (e.g., I hadn't even heard of SA or Pete Ossorio). However, whether or not I am now, at least, eligible to participate in that relationship is far more than an attributional change:

1) there are certainly no "objective" attributes that, unequivocally, prove my eligibility; and

2) most such attributes are, in fact, status assignments made (not necessarily consistently) by others.

But, in fact, I am writing this paper. And that could not possibly have been predicted ten years ago; and it will certainly cease to be true (in the active tense) when, if ever, it has been written. Moreover, although I have written other papers in the past, I may not write any papers after this. Schematically speaking, it is not the instantiation of predefined relationships that constitute change so much as it is the appearance or disappearance of the defined relationships themselves (and to say otherwise is to claim the ability to predefine every possible kind of relationship that could exist).

"That a state of affairs is related to another state of affairs ia a state of affairs" and these are delimited only by LC-I; i.e., everything that is the case or the real world.

15. ENTITY NAMES

Names of things (entities) in data modelling get inextricably bound up with semantic issues since we can't store the "things" themselves in a database. Hence, it has been correctly observed "... that we give them [the entities] names and store these instead." [25] In extremely simplistic "worlds" one can often get away with this tactic but, as we will see, it is inevitably doomed to fail.

15.1. Correspondence

The simplistic "world" in which traditional data model naming systems work naturally is when there exists a one-to-one correspondence between the names and the things in the world. In "business worlds", for example, the fact that lawyers and accountants were on the scene before the data modelers often causes this to be the case. Accounting systems are about accounts and these already have names (e.g., from the chart of accounts) and these names are, by definition, unique (e.g., from the chart of accounts) and these names are, by definition, unique (e.g., they exist as separate pages in the ledger book). In such situations, traditional data models fit nicely which is one reason they have (and will) remain popular in commercial computing.

However, it is sometimes the case that there isn't a one-to-one correspondence between names and what things they stand for. This may occur in several different ways, and we generally attempt to solve each in a different way.

1) Sometimes the names (the real names - i.e., the one's we use naturally) aren't unique among the things they represent. The classic case is a person's name in the company that employs that person. So we create new things that are in one-to-one correspondence (e.g., employee numbers) with the things and give these new things names in such a way that they are unique. Note, however, that the original names don't go away. It's highly unlikely that employees refer to each other by their employee numbers. But, in the computer, it is the new, unique, name that is used. Hence, the level at which we describe (i.e., name) something depends on how we intend to use that description. And the level at which it is most natural to name something is not, in general, a level at which we can uniquely identify it (even the employee number fails across multiple companies). This is, essentially, the issue of unelliptical descriptions relisted. In theory, there are no unelliptical names but, in practice, we can typically concatenate enough attributes to effectively achieve one in any LC-I limited state of affairs.

2) Sometimes, frequently as a result of doing (1), we have too many names. A person might have an employee number, a social security number, a membership number in a bowling league, etc. In older models the approach was simply to choose one that worked for the application (e.g., keeping the bowling records). More recently, as we have come to expect the data model to serve many different applications, the approach has been the development of name hierarchies coupled with attribute inheritance; and the replacement of the resulting unwieldy concatenation with a purely artificial (i.e., purely for the data model) name referred to as a surrogate; a place-holder for storing data about whatever thing (or kind of thing) the surrogate corresponds to. Of course, this technique (called generalization hierarchies) applies only when there is a well formed classification such as a=in biological species; and when the intent of such a classification is known and mutually agreed upon (more on this last point later). E.g., such a scheme would not work for naming pump stations in a network of pipelines.

3) Finally, there is the case that occurs when there simply aren't enough names for the things in question. This is nicely illustrated by Kent's example of

several open requisitions for the same position in the same department. The data model needs to distinguish which particular requisition is begin filled while the personnel manager makes no such distinction nor does he need to. In one sense, this is an issue of the LC-II state of affairs; e.g., in most contexts, it isn't important to name (hence, distinguish among) the grains of sand on the beach. In another, it is the explicit realization that the names we use are, themselves, "things" and, moreover, only a tiny subset of the things that exist (and, hence, may require names). I.e., it is generally impossible to have a one-to-one correspondence between the things in the world and only the tiny subset of them that are finite, machine storable, symbol strings.

15.2. Naming Systems In SA

SA recognizes that naming, as we do it naturally, is systematic but not in the mathematical (and paradoxical) way just discussed. As opposed to a reverential theory (with its requirement for one-to-one correspondence) SA approaches naming as essentially a non-referential theory in which the identity of things are coordinated as required, but not pre-empirically established.

