REIHE INFORMATIK
TR-2008-005

A FORMAL REPRESENTATION OF THE METHOD OF LEARNING

By

Hendrik O. van Rooyen - University of South Africa

Franz Stetter - University of Mannheim

“H O van Rooyen” < hopete@intekom.co.za>
“Franz Stetter” < stetter@uni-mannheim.de>

Mannheim, October 2008

A FORMAL REPRESENTATION OF THE METHOD OF LEARNING

by

Hendrik O. van Rooyen - University of South Africa

Franz Stetter - University of Mannheim

"H O van Rooyen" < hopete@intekom.co.za>

"Franz Stetter" < stetter@uni-mannheim.de>

Mannheim, October 2008

CONTENTS

1.	Introduction	4
2.	Scientific Method	5
3.	The Method of Learning	11
4.	The Method of Mathematics	17
5.	A Summary of Relevant Formalism	19
6.	The Knowledge Hypernet of a Formal Language	33
7.	Modes of Reasoning	39
8.	A Exercise in Speculative Reasoning	43
9.	Suggested Definition of a Plan	45
10.	Example of a Plan: An Information Ladder	50
11.	Conclusion	54

References 57

KEY WORDS

abstraction (inductive), analogy (structural), cluster, concept–relationship knowledge structure - CRKS (isomorphism of; translation of;), context schema, deduction, derivation, formal schema, graph (association), hypernet, information ladder, interpretation, knowledge hypernet – KH (abstraction to, interpretation of, isomorphism of, realization of), learning, meaning (in a context), model, mode of reasoning (association, construction, intuition, deduction, induction/analogical), noise, observable, primary (concept-name, sentence, statement), relation net, relationship (statement of), representation, rule of inference, theory (domain of, functionality of, falsification of), translation, trial-and-error, unpredictable process.

ABSTRACT

We show how the method of learning in science, education, and mathematics can be represented by a knowledge hypernet (KH) and its concept-relationship knowledge structure (CRKS) interpretations. We conclude that the method of learning is invariant over the three fields. Some background [from GVS1999, VGS2004] is required of the reader, but the crux of this is summarized in section 5. The paper is particularly suited for teachers of science, particularly physics, and of mathematics, and in the philosophy of science, but is also relevant for educators at every level of instruction. Those working in the fields of cognitive science and knowledge representation can also benefit from this paper and its main references.

1. Introduction

The background formalism for this paper, with some examples, is contained in [GVS1999] and [VGS2004]. In sections 2, 3 and 4 we deal with the method of science, of learning, and of mathematics respectively. All the material is well known, but we put a slightly new slant on it. In section 5 we present a summary of the formal background required, together with an illustrative example. In section 6 we apply this background to represent the structure of the formal languages of mathematical logic, and thereby the method of learning in mathematics. We then indicate how this formalism applies in the same way to the method of science and to the method of learning.

Since learning depends upon designing, modifying, and eventually replacing, plans, and designing a plan requires the application of reasoning, we deal with modes of reasoning in section 7 and suggest a formal definition of the notion “plan” in section 9. Section 8 presents a trivial example. In section 10 we introduce an idea for a particular “plan”, an information ladder, that stores information. Section 11 is a brief summary, and enumerates a few conclusions.

The overall purpose of our work has become a desire to propose universal basic education, in the specific sense indicated in this paper, as the stimulus to a movement of the human species toward an Age of Reason. This is our final and summarizing report on this topic. For a historical view see [VR1976] and [Gel 1993].

2. Scientific Method

An invariant observation is an observation that is common to a number of collections of observations. We choose some of these invariant observations, called the relevant observables. In each collection of observations there are usually relationships involving two or more observations - if not, no science. We single out, from these relationships, some of them, if any, that are invariant and involve observables. The chosen observables and those chosen relationships, more specifically statements of them, constitute our "model" of those collections of observations.

All that is different from an invariant in those collections constitutes the "noise" against which that invariant stands out; the contrast against which it "appears". An invariant is always given a name; the foundation of language development. In the case of an observable the name is a "word"; for a relationship it is one of the many statements of that relationship, a "sentence". Thus our "referent free" model of words and sentences can be represented by symbols, for words, with word combinations for sentences, in a variety of specific interpretations.

We now apply reasoning to the model to generate predicted observables, and predicted relationships among them, in a specific collection. This is followed by empirical test of those predictions. If a prediction falls within the limits of experimental error then we say this supports the "theory", where "theory" means the combination of the model and the reasoning applied to it to generate the predictions. Notice that the predictions are from the model, but are tested in the collection! If an observable or relationship is not predicted then we say that the theory is incomplete; it must be modified to extend its range of predictions. If any one or more predictions is found to be in conflict with empirical test then we say that the theory is false. It is the nature of science that every theory is falsifiable: observations and relationships change, the model is necessarily more simple than all the collections of observations it is supposed to predict, and the domain of observations that we want to predict is always growing. The domain of observations and relationships in which the theory makes "correct" predictions is thus always limited: it is called the "domain of functionality" of the theory. By falsifying a theory we not only find out about its domain of functionality, but we are forced to replace the model or the reasoning, or both, i.e. the theory. Thus science is an enforced and continuous process of learning.

Every invariant is abstracted against contrasting noise, and is named. The process of abstracting an invariant from a collection of things, or “examples” or “instances” of a thing, is called inductive abstraction. This leads into section 3.

In broadening these comments we repeat them for emphasis. An invariant is something common to a collection of otherwise different situations, or otherwise different occurrences of the same situation. In each situation considered, what is different from the invariant constitutes, together for all those situations, the noise against which that invariant stands out. Without noise there is no invariant! Every invariant is given a name – the “naming words” of a language - and naming invariants is the most basic human need. Given a collection of otherwise different situations, we use the most fundamental learning procedure, trial-and-error, to inductively abstract the invariant(s), i.e. to find out what is common to many different situations (the “process of induction”). This is called basic procedural learning: naming invariants is the basis of communication skill, of declarative learning.

Given a selection of situations we make observations. These are sensory perceptions, and we try to order, or structure them by finding relationships among them: for example, we invent (human) concepts such as the basic common sense concepts of direction – “towards that object” - space, and time. It is known that there is a smallest possible natural unit of space – about 6.625×10^{-34} meters – and a smallest possible natural unit of time – about 2.21×10^{-42} seconds – so, at the foundation of physics, one should use discrete mathematical modelling of space and time. At large scale this will “average out”, but it has important implications for the philosophy of science; in particular physics. For example this implies a minimum wavelength, maximum energy condition. (To place oneself in a favourable position to comment on the philosophy of science, one should have developed some new theory in science and have performed experimental work. The authors do not claim this distinction for themselves: nothing in this section is new; just perhaps the way in which it is stated.)

The world of our perceptions of the universe is complex – it is in and round us, and living forms are all part of it. We must divide it up, and even then every piece is complex! For every piece, every situation, we must simplify to try to “grasp” it, to partially “understand” it, i.e. to partially order it, find a structure, a pattern, of sorts within it. We must try to find relationships among these structures, if possible, to form a “bigger picture” of what is going on in and around us. How are we to do this? What method can we follow? We already see that we can never complete the task because we must simplify for example.

From all our observations of a sample of situations, we find only the invariant observations, if these exist. We choose some of them and these are called observables of the situations. Next we consider the observed relationships among the observables, and choose only some of those that are invariant. These are then the observables and relationships with which we work. The observables are our initial data, and those relationships our initial information, our structured data.

We now construct at least one model of the situations considered. This is a symbolic representation, where the symbols are signs, sounds or objects, each symbol depicting an observable, with the relationships among the observables holding between the relevant symbols that depict, or denote, the relevant observables.

We have captured, in a "map" or model, an inductively abstracted invariant part of the situation considered.

Next we apply one or more of the modes of reasoning - associative, constructive, deductive, intuitive, and inductive (i.e. analogical) – to the model, building a theory of the model, to try to predict observables and relationships, among them, from the model, for the situations considered. (The model and reasoning together constitute a "theory".)

These predictions are then empirically tested in the situations considered. If a test produces a predicted result within the limits of experimental error, then that test supports the (model and) theory. If not, we must modify or replace the model and/or reasoning i.e. the theory: they may be incomplete descriptions of the situation, or they may show at least one prediction to be false, i.e. falsify the theory. Within the domain of support the model and theory are said to be functional, i.e. they can be used as if correct. Outside of the domain they are false, and modification or replacement of them is necessary.

By force of circumstances, by definition of scientific method, every model and theory must be incomplete or be falsifiable. By falsifying a model and theory we find out its domain of functionality. If incomplete the theory must be extended: it will inevitably be falsified. By, of necessity, modifying or replacing it every time, we are committed, as scientist, to continuous never-ending learning!

There is no absolute notion of truth in science. The invariant relationships with which we start are called postulates, and "truth" in science means the following. If we assume that the postulates are "true" then the predicted consequences of them must be empirically tested for "truth within the limits of experimental error". The object of science is to find predictions that are not "true", i.e. to prove the relevant theory incorrect – to falsify it – so that we can modify it or replace it, thereby learning.

Every scientific theory is false, and falsification implies the necessity of continuous learning. There is no way to prove any scientific theory correct for any situation/phenomenon. The novelist Louis L'Amour, in "The Haunted Mesa", has the statement: "What is today accepted as truth will tomorrow be only amusing". In Open Vistas, Margenau remarks that "Science recognizes eternal problems but no eternal truth". There is not now, and never will be, a "theory of everything".

The whole business of science is a grand and (hopefully) controlled game of imagination, of trial-and-error, of questions "why ...? and "what if ...?" in an ever changing and joyful never-ending process of continual learning. It is vital to realise that science is not attempting to understand, to explain, how things work. It attempts only to anticipate collections of perceptions that will arise, by finding relational structures among perceptions.

In a collection of situations we choose (invariant) observables and (invariant) relationships among them. Every observer must find these in every one of the relevant situations, "everywhere and everywhen". This is called repeatability, or objectivity, as opposed to a subjective, idiosyncratic view of those situations. Thus, to summarise: repeatability, observables, relationships among them, model, reasoning**,** theory, predictions, experimental test and falsification. Without any one or more of these – no science! This determines the "domain" of science: outside that domain science can do nothing. One must distinguish, clearly, between faith and falsification; given any topic, faith precludes falsification and vice versa.

Finally, a scientist must be able to live with open questions. If asked a question, s(he) must first try to decide whether or not the question is meaningful. If it is, then s(he) must try to find an answer. If not, then the question must wait until a decision on meaningfulness can be made: in the meantime, live with it!

Given a theory, if an observation or collection of related observations is not predicted by that theory, then we say that theory is incomplete. We must try to extend the domain of functionality of that theory. If a prediction of the theory fails empirical test, then we say that theory has been falsified; that the theory is false. While we may continue to use that theory within its domain of functionality if convenient, it must be replaced. As we extend the domain of predictions, of every theory, it will eventually be falsified because it is, by definition via model and reasoning, always more simple than that piece of the real world modelled.

To close this section we look briefly at a trivial example, the so-called ***The Königsberg Bridges Problem,*** of scientific method in action. The first paper on what is now a vast and interesting topic in mathematics, called Graph Theory, was written by the Swiss mathematician Leonhard Euler, and appeared in 1736. He began his paper by discussing the so-called Königsberg Bridge Problem. The city of Königsberg is located on the banks, and on two islands, of the river Pregel, and the various parts were connected by seven bridges as shown below. On Sundays the populace would take their customary stroll about the town, and a much discussed question was that of whether it was possible to start from home and return there having crossed each bridge exactly once.

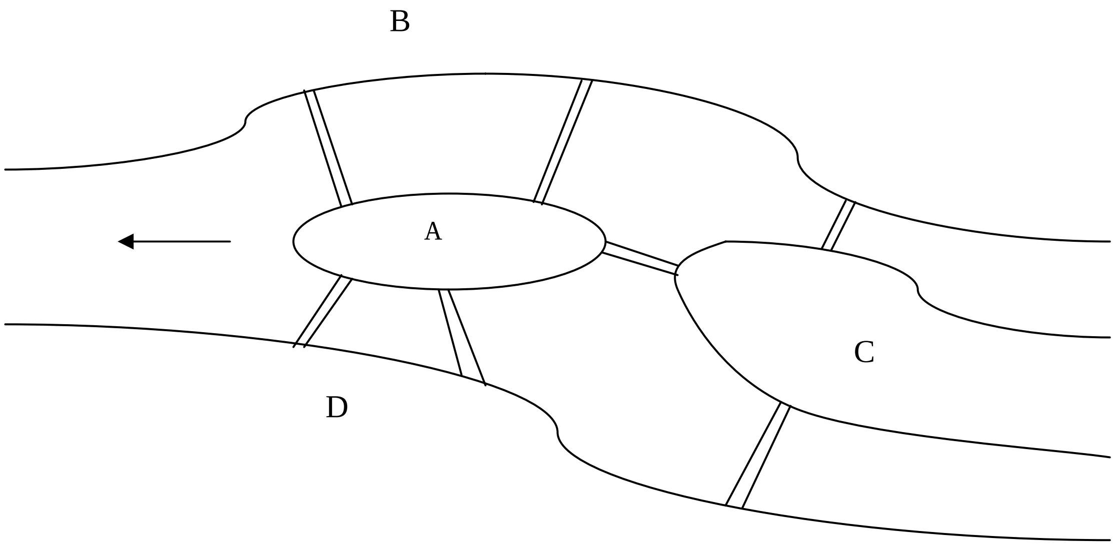


Figure 1: The Königsberg bridges

Here the observables are taken to be the four land masses A, B, C, and D, and the invariant relationships between them the 7 bridges. Euler represented the situation in the form of the point-line diagram, the reduction of the situation to a model.

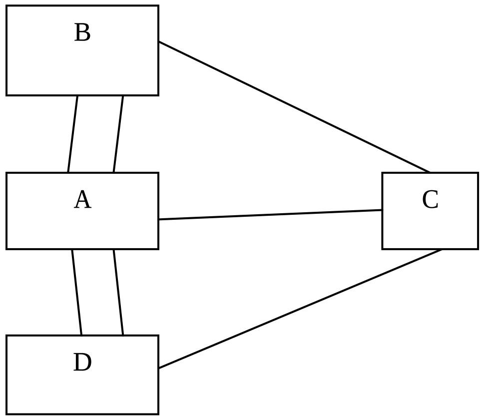


Figure 2: Model of Königsberg bridges

Since the number of bridges meeting each land mass is odd, no land mass can function as “home” for the proposed walk: every time we use a bridge once (only) we will have to return by another bridge, and then there will be a bridge left by which we must leave that land mass. Thus we have invariants and a model, and a theory which predicts that the journey is not possible. But now along come two people in a boat! The theory is falsified. The model was too simple to cope with the new observations: we must modify or replace the model and theory.

For students of Physics, for example, it is essential to know what scientific method is. Apart from this knowledge for general use in physics and multidisciplinary research, there are certain standard techniques that are part of teaching physics but are of broad use in life. For example, when trying to solve a numerical problem in physics the standard general approach is

- estimate the value of the answer, at least to the order of magnitude,
- check for consistency of units,
- use a theoretically justified recipe/algorithm to find the answer.

To close this section we should note the following about science. “Reality” is what is observed, and it depends upon the individual observer. “Theory” is the never-ending attempt to anticipate what will be observed by the individual observers, and it must necessarily start with what is common to the observations of a number of observers.

3. The Method of Learning

In order to teach effectively one must be aware of the basics of the method of learning. Much work has been done in this topic, and more must yet be done. Some of the main points are outlined in this section, followed by some consequences for the method of teaching.

Briefly stated, we have the following method of learning for both teacher and learner.

- present opportunities to find invariance by observation: the basic evidence. This involves trial-and-error. It constitutes the first part of "how to do it" learning, the procedural facet. "Look at these things and find what is the same in all of them". For example, we should learn to speak and to read by this method.
- mimicry, the first part of "what to do" learning, the declaration facet. A teacher should set a good example to follow.
- apply reasoning to the evidence, i.e. the data, to find relationships among the data, the information. Next find the patterns/structures in those relationships, i.e. the knowledge. This requires that we create, and follow, a plan.
- in all this, develop communication skills, communicate (declaration) first with yourself and then with others. Without communication there is no teaching and no learning!.

This process speaks of the same method as that of science: both require inductive abstraction of invariants, setting up a model, and making predictions, to be tested here by problem solving and examinations. In both one is driven, by curiosity and imagination and initiative, to ask questions. Question everything. Without questions there are no problems, so no solutions, no realisation that solutions, plans for solution, are temporary, and thus no learning. Reasoning is applied to the evidence in order to create a plan to solve the problem that follows from the first question: "what is the same"?

In the following three sections we will see how this same method is used in mathematics.

We need to nurture and teach the innate traits of humans, most importantly the following.

- Self-discipline, respect, and co-operation
- Curiosity
- Imagination

- Innovation
- Asking questions. Question everything. Without questions there are no problems, and thus no learning. Questions lead to problems, problems are a source of speculation, and speculation is an essential part of learning.
- Access current relevant evidence by observation, and apply reasoning to design a “plan” to make predictions to test.
- Realize that no plan is final. A plan must be modified, and eventually replaced, because the “current relevant evidence” is always changing.
- Develop precise communication skills in conjunction with learning: imprecise questions must be refined. This skill involves the sophisticated use of symbols: written, sound and pictorial.

We claim that changing plans is a major facet of learning and that learning and teaching is the “meaning of life” in the sense that it is an invariant of all life forms: DNA evolves, learning, and plans construction, teaching! Teaching is the process of learning together: to teach is to learn.

This is, we claim, what is fundamental to all education: its basis.

All learning starts with trial-and-error (procedural) search for invariants in seeking an answer to a question (which may not at first be verbalized). Every invariant must be named (declaration), the start of language. We apply reasoning to design a plan that generates predictions. We test those predictions to find out in what domain of experience, if any, the plan is “functional”. Every plan must be modified, and eventually replaced: that is learning!

This is education, not the data in syllabuses, which needs to be placed in a context, a structure, a pattern, to be meaningful.

Now in these remarks one concept is probably most critical: reasoning! What does the term mean? We are swamped with paradigms: for example democracy, the out of Africa, and the Big Bang. A paradigm is best regarded as a working hypothesis designed to stimulate questions, and to be questioned. Paradigms become the deadly enemy of reason when they become propaganda: a paradigm should not specify reasoning, but invite it. What “reasoning” does mean is approached in our work and dealt with in section 7. Reasoning is innate, and should be explained, nurtured, and trained by constant use.

There are some clear indications of the role of the teacher. One is that one should start with trial-and-error, procedure based, search for (invariant) observables and (invariant) relationships involving observables. Then apply reasoning to generate more relationships, and test the predictions. Modify or replace “plans”.