1) The naming (representation) of a thing has two parts: an identification and a description; e.g., "this technical report about State of Affairs," or "the RADC technical report." Each phrase serves both to identify and describe what is begin written.

2) In an ideal (in fact, unattainable) coordinate system, the identification and description components are orthogonal. E.g., "the class in BE205 on Monday at 4:45 is about State of Affairs" might be considered as approaching this ideal. Any subject could conceivably be taught in BE205 on Monday at 4:45; and State of Affairs technology could conceivably be taught anywhere at anytime. (Most existing data models attempt this approach.)

3) In extreme cases, the identification is the description; e.g. "the course on State of Affairs" or, more convincingly, "the color blue." Note that "extreme" here refers to the way of naming, not to the fact that such cases are rare. In fact, recalling the discussion of PCF, everything has a name of this sort which might, indeed, be called its primary or paradigmatic name.

4) Most coordinate systems have partially descriptive identify parts and partially identifying descriptive parts. (And as Senko pointed out, in these cases the roles are almost always interchangeable depending on the context, as in "EMP #1234 works for DEPT 420." Each is descriptive of the other.) This is equally true of almost all "fabricated" names of the sort previously discussed. E.g., the final report of Air Force Contract F30602-85-R-0012 contains, at least, descriptive content about the organizational unit for which the report was prepared and when.

Again, while SA doesn't purport to provide a way of solving the naming problem, it does offer an understanding of names and, through that, an understanding of the consequences of using a particular naming strategy. These consequences can be summarized in three basic principles.

1) A naming system tends to determine rather than reflect its referents. E.g., "computer science" is (and probably will always be) a branch of Mathematics as far as the library is concerned.

2) Changes to the referents of a naming system tend to create artificial events in which the referents are involved. E.g., if the department changes the number of a course, students will have to drop the old course and add the new course even though it is still the same course.

3) The presentation of name-referent correspondence tends to be symmetric. E.g., if the license number of my car is erroneously reported as that of a stolen vehicle, I will surely be apprehended by the police (and in a real case tat occurred in Florida, shot and killed).

16. RELATIONSHIPS

Perhaps one explanation as to why data models require distinctions between entities and relationships is the mathematical reality that you can't define a relation until you have defined its domains. (And, hence, further evidence of Levin's claim that it's a case of which thing you notice first.) This kind of segmenting tends to make relationships second-class concepts in such models and this half of the problem has remained largely unexplored. In this section it is difficult to begin with traditional data modelling technology as a counterpoint for all such models can deal with relationships only after they have been converted into entities. In SA relationships are also first-class concepts and must be discussed as such (N.B. Kent's enormous motivation to eliminate the distinction is admirable but he could never quite succeed. This is one "basic assumption" he apparently didn't reconsider.)

16.1. Domains

Inarguable, it is the participants that reify a relationship and not the other way around. But relationships exist in concept regardless of their participants and, in fact, typically provide the basis for reifying entities. This key equivocation is notably absent in the present technology of data models. But without relationships as first-class concepts, it is difficult,

if not impossible, to recognize that one of he ways that distinguishes one sort of entity (or kind of entity) from another is, in fact,the relationships it is eligible to enter into. For example, a primary way of distinguishing between Goguen's chair and table [26] is that a glass and table can be related by "placed upon" while a person and chair can be related by "sitting in", but other combinations are not eligible, i.e., the glass cannot sit in the chair, and that is one way to distinguish it from the person. But that way of distinguishing is only available if the relationship "sitting in" is available regardless of whether any entities are, at the moment, participating in that relationship (i.e., reifying it).

SA can succeed here precisely because it is non-reverential. "Sitting in" is well-defined even if it is not currently begin exemplified or even if it has never been exemplified but has merely been conceived as possible in principle. Possible ways in which things might relate is a fundamental ingredient in "natural intelligence" or creativity and its absence in data models is a notable deficiency. Or, stated another way, existing data models force the choice of ultimate objects (LC-II objects) as a pre-empirical conceptualization which becomes a fundamental assumption in these models and most difficult to change. The initial choice of those entities that are not relationships (states of affairs) delimits, a priori, the kind of relationships (states of affairs) that could never be obtained in such a model. And, by

the limitation on relationships, the kind of objects that can be distinguished in such a model on the basis of the relationships in which they can participate is also so delimited. Thus any attempt to make a first-order distinction between entities and relationships leads quickly to a closed-world model.