Consider a fundamental example; establishing the sound symbol “red” for a child. In brief, one would make statements such as “red ball”, “red car”, “red roof”, “the gate is red”, “a red jersey” and so on, pointing each time to the object in question and making sure that the only common sound is “red”, thus providing the contrasting noise which should preferably be meaningless to the child. We establish an invariant and a name for it. Thus we establish “naming words”, and then, later, “connecting words” so that we can establish the invariant “sentence” in a similar way. Sentences describe relationships, and one can then introduce secondary concept-names, also invariants, such as “colour”: “red is a colour”, “this colour has the name green” and so on. A relatively standard approach to teaching emerges, and with imagination it can be applied to any topic at any level.

Whatever the study material, your approach as teacher should thus be as follows.

- ask initial questions.
- suggest a general way to find answers.
- seek relevant invariants and relevant invariant relationships, by trail-an-error, to start with.
- make an initial plan, i.e. structure the study material, perhaps in the form of a concept-relationship knowledge structure (CRKS) - see [GVS1999] and section 5. (A CRKS is a plan: designing it together involves reasoning, and it provides a presentation tool.)
- ask questions at each stage of construction of the plan, and let the answers build the structure: speculate with each question. If the structure being built is a CRKS then, by assisting with the precise statement of relationships and questions one teaches communication skill and reasoning. By adapting the choice of “teaching routes”, i.e. presentation strategies, one can adjust teaching to suit classes and individuals. See section 5.
- at the end of a lesson/course, look ahead by asking questions.
- question the structure: notion?, alternative structures?, refine questions in view of “new” knowledge?, what new questions now arise?
- empirically test, by examination for example, to see whether the structure/plan has been assimilated - see [GVS1999] - by the students and can be used to answer appropriate questions.

One is exposing what education really is, regardless of the actual topic being taught, and this method should be exposed and made specific in every topic taught; in every lesson. "Do you see how we are learning?": it must be constantly pointed out, until it becomes automatic for everyone.

As to actual study material, one needs to motivate the student. Certainly exposure to the method, the approach above, in an explicit manner, pointing it out specifically to students as the essence of learning, including specific exposure to the modes of reasoning, will move us towards an Age of Reason, but this is motivation for the teachers. What of the students? We must, as teachers, provide information on the basis of which the student will be able to ask questions: does the study material qualify? If the structure being built is a CRKS then, by assisting with the precise statement of relationships and questions one teaches communication skill and reasoning. By adapting the choice of "teaching routes", i.e. presentation strategies, one can adjust teaching to suit classes and individuals. One may say that certain things must be known so that the students can "fit into society as it is at the present". But to say "learn this by rote because you are going to need it" is not a student motivation; is definitely not education. If a teacher says "2 + 3 = 5" there must be a reason for the statement; a structure into which it necessarily fits as part of a plan, or the statement is not meaningful and thus not motivated. "Rote learning" is usually of data: there is little relational information, so there are no patterns of information, i.e. no knowledge. Thus "rote learning" is not learning: it results in assumptions of "absolute truth" and prohibits certain questions, so it is destructive and devoid of real motivation. Real motivation of study material is "we are learning this together so that we can ask questions and speculate: further, it will be temporarily useful to us in real life till we learn more".

In all this, teachers must bear in mind that the situation changes, so the relevant information changes: we must all continuously learn together in co-operation. And the phrase "in co-operation" implies a multi-disciplinary approach to all questions and the resulting problem solutions. Thus:

- ask, and encourage, questions such as "why?"
- encourage speculation such as "what if ...?"
- to state "because it is so" or "because I say so" is a non-motivational taboo.

The current craze for teaching "mathematics", i.e. elementary, structurally unmotivated "numeracy", before adequate development of social and communication skills, it horrifying. One should start with the practical notions of observer, direction and position, and move to measurement by pacing out distances, introducing “unit”, and “names” for the “number of paces”. Then move to addition as successive pacing as a “discovery by pacing”. Later, one can introduce the set definition of the counting numbers and the notion of strings to associate with “pacing”, leading to addition and counting. To get at the set definition of the numbers one would approach it via “collection” and “matching”, notions that are fundamental, practical, and easy to establish through play situations – see the example of a CRKS in section 5. One can approach all this through play situations, and questions leading to the discovery of, and naming of, invariants, and one can associate all the resulting relational information into a simple pattern: a question stimulated and directed extended discovered play pattern. In this sense, all learning can be regarded as an extended play pattern, when viewed appropriately, which never stops.

It is futile to claim that this “extended play” scenario does not apply to certain study material: it is the teachers task to use imagination to show how it applies!

The mission of every teacher at every level in every topic is to teach reasoning using leading questions and play, by trial-and-error, with objects and ideas. In this, speculation is an essential step. In teaching one should perhaps make the following distinction in terminology:

- “Objectives” - specifically stated lesson and topic goals
- “Outcomes” - general goals of education, which should be always the same and should start with self-discipline, mutual respect and co-operation.

A teacher’s task in general is to guide discovery and to introduce names for invariants, to ask and encourage questions, and to explicitly expose speculation, planning and reasoning. Trial-and-error is a vital aspect of learning, and is the initial source of creative play; of the creativity which must be nurtured, of innovation.

Since all parents will be teachers in an Age of Reason, we give a hint about guiding your child to mental maturity.

- encourage questions
- with every answer, and with every instruction, give a cogent reason. The child will not fully understand the reason, but will grow up looking for, and trying to express, a reason for everything.

We briefly mention some taboo's in stating objectives ("outcomes"): some "reasons" are also taboo. We must guide trial-and-error experimentation, and we must never say the following.

- you should know the meaning of ...
- you should understand ..., since nothing is ever known or understood except within a specified context, and for reasons
- because I say so
- because it is so
- God made it so; feeble excuses for not providing a reasoned answer.

Rather say "this is the best answer/reason that I know of at the moment: ask again later", or "I don't know, so lets try to find out". Mistakes are inevitable, and we must learn from them: this is scientific method in a nutshell! Do not make the mistake of having too little time to guide one's children: rather not have children than treat their mental futures as a lottery!

Finally we must mention the most difficult aspect of teaching: each student is an individual, so, especially in the early years, individualization of approach is required. One should exploit individual characteristics rather than trying to force everyone into a standard mould. Individualization is closely related to innovation, so the teacher should adapt to it. Study material can be standardized, but not people!

4. The Method of Mathematics

We will expose the method of mathematics in the course of this section and the next. Mathematics is a creative art: we create and describe symbolic structures, so we are dealing with languages. It is with the method built into these languages that we deal here; the method of mathematics. We start with the basis of mathematical logic in simplified form.

By a formal language L we mean the following. L has a countably infinite set of symbols called the alphabet of L, where a set is a collection of distinct symbols. L has a countably infinite set of expressions X, where each expression is a finite string of symbols of the alphabet. L has a proper subset W of X; each element of W is called a word of L. A word is a finite string of symbols. L has a set of sentences S. Each sentence is a string of words, with each pair of contiguous words separated by a symbol, from the alphabet, called a line symbol. A sentence is a finite string of words and appropriate line symbols.

Which expressions are words? Which strings of words are sentences? We start by remarking that to answer both questions, i.e. to specify W and S, we must give membership conditions. We will anticipate that “words” are going to name “things”, and that “sentences” are going to name “relationships”. Both W and S are going to be countable sets.

Now to specify the words, i.e. what forms of expressions are words, it is best to see, in each case, what we are going to name with the words. We will deal with this later, but we can deal with the specification of which strings of words are sentences now. We do it by laying down sentence specification conditions called rules of inference (or deduction). These constitute the grammar of L, and L is a “formal language”, i.e. the rules are fixed; the grammar is absolutely specified for L. This does not mean that we can now write down all these rules. What we do here is to define a rule of inference as a finite string of “word spaces”. To specify S we now proceed as follows.

A countable proper subset A of W is called the set of axioms of L. It is from this foundation that all sentences of L are constructed. (Note that an axiom in mathematics is not a “self-evident statement”).

A primary sentence of L is defined as follows. Given any rule of inference, if all spaces but the last are filled by axioms then the final space is filled by a non-axiom word “inferred” from the previous words in those spaces. The final word is called a theorem. (It is an inferred word).

In general, a sentence of L is defined as follows. Given any rule of inference, if each of the spaces but the last one is filled with an axiom or a theorem then the final space is filled by a non-axiom word inferred from the previous words in those spaces. The final word is called a theorem.

In every case, the sentence is called a proof of the theorem. (We usually meet arguments intended to convince the reader that at least one (formal) proof of the relevant theorem exists.) As in science, there is no notion of absolute truth in mathematics: in both truth is relative. Thus, if the axioms are “true”, then the theorems are “true”, where, as for science, we note the “if” in the statements.

To define W we start with the alphabet: each symbol is a word. We then specify a few "word forms", forms with spaces into which we fill words. In the same way as we recursively defined the theorems using a base of axioms and the rules of inference, we now recursively define the word set W using as base the alphabet, and the word forms. The axioms are defined by "axiom forms", forms with spaces into which we fill words. There are again just a few of these forms, and they specify a proper subset of W: the axiom forms generate a proper subset of W.

Thus the words are defined from the alphabet using word forms, the axioms using special word forms called axiom forms, and the sentences using sentence forms called the rules of inference. The theorems (words) are defined from the axioms (words) using the rules of inference (sentence forms). (Note: for "spaces" read "variables".) We will meet an example, proposition logic, in section 6, but first we revise some required background.

5. A Summary of Relevant Formalism

A Relation Net is a pair <A, T> where A is a set of vertices, T is a set of “tuples”, i.e. sequences, $<a_0, a_1, ..., a_{n-1}, a_n>$ over A, and each tuple is an edge from a_o to a_n labelled with $<a_1, ..., a_{n-1}>$. The set of members of A that occur at least once in $t \in T$ is called the tuple set of t. Each different $t \in T$ is regarded as specifying a different relationship among the members of the tuple set of that t.

A Formal Schema is a relation net <A, T> for which every vertex is a concept-name. In each $t = <a_o, ..., a_n>$, any member of the tuple set of t can of course occur any finite number of times. Finally, the tuple set of t has cardinality of at least 2, and every vertex occurs in some tuple with at least one other vertex, and <A, T> has at least one primary and at least one goal – see later and [VGS2004] - and has no circuits. A formal schema is said to be complete if it has no isolates - [VGS2004] and later.

A Concept-Relationship Knowledge Structure (CRKS) is a complete formal schema <A, T> in which every vertex is derivable, where a vertex is a derived vertex iff it is a primary or there is a tuple that ends with that vertex and is such that every other entry in that tuple is of a derived vertex. Every edge in a CRKS is a derivation edge, and every path from a primary is a derivation path [VGS2004]. Each tuple t of a CRKS <A, T> represents a specific statement of relationship among the members of the tuple set of t, which statement corresponds with t in the sense of mentioning each of those members at least once. Each statement can be regarded as a rule of inference.

Given any $a \in A$ of a CRKS <A, T>, the context schema of a in <A, T> is that formal schema induced by the subset, of T, of all its tuples which have a least one entry of a. The context schema of a in <A, T> specifies the meaning of a in the context of <A, T>. Its vertex set is the set of all vertices that occur at least once in any of its tuples. If $a \in A$ is deleted from a CRKS <A, T> then all the tuples of the context schema of a in <A, T> are deleted from <A, T> as no statement of relationship involving a can then be written: if we delete a we delete every tuple in which a is at least one entry.

By a fast access cascade from a subset $A_0 \subseteq A$ in a CRKS <A, T> we mean a sequence of sub-relation nets of <A, T>, $<A_0\ \varnothing>$, $<A_1, T_1>$ ---. T_{i+1} is the set of all $t \in T$ that start with any $a \in A_i$ and

A_{i+1} is the union of all its predecessors together with all vertices incident from any $a \in A_i$ using T_{i+1}, i = 0, 1, …. By a limited access cascade we mean a similar sequence in which T_{i+1} is the set of all $t \in T$ that start with any $a \in A_i$ and for which every entry, but possibly the last one, is an entry of a member of A_i. A cascade can be stopped at any stage. It will stop automatically when all the goals of <A, T> have been accessed, as it then just continues to reproduce the same result. It can be shown that a complete formal schema is a CRKS iff it can be generated by a limited access cascade for which A_0 is the set of all its primary vertices [VGS2004]. (Recall that $a \in A$ in a formal schema is a primary iff it is not the last entry in any tuple, and a goal iff it is not the first entry in any tuple. It is an isolate iff it is neither a first, nor a last, entry in any tuple, but it must occur in at least one tuple [VGS2004].)

One use of CRKS's is in organizing study material in teachable/learnable form – see [GVS1999], and below.

By a Hypernet we mean a pair <A, E> where A is a set of vertices and E is a set of subsets of A, the edges of <A, E>.

Given a CRKS <B, T> we define an abstraction of <B, T> to be a 1-1 correspondence of B and A of a hypernet <A,E> such that the tuple set of every $t \in T$ is mapped, vertex by vertex, to a separate edge in E and there are no other edges of <A, E>. Such a hypernet is called a Knowledge Hypernet (KH), and the same KH can be abstracted from different CRKS's . Given two of these CRKS's we can have two tuples, one from each, that abstract to the same edge $e \in E$: they are different tuples, arising from different statements of relationship among different sets of concept-names, i.e. the tuple sets are generally different, but both abstract to e; they have something in common, an invariant. We think of e as a relationship among its members, and the two tuples as specific statements of relationship that abstract to the same general relationship encapsulated in e. A CRKS edge from a_0 to a_n becomes the tuple set of that edge, which becomes a derivation edge, from the vertex corresponding to a_o to the vertex corresponding to a_n, in <A, E>. See [VGS2004].

By an interpretation of KH <A, E> we mean a 1-1 correspondence of A and the set of vertices B, of a CRKS <B, T>, which mapping is such that every $e \in E$ is mapped, vertex by vertex, to the tuple set of precisely one tuple $t \in T$, and every tuple of T is covered by the mapping. Each t is then associated with a specific statement of relationship. Thus in an interpretation of a KH <A, E> we will, given any e $\in$ E, choose any two members of e as the first and last entries of a corresponding tuple t, and in t there will be at least one entry of each member of e.

Given a CRKS $<B_1, T_1>$ that abstracts to a KH <A, E>, and an interpretation $<B_2, T_2>$ of <A, E>, it is easy to see that the abstraction followed by the interpretation constitutes a 1-1 correspondence of B_1 and B_2 that preserves (single member) relationships (but not statements of relationship). We call such a mapping, from $<B_1, T_1>$ through <A, E> to $<B_2, T_2>$, a CRKS isomorphism.

Given an interpretation of a KH <A, E> that produces a CRKS <B, T>, we call <B, T> an application of <A, E>. The set of all applications of a given KH <A, E> is called the realization of <A, E>. Only a finite number of members of a realization can be thought of as meaningful in the sense that the statements of relationship involved can be thought of as meaningful. The members of the realization of a given KH are pairwise CRKS isomorphic: we say that each is an analogue of all the rest. In this way we formalise (structural) analogy. A CRKS isomorphism can be seen as a translation between the two relevant CRKS's.

Given two CRKS's $<B_1, T_1>$ and $<B_2, T_2>$ we can have a partial translation between them, i.e. between two sub-CRKS's. Suppose tht $<B_1, T_1>$ has been "taught" and that $<B_2, T_2>$ is to be taught/learned, and that <B, T> is a sub-CRKS of $<B_1, T_1>$ that is translated to a sub-CRKS of $<B_2, T_2>$. One can now teach $<B_2, T_2>$ by analogical modelling: we make an accommodation [VGS 2004] of <B, T> in $<B_1, T_1>$ and try to extend the translation of <B, T> into $<B_2, T_2>$ to translate this accommodation to $<B_2, T_2>$. If successful, we have extended the translation, i.e. extended the usefulness of $<B_1, T_1>$ as an analogy of $<B_2, T_2>$. If not, try another accommodation. In this way we will be able to gauge the effectiveness of the analogical modelling, and to what useful limit it can be "pushed".

To close this section we present a study material CRKS for the counting numbers with addition. It is not complete, but serves as an illustration. A CRKS can be presented in both diagram and table form: the diagram arises by plotting a point (vertex) for each concept-name involved and an arrow (directed edge) from the first entry vertex to the last entry vertex, labelled with the middle entries, in order of occurrence, for each tuple. The table is just a listing of all the relevant tuples. The following example shows how symbolic representation and precise description can help to clarify intuitive notions. It yields, by virtue of the representation technique (a CRKS), an ordering of the data (vertices), the information (edges/tuples/statements of relationship), and the knowledge (the collection of all paths, i.e. the structure of the information). To begin we present some experimental background and some terminology to set the scene.

Given lots of examples of "distinct objects together", we inductively abstract the invariant "togetherness". We name it "collection": each example is a collection of distinct objects. Given a collection of different squiggles on a piece of paper, we inductively abstract the invariant "singleness" of each squiggle. We name this invariant "written symbol": each example is a (written) symbol. "Collection" is, here, a primary concept-name, and "symbol" is, here, a derived or secondary concept-name.

Now consider a collection of collections of objects, all these objects being distinct. Arbitrarily choose any one collection. We call it the standard collection for this collection of collections. Now we "measure" the other collections by comparing each of them with our chosen standard collection as follows. Choose any other collection. Take any object from the other collection and place it next to any object from the standard collection. Pick another object from that other collection and place it next to another object in the standard collection. Continuing this process, we will run out of objects in one or both of our two collections. If we find that we are left with unplaced objects in one of the collections, then we say that that collection is the "bigger" of the two. If, at the end of the experiment, we find that there are no objects left in either collection, then we say that the two collections are "matched", or, more appropriate to the procedure, that they are in "one-to-one correspondence" (1-1 correspondence).

Next suppose that we continue to measure each of the collections against our standard collection, and find that each collection is matched with our standard collection, i.e. is in 1-1 correspondence with it. Then we can easily show by experiment that matching is transitive, i.e. that if two different collections both match the standard collection then they match each other. We conclude that, in such a case, the collections are "pairwise matched", i.e. if we choose any two of these collections then, since both match the standard collection, they will match each other. All these collections then, being pairwise matched, have a common property; an invariant. We call it the "number" of objects in each and every one of these pairwise matched collections. Before we move on, let us play a bit more with 1-1 correspondence. Suppose we have a collection L on our left, a collection M in front of us, and a collection R on our right; and that L and M are in are in 1-1 correspondence and M and R are in 1-1 correspondence. Then if we go to M from L by 1-1 correspondence, and then from M to R by 1-1 correspondence, we have object matchings of the kind object l in L matched to object *m* in M and object *m* in M matched to object *r* in R. It is easy to verify then that object *l* is matched to object *r* via *m*, i.e. there is a 1-1 correspondence between L and R. Thus a 1-1 correspondence form L to M followed by a 1-1 correspondence from M to R makes up a 1-1 correspondence been L and R. We say that 1-1 correspondence is transitive: this property is called transitivity.