16.2. Roles and Degree

It is critical, in a system such as SA that treats relationships as first-class concepts, to recognize that the definition of a possible relationship, pre-empirically, (or a possible entity for that matter) is merely a status assignment made by the construction of the model. It contains no objectivity and no permanence. Moreover, the status assigned to either the entity or to the relationship may be either a requirement or a permission. An example of a requirement would be that "sitting in," in order to occur on a given occasion, requires a participant to do the sitting, and another to be sat in. A permission would be exemplified by a person begin eligible to do the sitting, but a dog, for example, not begin eligible. And both of these status assignments are typically part of the missing context in the elliptical description. I.e., one can refer to "the person in the chair" without problem since "sitting" is the only reasonable way (familiar way) in which a person can be "in" a chair (as opposed to having been sewn in the upholstery).

Another status assignment, the one most frequently not treated as such in data models, is the degree of a relationship; i.e., how many eligible entities are required to participate in a relationship in the ways in which they are eligible to participate to satisfy that a version of that relationship has actually occurred on a given occasion. The way in which degree is typically

mistreated in data models is to give a "covering law" interpretation such as functional dependency. For example, most data models require one to take the position that it is not that I am teaching a course about State of Affairs; but that I am a professor, and State of Affairs is a course, and professors teach courses (and, of course, that I am the professor is simply a status assignment). This kind of interpretation fails in two distinct ways.

1) That a particular co-occurrence occurs as part of a relationship does not imply that whenever it occurs, it is part of the relationship. For example, a grant, in order to have occurred, requires that (typically) the Principal Investigator writes a report for the sponsoring agency. However, I have written many reports for a sponsoring agency, only a few of which have been in connection with a grant (or a contract or proposal or any other business arrangement for that matter). And no matter how many of the required (part-whole) co-occurrences have occurred that are required for a grant, it is only the fact that it is a grant that makes it a grant.

2) That a particular co-occurrence occurs as part of a relationship does to imply that it always occurs whenever the relationship does. This is the opposite

of (1). There exists many cases of grants in which the PI doesn't write a report for the sponsoring agency and, perhaps, does few of the other things that typically occur as part of a grant, yet they are still genuine cases of a grant.

Such "causal regularities" hold only among the theoretical or hypothetical - not among what actually happens. To invoke such regularities in a data model (in the way that we use data models) invites disaster.

16.3. Normalization

It is no wonder that relational normalization theory in general, and the Universal Relation (UR) assumption have been as difficult and controversial as they have. For, in the preceding context, it is easy to see that normalization is no more than an attempt to force casual regularities to hold for the empirical. No doubt, such regularities are desirable for computerized processing (discussed in detail later) but they can be obtained only with a proper understanding of the circumstances. In SA one might say there are three "normal forms" (not to be confused with relational 1NF, etc.) or 'degrees of normality" that can be obtained.

1) Evaluative Normal Form is an evaluation that the roles and degree of a relationship on a given occasion hold regularly only inasmuch and for so long as they do (and such regularity in role and degree is useful to some other end). For example, a loan officer in deciding whether or not to make you a loan (the further end) may evaluate that for you (and all the other applicants to be processed that day) the relationship "credit worthy" can be regularized as your net liquid assets at the bank and your payment history on previous loans at the bank. And he will make his determination on the attributional and co-occurrence constraints he

has established while knowing full well that the issue of whether or not your loan would be a profitable venture for the bank on this particular occasion depends on a much larger set of circumstances regarding both you and the bank. (i.e., in light of limited information, time and other resources, a person "takes it to be the case" and acts accordingly).

2) Experimental Normal Form is a hypothesis that under these kinds of circumstances (or circumstances like them in some significant way) certain regularities in role and degree of a relationship can be expected to hold regardless of the others that may not be regular. E.g., in deciding to make a venture loan to a business developing a new product, the officer might consider innumerable situation-specific facts, but if the applicant is presently in Chapter 11 bankruptcy, no other facts will be considered at all. This might be the result of an empirical history or reasonable speculation; but in either case will be used as a "rule-of-thumb" so long as there isn't sufficient contrary evidence to contradict the rule.

3) Generalized Normal Form is a generalization that certain regularities in role and degree of a relationship hold and this will always be the case. This

resembles "rules-of-thumb" in that the rule is applied independently of the situation but differs in that:

a) the rule is considered extremely safe to apply; and

b) there is little time in which to construct any redescription of the situation to which some other behavior might be more appropriate.