We can introduce a special 1-1 correspondence on any collection. We let each member of the collection be matched with itself. Such a 1-1 correspondence is called the identity correspondence on that collection, and for each collection there can only be one identity correspondence on it.

Finally, consider any collection and any 1-1 correspondence of that collection with itself - a reshuffling of the objects in the collection. Apply this 1-1 correspondence and follow it with the identity correspondence on the collection. The result is just the same as our 1-1 correspondence by itself, i.e. C followed by I is C where C is the 1-1 correspondence on the collection and I is the identity correspondence on the collection. Similarly, I followed by C is C! This is called the identity property of 1-1 correspondence on a collection.

How do we represent, symbolically, by written symbols, this data and information in a precise and succinct manner? To do so is to move into the realm of mathematics, and our first entry into that realm in this context is with the introduction of the term “set”, by definition.

By a set we mean a collection of distinct symbols. Before expanding on this beginning in our sample CRKS, notice that “symbol” can also mean an object, sound, gesture and so on. Further, a symbol can be used to name anything at all; for example a collection or something that is in a collection.

In our CRKS statements the concept-names involved are marked by bold script. From these we get the tuple for each statement. The main point in the design of a CRKS is to ensure derivation at each phase of design.

An example: sets and counting numbers.

1. A ***set*** is a collection of distinct, different ***symbols***.
2. Each ***symbol*** in a ***set*** is, for that ***set***, a ***member***.

3. A ***set*** can be described by writing down a list of its ***members***.
4. In describing a ***set***, the order in which we list its **members** does not matter.
5. No ***symbol***, for any ***set,*** in the list of its ***members***, occurs more than once.
6. If the ***set*** denoted by ***symbol*** A and the ***set*** denoted by ***symbol*** B have precisely the same list of ***members***, i.e. if ***symbol*** a is a ***member*** of A then ***symbol*** a is a ***member*** of B and if ***symbol*** a is a ***member*** of B then ***symbol*** a is a ***member*** of A, then we say that ***symbol*** A and **symbol** B denote the same ***set***, which we express by saying that ***sets*** A and B are ***equal.***
7. Two ***equal sets*** have precisely the same list of ***members***.
8. Given a ***set*** denoted by a ***symbol*** A, we write “$a \in A$” to indicate that ***symbol*** a is a ***member*** of A.

So far we have the following table of tuples of concept-names, and an obvious a look-up table for statements. Notice that the sometimes convoluted, but precise, wording is designed to achieve derivation of concept-names.

1. <set, symbol>
2. <symbol, set, set, member>
3. <set, member>
4. <set, member>
5. <symbol, set, member>
6. <set, symbol, set, symbol, member, symbol, member, symbol, member, symbol, member, symbol, member, symbol, symbol, set, set, equal>
7. <equal, set, member,>
8. <set, symbol, symbol, member>.

A diagram of this CRKS is as follows.

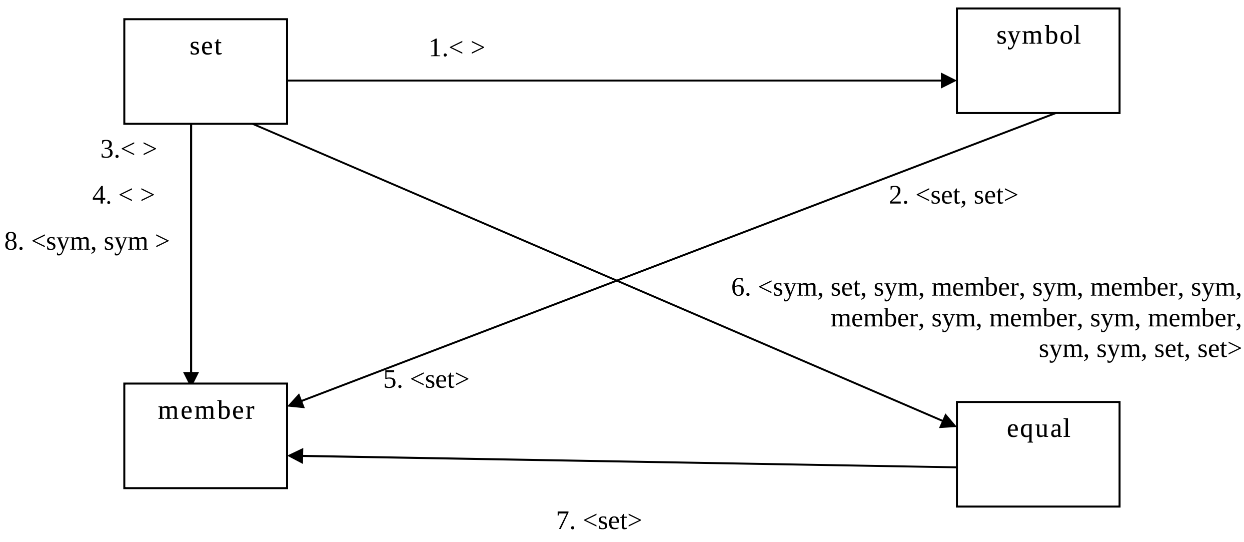


Figure 3: Diagram of a CRKS

Check manually that every path is a derivation path.

A statement of relationship like 6 is quite long and complex. We could break it up into several statements, but two comments are relevant. First, to formulate such statements is excellent training in learner communication skill. Second, it requires concentration on precision, and leads to the precise formulation of questions. In essence, teaching requires the exposure of the student to learning experiences, and guidance in helping the learner to formulate “discoveries”. Recall that every relationship can be stated in a potentially countably infinite number of different ways.

Continuing, we have

9. If the ***set*** denoted by ***symbol*** A and the ***set*** denoted by ***symbol*** B have the same list of ***members*** then we write “A **=** B” to indicate that the two ***sets*** are ***equal***.
10. If the ***set*** denoted by ***symbo***l A is **equa**l to the **set** denoted by ***symbol*** B, i.e. A **=** B, then we are saying that ***set*** A and ***set*** B are identical, i.e. that A and B are different ***symbols*** for the same thing, which is the general meaning of **equa**l.
11. If the ***set*** denoted by ***symbol*** A is identical with the ***set*** denoted by ***symbol*** B then we express this by saying A **=** B. (Notice that we use “equal” and its representation “**=** ” interchangeably after 9.)

The tuples of concept-names, to be tabulated, are respectively

9. <set, symbol, set, symbol, member, set, equal>

10. <set, symbol, equal, set, symbol, **=**, set, set, symbol, equal>

11. <set, symbol, set, symbol, **=** >

With a diagram

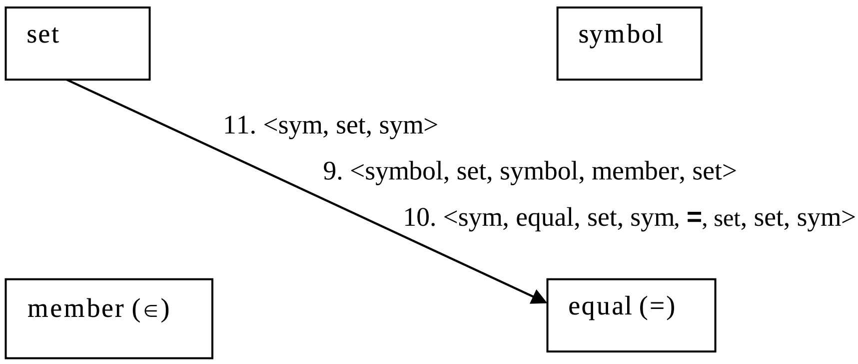


Figure 4: Diagram

Confirm that superposing this diagram upon the first still yields CRKS form. For example, we can derive ***symbo***l and ***member*** by 1 and 3, ***set*** being primary, and then **equal** by 6, so we then use **equal** to derive ***equal*** by 10! At first sight it may seem circular to derive a concept-name in terms of itself, but this illusion is dispelled by the notions of the changing meaning of a concept-name as its context grows with a CRKS – see later in this section and [GVS1999, 45f, 47], [VGS2004, 13, 66] – of first derivation path and path tree – see [GVS 1999, 51, 80f], [VGS2004, 89, 90f] - and of spiralling – see [GVS1999, 87], [VGS 2004, 89]. The initial meaning of "equal" in the context of this CRKS is introduced by statement 6. Statement 10 extends this meaning. The full meaning of the concept-name "equal" in the context of this CRKS is captured in the subset of all those of its statements that mention "equal" at least once.

12. A ***set*** with no ***members,*** usually denoted by the ***symbo***l ∅, is called an ***empty set***.

13. If ***symbol*** A denotes an ***empty set*** and ***symbol*** B denotes an ***empty set***, then, since ***sets*** A and B have the same list of ***members***, we have A **=** B, i.e. the ***empty set*** is unique.

This is our first theorem!

Notice that in trying to prove a theorem, the first step should be to expend a reasonable effort to find a counter example. This is a way in which the equivalent of falsification of a theory appears in mathematics, but with a different interpretation and in different circumstances.