E.g., a doctor admitting a patient with chest pains will treat the case as a coronary, at least at the outset. But this kind of normalization, while very similar to Relational Normalization, arises for entirely different reasons. There is no attempt to force a regularity regardless of the situation, it is the situation itself that motivates invoking the normalization, and not the processing of the data (although processing the data may, in fact, be viewed as the situation motivating the normalization).

17. ATTRIBUTES

Traditional data models require, in most cases, yet a third distinction beyond that to be made between entities and relationships. That is the distinction of an attribute which is neither an entity or a relationship per se', but, at best, a degenerate case of either. (In fact, post-relational models recognize this to some extent by introducing the distinction between a concept-based entity and a value-based entity; the latter standing for an attribute). In so treating attributes as degenerate entities and relationships, much of the richness in real situations cannot be accommodated. SA requires no such distinction. In the same sense that entities and relationships are both first-class concepts, the notion of attribution in SA refers to the way in which these concepts are applied.

17.1. Contingency

To say that something is an attribute is to say that it serves as a constraint that the occurrence of a contingent entity or relationship on a given occasions contingent upon. This is considerably different from existing data models. For example, if, as previously discussed, the participation of an entity rather than another, then that relationship is an attribute of the entity whenever that means of distinction is employed. Thus, if the "writing" of this report is a relationship in which the

author is a participant,then it is the relationship "writing" that attributes being an author to the writer; i.e., is an attribute of the person. But this is the case only if that is the means of distinction. It might also be the case that a parson who has never written anything but aspires to be an author is accorded the status of author and, in this case, the relationship "writing" is not the means of distinction and so, on that occasion, is not an attribute of author.

In fact, in SA, it will generally be the case that attributes, as a concept, are not distinguishable from entities, relationships, names, co-occurrences, etc. except inasmuch as they are attributes (on a given occasion).

In this light it can be seen that what existing data models mean by attribute is a combination of:

1) invoking LC-II to create atomic (degenerate) entities; i.e., entities for which no further description is provided; and

2) invoking generalized Normal Form pre-empirically to regulate that all entities of a given kind always have (are always distinguished on the basis of) those attributes (and, typically, no others).

By contrast, in SA, the fact that an entity or relationship is attributional to a State of Affairs (or potentially so) is treated, itself, as a State of Affairs and, consequently, "attribute" is, in SA, a first-class concept as well.

17.2. Ambiguity

As discussed previously, without some form of convention, the identification of an entity, and the description of an entity, will generally be interchangeable; and they will somewhat overlap. And this fact, in conventional data models, inevitably leads to ambiguity. From the SA perspective, this ambiguity arises in two cases.

1) The description may be transitively related to the identification; i.e., on the surface, the identification and description may appear orthogonal but, in fact, are not. For example, using Ossorio's file cabinets, the assertion that "the orange file cabinet costs more than the black one" it is easy to overlook that the identification (color) and the description (price) are not necessarily independent. E.g., the formulation and/or application of orange paint may, in fact, be more expensive than for black paint. When this occurs in the schema design of a model, an extraneous regularity in the attribution (constraining) of a class of entities is the typical result; and anomalies of the sort that Relational Normalization seeks to avoid are the typical consequence. Moreover, if it is in fact the case that there is always some transitive relationship between the identification and

description of an entity (or kind of entity); e.g., a Gross Vehicle Weight is an attribute of a truck but it is also the having of this attribute that essentially makes it a truck; then the elimination of Relational transitive dependence at the schematic level is, in principle, impossible to achieve. Many studies in Relational normalization over the past several years tend to confirm this observation although they do not offer this explanation of the failure.

2) The description may be embedded in a completely or partially presupposed or historical context. In fact, to avoid completely any such embedding is equivalent to claiming existence of the unelliptical description (or identification) which is, of course, generally impossible. E.g., there is nothing innate to a truck (but not a car) that gives rise to the use of Gross Vehicle Weight as a constraint (attribute); it arises in an historical context of regulatory practices that have been applied to trucks but not generally applied to cars. Thus, we typically presuppose certain attributes based on reasons with which we are all familiar (or accept by convention) without ever specifying that basis.

17.3. Generalization Hierarchies

Post-relational data models have attempted a solution to the attribute issue by introducing an hierarchic arrangement of classes and any sub-classes (etc.) of entities such that a subclass inherits the attributes of the class of which it is a member. From the SA perspective, there are two significant concerns with this approach.