Thus

12. <set, member, symbol, empty set>
13. <symbol, empty set, symbol, empty set, set, member, equal, empty set>

With a diagram

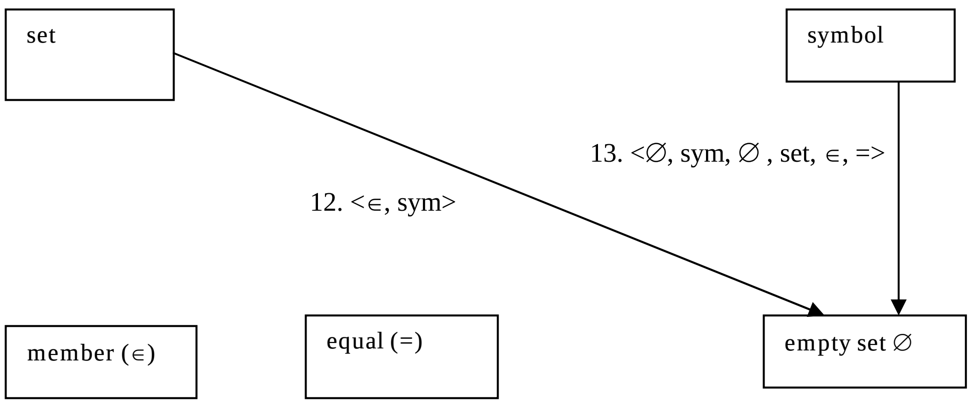


Figure 5: Diagram

Confirm that if this diagram is superimposed upon the first two diagrams we get a CRKS as a result.

We could add more statements of relationship to further cement the relationships among these 5 concept-names, but instead we extend the CRKS in another direction.

14. A collection of ***symbols*** written down one after another in a fixed order, without commas between the ***symbols***, is called a ***string***
15. In a ***set***, no ***symbol*** may occur more than once, but a given ***symbol*** can occur any number of times, and anywhere, in a ***string***
16. By a ***set*** that is representative of a given ***string*** we mean a ***set*** made up of precisely one **membe**r for every occurrence of any ***symbol*** in that ***string***.

14. <symbol, symbol, string>
15. <set, symbol, symbol, string>
16. <set, string, set, member, symbol, string>.

Draw a diagram of 14, 15, 16, superimpose it upon the previous combination of diagrams, and confirm that the result is a CRKS.

17. Given a ***set*** denoted by ***symbol*** A and a ***set*** denoted by ***symbo***l B, if we can match every ***member*** of ***set*** A with precisely one ***member*** of ***set*** B, and there are no ***members*** of ***set*** B left, then we say that ***sets*** A and B are in ***one-to-one correspondence*** (written "1 – 1 correspondence").
18. Suppose we have any collection of at least two ***sets,*** and that every one of these ***sets*** is in ***1 -1 correspondence*** with every ***set (***including itself) in this collection, then all these ***sets*** must have a common property which we will call the ***sets member number***.

17. <set, symbol, set, symbol, member, set, member, set, member, set, set, 1-1 correspondence>
18. <set, set, 1-1 correspondence, set, set, set, member, number>

Again, draw a diagram, superimpose, and confirm CRKS form.

19. The ***set*** of natural, or counting, ***numbers*** is defined in terms of ***sets*** as follows: 0 = ∅, 1 = {∅} = {0}, the ***set*** with 0 as its only ***member,*** 2 = {∅, {∅}} = {0, 1}, the ***set*** with 0 and 1 as its only ***members***, and so on, so that, in general, we have, if we have defined m then the next one is {0, 1, 2 ..., m}, so n = {0, 1, 2, ..., $\bar{n}$} for the ***number*** n where $\bar{n}$ is the ***number*** defined just previously in this ***set*** of ***numbers***, so ***number*** n is the ***set*** of all its predecessors in the ***set*** of all counting ***numbers***. (Note that 0 = ∅ has no predecessors.)

19. <set, number, set, set, member, set, member, number, number, set, number, number, set, set, number>.
20. To "count" a ***set*** means to find the unique counting ***number*** that is in ***1-1 correspondence*** with that ***sets members***.
21. To add any two counting ***numbers*** m and n we write out the ***string*** of ***set*** m and write out the ***string*** of ***set*** n and make a ***string*** of all the ***member symbols*** written down. Next we

construct the ***set*** representative of that ***string***, and "count" the ***set***. The result is a counting ***number*** p that is in ***1-1 correspondence*** with that representative ***set,*** and we write the result as p **=** m **+** n.

Example: 2 + 3 $\longrightarrow$ {0, 1} + {0, 1, 2} $\longrightarrow$ 01012 which is 1-1 correspondence with {0, 1, 2, 3, 4} = 5. Thus 2 + 3 = 5.

20. <set, number, 1-1 correspondence, set, member>
21. <number, string, set, string, set, string, member, symbol, set, string, set, number, 1-1 correspondence, set, equal, +>

22. It is easy to show that, for any counting ***numbers*** a, b and c we have

$$((a + b) + c) = (a + (b + c))$$

$$(a + b) = (b + a)$$

$$(a + 0) = a$$

as the properties of **+.** [In each case the representative set of the string is the same (identical) both sides of the **=**].

23. Using the meanings of the terms representative ***set*** and the counting ***numbers*** we can show that for any counting ***numbers*** a, b and c we have that

$$\text{if } a + b = a + c \text{ then } b = c,$$

called the ***cancellation law***. Our second theorem. (Challenge: prove the cancellation law)

24. The counting ***numbers*** together constitute a ***set***, written {0, 1, 2, ...}, that is never ending: it is said to be ***countably infinite***.

25. Multiplication of two counting ***numbers*** m and n means adding n to itself m times, and is written m ***x*** n.

22. <number, +, +, =, +, +, +, =, +, +, =, + >
23. <set, number, number, +, =, +, =, cancellation law>
24. <number, set, countably infinite>
25. <number, ***x*** >.

Confirm that, in the end, we have a CRKS. This is the end of our example.

A primary vertex is a vertex with no incoming arrows. A goal vertex is a vertex with no outgoing arrows.

What is the "meaning" of the concept with concept-name a, for example "member", in a CRKS K? By the context schema of vertex a in a CRKS K we mean the set of all tuples that have a as at least one entry in them, i.e. all those statements of relationship, in K, that involve vertex a together with all the vertices that are entries in at least one of those tuples. That is all we know about a in the context of K! Deleting a from K entails deleting all the tuples in the context schema of a from K. Try it for "equal" and "member" and "set" in the CRKS.

It is easy to show that, given any non-primary, non-goal vertex p in a CRKS, there is at least one derivation path for p that extends to a derivation path for some goal of that CRKS. This hinges upon noticing that every arrow in a CRKS is such that every vertex in its tuple is a derived vertex, so it is a "derivation arrow".

Every concept-name in a CRKS is associated with primaries and/or with other concept-names that can be previously met in that CRKS, for example by running a limited access cascade from its primaries.

We have, in CRKS's, a way of organizing sections of study material and of joining these sections together. Derivation means that every "new" concept-name met in a CRKS is related to primary and/or other concept-names that can be previously met in that CRKS. We are forced to have repetition – the various statements of each relationship, chosen (easily) to enforce derivation, give us different "views" of that relationship. We have various ways of presenting/teaching/learning the study material by following collections of derivation paths from primaries in a CRKS. We can gauge the "complexity" of each concept-name in various ways, for example by examining its concept schema in the given CRKS. We have ways of gauging the complexity of each path – see the appendix of [VS 2008] for some modifications of [VGS 2004]. We can build a CRKS from "clusters", where a cluster for a tuple is any minimal CRKS that contains that tuple. And we can use CRKS's, in general referent free form, to define each of five modes of reasoning: associative, constructive, deductive (vertical), intuitive, and inductive (analogical/lateral). In general, CRKS's also have applications in any field in which we can write down relationship statements involving two or more “things”.

A parallel approach to basic numeracy was mentioned in passing before: that of “pacing”. This depends upon the key concepts of direction and measurement. Start in a reasonably large open region. Mark a spot, and observe some obvious object some way off in that chosen region. Introduce the idea of “in the direction of that object” from our marked spot. Choose a child an let s(he) take a pace from this “home spot” in the direction of the object. Mark the place where the child now stands with a marker labelled “pace”. Re-mark the “home spot” with label “no pace”. Now let the child take another pace in the direction of the object. Mark where the child is now with the label “pace, pace”. We see that, in this direction, “no pace” comes before “pace”, which comes before “pace, pace”. Now to be very brief, one can see how the notions of distance, measurement, and standard pace can arise, as well as the introduction of the nouns, 0, 1, 2, ... We get the natural number set and countable infinity via “measurement”. (Bearing in mind smallest possible distance there is no actual necessity to introduce fractions: 2.75 kg is 2750 grams!)

To extend this approach in a natural way, turn the child, at “no pace”, to the opposite direction. Repeat the pacing, but now we progress from “no pace” to “opposite pace”, “opposite pace, pace” and so on, where “opposite” is to become “minus” to indicate the direction in which the child must face before pacing. Then we reach the point where +3 +2 means go three paces, from where you are, towards the object, the plus, then another two paces toward the object, the plus, and we get to the pace, pace, pace, pace, pace marker, so one gets to discover that +3 +2 = 5. For +3 +(-5) we go three paces, from where one is, in the plus direction, see the + which means carry on pacing, then see the – which means face about, and then take five paces from where one was after the +3. We end up at the “opposite pace, pace” marker, and we have discovered that +3 + (-5) = -2. Of course one should start these experiments from the “no pace” marker in order to establish the operation + and operator -, and then generalize to a start from any position. For “distance” one would introduce the notion of any minimal value measurement in standard paces.

One can see how this approach meshes with the one indicated in our example CRKS. We have “parallel approaches”.

Vulnerability of relation nets and hypernets is dealt with in [GVS1999, Part II, 195 and 215f] and in [VGS2004, section 2.2, 44-47]. Presentation strategies are introduced in [GVS1999, 67f] and in

[VGS2004, in various places]. Menger's theorem for these structures is covered in [GVS1999, 69f], [VGS2004, 78f] and [VS2008]. Some gauges of complexity are defined in [VGS2004, 92f]. Accommodation and assimilation are discussed in [GVS1999, chapters 6 and 10] and in [VGS2004, 95].

Of course organization of study material is instinctive for every good teacher: here we merely present one possible formalism for organizing the material in learnable and teachable form – see [GVS1999, Part I].

We claim that graphs, hypergraphs and networks, together with those applications, in various fields, of which we are aware, are, in a sense, special cases of relation nets and hypernets. A relation net is, after all, essentially just an edge labelled multigraph, but has the strong vulnerability property which sets it apart from graphs – see [GVS1999], [VGS2004].

6. The Knowledge Hypernet of a Formal Language

Given a formal language L, we construct a knowledge hypernet <V, E> for its theorems as follows. The vertices of <V, E> represent the words of L, where the primary vertices represent the axioms of L. The edges of <V, E> represent the sentences of L as follows.

A primary sentence is represented by a derivation edge that starts with an axiom in the first space of the relevant inference rule, ends with a theorem, and has as label the set of axiom words, the primaries, involved (i.e. their vertices) and that theorem.

A general sentence generates a derivation edge in a similar manner. The axiom vertices (the invariant, observable initial data) and (invariant) primary sentences are our “model”, the sentences being the initial information. The pattern of relationships among the derivation edges, as displayed by the derivation paths, is our knowledge. The KH is our theory.

Mathematics defines and describes symbolic structures. For each structure/pattern we define a language to describe that structure: i.e. to write down the words, the axioms, and the rules of inference. The rules are always the same and are rigid, i.e. the grammar is fixed for all those languages. Thus they are “forma**l**” languages. Now, in each, the words “mean something”, they are each assigned a “value”. Further, each sentence “says something”, but always what is “meant” and what is “said” are about symbols, not about observations or objects. How do we tie all this up? The answer lies with the KH of L.

A hypernet is realized in terms of relation nets. A KH is abstracted from a concept-relationship knowledge structure (CRKS) [VGS2004], which has concept-names as its vertices, has primaries, has goals, statements of relationship generating its derivation edges, and derivation paths. A CRKS can potentially represent any body of knowledge that is describable in terms of nouns and statements of relationship among them, in a way ordered by derivation paths, be it a scientific theory in which the primary vertices and primary edges are the model and the derivation paths the reasoning, or be it a piece of study material for example. See [GVS1999]. Notice that if we have a countably infinite set of vertices then we must think in terms of a countably infinite set of primaries, of goals, of non-primary non-goal vertices, and of edges, but of every derivation path being of finite “length” i.e. each uses a finite number of edges. Indeed, one could define a formal language as such a KH.

Given a collection of CRKS's with possibly different concept-names, but all of the same number of concept-names, we may be able to inductively abstract from each the same referent free KH, if all those CRKS's have the same relational structure. An edge of that KH is a member of a relation on its vertices. We thus have the following situation. That collection of CRKS's is a realization of that KH: the KH is an inductive abstraction from every one of them. Any two of those CRKS's are "isomorphic via that KH": CRKS isomorphism, called a (complete) translation. Notice that each relationship has a potentially countably infinite number of different statements.

Think of the concept-names as "words" and the statements of relationship as "sentences" and one can see that, given the KH of the relevant formal language L, we can regard each mathematical structure described by L as a CRKS in a realization of that KH.

It is now possible to distinguish the same method in science, teaching and learning, and mathematics. Everything is really just description of structures in an appropriate language, and this method can be formalized in the same way in each case.

Derivation in a KH is a property inherited, by definition, from CRKS's [VGS 2004]. This is effectively as follows; every primary is, trivially, a derived vertex, and every other vertex is a derived vertex by virtue of a path from a primary to that vertex on which every vertex involved in each edge on that path is a derived vertex. That vertex is then, too, a derived vertex by virtue of that path, a derivation path. It is easy to show [VGS 2004] that every edge in a CRKS is a derivation edge, and every path is a derivation path (because, by definition, every vertex in a CRKS is a derived vertex). This is of course a now familiar situation! It occurs in science, teaching and learning, and in mathematics.

Note in passing that this "definition" is recursive and that every recursively defined set has an associated "principle of induction".

A formal language is rigidly defined and has effectively a countably infinite vocabulary. It is precise, but its descriptive ability is limited by its rigid definition. In every interpretation in which there is a reflexive and substitutive relation we can partition W into equivalence classes: in each class every word "means the same thing". This gives us flexibility of description, both with words and relationships. However, while a natural language may contain a "partial interpretation" of a formal language, with a finite but extendable set of words as the natural language grows, it is very descriptive, and can become more so by adopting words. This is why we can use a natural language, as we are doing here, as a meta-language in which to describe formal languages and write out arguments in place of proofs. It is why a (good) natural language can be used as a meta-language, a teaching language, in teaching anything. Note that natural language communication skills are thus vital, and that they include such things as pictures, speech, demonstrations and so on. Another way to "enhance" a natural language is by introducing definitions. A definition in a natural language introduces a concept-name that denotes one only precise paragraph of one or more sentences. It is used in a specified context, and enables shorter descriptions.

From the KH for a formal language L we move to a realization in terms of CRKS's for each formal language, usually by only partial interpretations. Given one of these CRKS's, we can assign "real world" concept-names to its vertices and "real world" statements of relationship result. We start with the primaries and primary sentences of the CRKS's, our model, and then predict that that entire CRKS is isomorphic with, is completely translatable to, that "real world" CRKS – our theory. The predictions of that theory are then empirically tested. We have used a formal language as a "mathematical modelling" technique. Such an approach can be seen to be operating in science and in teaching/learning, where, in the latter case we compare the teachers study material CRKS with that of the learner by examinations, for example, as empirical test of the predictions, by the teacher, of what the student has assimilated/learned (in terms of structure, pattern; not by rote!). See [GVS1999, Part I)

Recall that the notion of derivation is defined by regarding each application of a rule of inference as a statement of relationship among words; a sentence. The primaries are derived words, every other word is a derived word, a derived word is call a theorem, and any derivation path to a word is called a proof of that word

The sub-hypernets of the KH, K, of L constitute a Boolean Algebra with operations $\cup$ and $\cap$ [VGS 2004] and the obvious complement. Each of these has a realization; a set of relation net interpretations. Those that are sub-KH's of K produce a set of CRKS's. By a partial interpretation of K we mean the restriction of an interpretation of K to a sub-hypernet of it. The significant ones are those for which the domain is a sub-KH of K, but the range will not usually be a "meaningful" CRKS. The meaningful partial interpretations are those for which the domain is a sub-KH that contains [VGS 2004] a sub-KH of the primary KH of K, the (scientific) model. We interpret, then, only a subset of the axioms/primaries and a subset of the primary sentences, in such a way as to constitute the primary CRKS, the (scientific) model, of the range CRKS. We may then speak of the range as a partial (mathematical) model of K; of the formal language represented by K. There will be a countable infinity of partial interpretations in each case, but they differ only in notation.

To briefly indicate how this view applies, consider the example of proposition logic.

The word building CRKS: The primaries are the proposition symbols. The (sufficient) word building rules are $\neg$ and $\vee$. For every proposition symbol P we generate ($\neg$ P), which yields, in each case, a binary relationship statement “by starting with P and applying $\neg$ to P we get ($\neg$ P)”, displayed by a binary edge

$$P \xrightarrow{\quad i;\varnothing \quad} (\neg P)$$

For every two proposition symbols P and Q we generate (P$\vee$Q), which yields, in each case, a statement of relationship “by applying $\vee$ to P and Q we get (P$\vee$Q)”, displayed by an edge

$$P \xrightarrow{\quad i;\, \langle P, Q\rangle \quad} (P \vee Q)$$

where P and Q are written “in detail”, i.e. in terms of proposition symbols. In this way we build all the words, displayed as vertices, if we use only legitimate applications of $\neg$ and $\vee$. Notice that we first build the primary CRKS, and then extend it. In the extension, P and Q above now become variables for words already “found” by a limited access cascade from the primaries, which yields a partial order in the generation of the set of words. See [VGS 2004] on presentation strategies. The CRKS is countably infinite, but that difficulty disappears when we consider that in “real world” applications the relevant interpretations are of finite sub-CRKS’s only, i.e. they are partial interpretations.

The theorem building CRKS: Taking one set of axiom word forms to be

$$((u \vee u) \Rightarrow u)$$

$$(u \Rightarrow (v \vee u))$$

$$((v \vee u) \Rightarrow (u \vee v))$$

$$((u \Rightarrow w) \Rightarrow ((v \vee u) \Rightarrow (v \vee w))),$$

the primaries are every word, of one of these forms, in which u, v and w are proposition symbols. These are the primary theorems. Choosing an inappropriate collection of (invariant) axioms will result in a contradictory theory (structures) in which both P and ($\neg$ P) can be proved as theorems – another facet of “falsification” in mathematics!