1) Generalization hierarchies need to be recognized as being no more regular than any other attribution schema, since, in general, an entity will be a constituent of different states of affairs on different occasions. Thus, in reality, an entity inherits the attributes of a state of affairs of which it is a constituent only when it is such a constituent. To claim otherwise, of course, would be equivalent to declaring that one and only one taxonomy (classification) can exist for all the entities in the model. For example, a data model may specify that a truck is a constituent of the capital assets of a company and, therefore, inherits the attribute of depreciation schedule. But it may also be the case that a truck is a constituent of shipments and thus inherits the attribute destination. It will clearly be the case that depreciation does not serve as a constraint in

assigning a truck as a constituent of a shipment and thus is not an attribute of the truck on such an occasion. The only alternative position is to specify that every fact that could, in principle, be discovered about a truck is already an attribute of a truck and that data modelling is really an exercise in epistemology.

2) Generalization hierarchies are explicitly classifications and may be used at different times and by different persons for at least three significantly different ends.

a) Mere description - it is simply parsimonious, on occasion, to construct descriptions by constructing a classification that avoids needless redundancy. Generalization hierarchies in data models seem to have been introduced with this end in mind and when their use is strictly (unambiguously) limited to this end they are not problematic. E.g., biological taxonomies achieve economy in that we do not have to repeat, for each kind of bird, that it has wings and lays eggs.

b) Status assignment - this is a stronger use of classification in that we assign entities to classes

according to the ways in which it is appropriate to treat (or not to treat) those entities. For example, when my bank classifies my account as "overdue" or "overdrawn" it is a statement that, in effect, says it is not appropriate to grant me additional credit. This use of classification can be problematic in that there is nothing in the data model schema that represents the bank account to declare that the classification is to be used in this way rather than as a mere description. Any agreement between observer and decision maker to use the classification in this way is purely by convention and totally outside the scope of the model.

c) Appraisal - This use of classification gives yet another step beyond status assignment in that it carries motivational significance; i.e., not only is it appropriate to treat an entity in certain ways, but the entity is treated in one of those ways. E.g., when my mortgage is appraised to be "in default" the bank will initiate a foreclosure. As with status assignment, any such use of a classification is elliptically embedded in the schema by convention and agreement.

To more vividly exemplify the problems that occur when a classification is understood differently by observer and decision maker, one example should suffice. Consider a person admitted to a hospital with symptoms of a serious virus, and his condition being described in a data model by his insertion in a classification schema. The epidemiologist might use this information as a mere description to update records concerning the incidence of the disease. The physician might interpret the classification as a status assignment in deciding what tests or treatment might be appropriate on furthering his diagnosis. The night nurse might interpret the classification as an appraisal by the physician and administer certain drugs to the patient. Depending on who made the initial observation and classification; and who accesses and interprets its intention; the results could obviously be catastrophic. And there is no place in the model (i.e., there is no construct) for recording what the intended use of the hierarchy is.

18. A NEW SYSTEM?

In the same vein as Kent concludes his book, it is equally appropriate to conclude this report by considering the possibility of a new data model. It is, of course, only a possibility since a model based on SA has yet to be developed. But it is not premature to consolidate what has been discussed in this report in terms of the characteristics such as implementation would exhibit, and in particular, how those characteristics would differ from existing and other proposed implementations.

18.1. System Architecture

A system based on SA would exhibit an architecture squarely in between that of database systems on the one hand, and so-called knowledge-based systems on the other. And, hopefully, this would help to clarify rather than further confuse an already confusing distinction by showing this distinction to be no more than that between process and structure which, as already pointed out, are equivalent. We can describe a knowledge-based system (expert system), as proposed by most practitioners, to be a network of rules (e.g., if <A1, A2,...,An> then C) in which the nodes represent logical implication; and the system operates as an inference engine, traversing the network according to the rules of logic to compute conclusions. Using the same topological paradigm, an SA system could be envisioned as a network in

which the nodes represent States of Affairs; the edges represent the part-whole relationships; and the system operates as a "distinction engine", continually recording which States of Affairs exist and which do not by virtue of having the necessary constituency. The difference is, again, a practical (behavioral) one. The SA system takes States of Affairs to be the case unless there is a reason (stimulus or input) to make it otherwise. The inference system considers everything in doubt until proven true.

18.2. System Protocol

Both the database and knowledge-based systems operate by the familiar question/answer protocol. Updates to the rule clauses and the data (ground clauses) are routinely collected but are only considered a response to an open-ended question (open-ended in that any question has, in principle, an infinite number of possible answers). By contrast, an SA system is closed; i.e., it can only make a predetermined number of distinctions among the States of Affairs defined, regardless of how the data are permuted. Thus, rather than a Q&A protocol, an SA system is better suited to a "cogitating" protocol; i.e., acting on each input to continually refine its distinctions, and always having its latest distinctions available for answering the unit question "What is now the case?

18.3. System Applications

As rule-based and SA-based are logically equivalent, each could, in principle, be applied equally well in any problem domain. But there are clearly practical, behavioral differences. Domains in which the rules are extensive, and there is a premium on completeness and correctness (e.g., medical diagnosis) will probably continue to be better served by rule-based systems since, being open-ended, they deduce only that which can be proven true and nothing else. However, the price for this logical precision has been and will probably continue to be very slow response. An SA-based system, being closed-ended, is much more tolerant of domains in which the part-whole relationships are generally incomplete and imprecise (e.g., intelligence analysis) since the system needs only enough data to distinguish among possible conclusions, without having to rigorously prove its results. It is consequently likely to be quite fast but at the expense of being error-prone. In this regard it will probably resemble the phenomenon of human judgement (particularly "snap" judgments) more than its counterpart.

19. REFERENCES

[1]. Kent, W., Data and Reality, North Holland Publishing Co., New York, N.Y., 1978.

[2]. Ossorio, P.G., "What Actually Happens:" The Representation of Real World Phenomena, University of South Carolina Press, Columbia, S.C., 1978.

[3]. Senko, M.E., personal correspondence, 1975 until his untimely death in December 1978.

[4]. Bernstein, P.A., "What Does Boyce-Codd Normal Form Do?" in VLDB, 1980.

[5]. Chen, P.P.S., "The Entity-Relationship Model: Toward a Unified View of Data," in ACM TODS, March 1978.

[6]. Smith, J.M., Smith, D.C.P., "Database Abstractions: Aggregation and Generalization," ACM TODS, June 1977.

[7]. Codd, E.F., "Extending the Database Relational Model to Capture More Meaning," ACM TODS, December 1979.

[8]. Hammer, M., McLeod, D.J., "The Semantic Data Model: A Modeling Mechanism for Database Applications," in ACM SIGMOD, 1978.

[9]. Op. Cit. [1].

[10]. Ossorio, P.G., "Ex Post Facto Formulation," in Advances in Descriptive Psychology: Volume 5, JAI Press, Inc., Greenwich CT, 1986.

[11]. Ossorio, P.G., State of Affairs Systems: Theory and Techniques for Automatic Fact Analysis, RADC Technical Report 71-102, Rome Air Development Center, Griffiss AFB, 1971.

[12]. Wittgenstein, L., Philisophical Investigations, The Macmillan Company, NY, 1953.

[13]. Op. Cit. [6].

[14]. John Wisdom.

[15]. Ossorio, P.G., lecture at the University of Denver, 1983.

[16]. Webster's Ninth New Collegiate Dictionary, Merriam-Webster, Inc., Springfield MA, 1983.

[17]. Op. Cit. [2]. Note that this definition forestalls the paradox about the observer influencing the event (e.g., if a tree falls in the forest and nobody hears it then did it make a sound?) In a world of persons there is no paradox. If nobody observed the falling of the tree, then nobody will act differently because of the event and thus, for all practical (behavioral) purposes, it may as well never have happened. And if only later someone observes that the tree did fall and acts accordingly, then that is the observation of he event and, therefore, that is what we act upon. There is no need for meta-physics.

[18]. Jeffery, H.J, "A New Paradigm for Artificial Intelligence" in Advances in Descriptive Psychology: Volume I, JAI Press, Inc., Greenwich, CT, 1982.

[19]. Meyers, P.M., Data in the Context of Behavioral Potential, Final Term Paper, Seminar in Descriptive Psychology, University of Denver, Denver CO 1983.

[20]. Op. Cit. [1].

[21]. Levin, M., Schneider, L.S., "A Dual Model Approach to Integrating Data," Technical Report, Department of Mathematics and Computer Science, University of Denver, Denver CO, 1981.

[22]. Op. Cit. [10].

[23]. Ossorio, P.G., Place, LRI Report No. 30a, Linguistic Research Institute, Inc., Boulder CO, 1982.

[24]. Op. Cit. [1].

[25]. Senko, M.E. (et al), "The Data Independent Accessing Model," in IBM Systems Journal, IBM, Inc., December 1973.

[26]. Op. Cit. [1].