Applying the rules of inference produces sequences of words p_o, p_1 ..., p_{n-1}, p_n. If all of these but p_n are axioms, i.e. primaries, then p_n is a theorem (which is not a primary) and we have a statement of relationship “applying the following rule of inference ... to p_o, p_1, ..., p_{n-1} we derive the theorem p_n”. This is displayed as the edge

$$p_o \xrightarrow{\quad i;\ <p_1, p_2, \ldots, p_{n-1}> \quad} p_n$$

and in this way we generate the primary CRKS, the first step in a limited access cascade from all the primaries. Extension of the generation is obvious, getting an edge, for each proof of a theorem p_n, in which all the other p_i are axiom or theorem words already found by a limited access cascade from all the primaries. The corresponding argument is the statement of relationship.

Again “real world” applications arise from partial interpretations, of sub-CRKS’s. The non-primary vertex theorems in the interpretation provide our predictions, and must be empirically tested. The primary sub-CRKS is our model, and the sub-CRKS our theory, in each particular partial interpretation.

A rule of inference is a tautological implication, a theorem of proposition logic, which enables us to infer a word, a theorem, from a sequence of axioms and theorems. In the theorem building CRKS, every statement of relationship, i.e. every edge, is a valid argument.

In any CRKS, every tuple $<a_0, a_1, \ldots, a_{n-1}, a_n>$ arises from a statement of relationship. Every statement of relationship is a specific instance of one only CRKS rule of inference that has the general form “a combination of the occurrences a_0, a_1, ..., a_{n-1} of concept-named invariants, in that order, yields the invariant with concept-name a_n”. The precise form of the “combination”, together with a_0, a_1, ..., a_{n-1}, determines the individual statements of relationship. For example, in the case of $\langle P, (\neg P)\rangle$ above the “combination” is determined by $\neg$ while in the case of $\langle P, P, Q, P \vee Q\rangle$ the “combination” is determined by $\vee$. For the theorem building CRKS, each “combination” is determined by a rule of inference of proposition logic. Here the concept-names are words.

Consider the word form

$$((P_1 \wedge P_2 \wedge \ldots \wedge P_n) \Rightarrow C)$$

where the P_i are premises and C is the conclusion of an argument when specific words are assigned to each variable. The argument is valid iff all the premises are true and the implication is true, so that the conclusion is true. If this form is tautological, i.e. a theorem of proposition logic, then it can be regarded as a rule of inference: if each of the words assigned to the premises is (an axiom or) a theorem then, since the implication is true, C is a theorem. Thus a valid argument and a formal proof are equivalent for specific assignments of words to the variables in those cases in which all the premiss words are true, are theorems, and when the implication is also true (which is why the form must be tautological). Valid argument, formal proof, and derivation path are equivalent notions.

The case of Modus Ponens, i.e. the tautology $((P \wedge (P \Rightarrow Q)) \Rightarrow Q)$, is straight forward: assigning specific words to the variables we will arrive at a valid argument, and formal proof, if the assignments for P and $(P \Rightarrow Q)$ are both true. The case of Contradiction, the tautology

$$(((P \Rightarrow (Q \wedge (\neg Q)))) \Rightarrow (\neg P)),$$

is more convoluted. If the word assigned to premiss P is true, and we then prove that the word assigned to the premiss $(P \Rightarrow (Q \wedge (\neg Q)))$ is true, then since the main implication is true for all assignments, the word assigned to $(\neg P)$ is true. Since we cannot have the assignment for P such that it is both true and false – $(\neg (P \wedge (\neg P)))$ is a tautology – our assumption that the word assigned to P is a theorem was incorrect. It is really a simple case of valid argument, and, equivalently, formal proof, of the word assigned to $(\neg P)$. Another view of the situation is to note that $(Q \wedge (\neg Q))$ is universally false, so $(P \Rightarrow (Q \wedge (\neg Q)))$ can only be true if P is false.

7. Modes of Reasoning

It is true that inferential reasoning is embedded in the method of mathematics. It is true that we have used the term “reasoning” without specifying what it means. In this section we try to clear this up by briefly discussing five modes of reasoning. We do it in terms of CRKS’s.

7.1 The Association Mode

Basically we have a collection of CRKS’s, plot a vertex for each CRKS, and join a pair of these vertices if the intersection of the two relevant vertex sets is non-empty. Label that edge with the appropriate intersection, and we have a hypernet. This hypernet is called the association graph for that collection.

More particularly, suppose that we have observed, and stated, a relationship among two or more observations. The statement produces an n-tuple of the occurrences of concept-names in that statement, generating a directed edge, and a vertex for each concept-name involved in the statement; a relation net. By a cluster for that n-tuple we mean any minimal CRKS that contains that n-tuple, where by minimal we mean that if any one vertex or any one edge is deleted from the cluster we no longer have a CRKS that contains that tuple. A cluster, which is of course not generally unique, for a statement of relationship tells us a more complete “story” of the relationship than just the statement. Its “cluster set” is the set of its vertices, which is precisely the set of those vertices that occur (at least once) in the n-tuple that generates it. Clusters tie up easily with observations of relationships as stated, and plotting the association graph for a collection of clusters produces a “finer detail” graph. Following edges in an association graph constitutes the association mode of reasoning.

7.2 The Construction Mode

The association graph can be used, in part, to construct a CRKS in most cases. If its vertices represent lessons for example, then we get a “course CRKS”, if courses then we get a “curriculum CRKS” – see [VGS 2004]. How can we construct a CRKS from (parts of) an “appropriate” association graph?

A cluster is more descriptive of a statement of relationship than the bare statement, so we consider the "bottom line" case of a clusters association graph. Bear in mind that, in general, every cluster is idiosyncratic.

- Look for primary concept-names in association with statement tuples which are such that all but the last entry are primaries. These form our "model"; our beginning, or primary CRKS. There must be at least one "primary tuple", thus producing at least one "theorem concept-name/word".
- Look at all clusters associated with the clusters of these primary tuples. By reversing edges if necessary, by re-wording statements of relationship, choose associated clusters that can be joined [VGS 2004] to our primary CRKS to form a CRKS, i.e. accommodations which can be assimilated [VGS 2004].
- Continuing by appropriately repeating step (b), we construct an idiosyncratic CRKS from the (cluster) association graph.

This process is called the construction mode of reasoning. First we find associations, and then we build an idiosyncratic knowledge structure using some, or all, of those associations. (Note that the CRKS's involved need not of course all be clusters.)

7.3 The Intuition Mode

Given a CRKS, we run a fast access cascade [VGS 2004] from its primaries. In a fast access cascade we access every concept-name that occurs in a tuple as we "run" that tuple. This can produce vertices that are "found" with no edge incident to or from them at that stage. We "know something about" each such vertex, but only inasmuch it is involved, as an "unknown", or hypothesis, in at least one statement accessed in previous steps in the cascade. Such vertices are said to represent intuited concept-names at that stage. (They must eventually be "connected in" by "finding" derivation paths to or through them, or we do not have a CRKS!).

The process of running a fast access cascade from the primaries of a CRKS is called the intuition mode of reasoning.

7.4 The Deduction Mode

Given a CRKS, we run a limited access cascade [VGS 2004] from its primaries. Here we can only "run" along an edge if every concept-name, but possibly the last, in that tuple is primary or has already been “found” by the cascade: it is a derived vertex. Every concept-name met is either primary or is related to (primary and) previously “found” concept-names – the formalization of Ausubel’s main criterion for learning and teaching, see for example [GVS1999] and its references. It is easy to show that a relation net with primaries is a CRKS iff it can be generated by running a limited access cascade from those primaries [VGS 2004].

The process of running a limited access cascade from the primaries of a CRKS is called the deduction, or inferential, mode of reasoning. We must prefer “deduction” here, because all five modes of reasoning are inferential.

7.5 The Induction Mode

Consider a collection of otherwise different red objects. From it we inductively abstract the invariant “red”, a primary (or primitive) concept-name because it is abstracted from a collection of examples or “instances” – a familiar situation, see for example “observable”. From each object we can imagine a “mapping” assigning the object to the one invariant “red”. Visualise a graph with a vertex for each object, a single vertex for “red”, and a single edge from each object vertex to the “red” vertex. Each edge can be regarded as an assignment in an abstraction mapping; as a “trivial isomorphism” with one member. Reverse the edges and each edge can now be regarded as an assignment in an interpretation mapping; as a “trivial isomorphism” with one member that “algorithmically constructs” the object involved. Now broaden this situation.

Consider the realization of the KH, K, of L. It consists of a countably infinite set of CRKS’s each of which is an interpretation of K. Looked at the other way, K is the invariant referent free structure inductively abstracted from this set of CRKS’s. In any subset of this set, any two CRKS’s are isomorphic – they have the same relational structure. Visualise our graph again. There is a vertex for each CRKS in the (usually finite due to partial interpretation of K in “real life” situations) subset of the realization of K, and a vertex for (the domain of interpretation in) K. Each edge displays an abstraction of K from the CRKS involved. Reversing direction, each edge displays an interpretation of K. The CRKS’s are pairwise isomorphic. Each abstraction edge actually displays an “abstraction isomorphism”, and reversed edges actually display “interpreting isomorphisms” each of which is “algorithmic” inasmuch as it constructs the relevant CRKS using K as the “plan”. The CRKS’s are pairwise analogues. A different KH, even though it may have the same model sub-KH as K, is a “new” plan, with its own realization.

These CRKS’s could “be” those of mathematical structures, or they could “be” scientific theories about “real world” structures. For a particular K they are analogical, and we compare structures for analogue, or generally partial analogue (when two different KH’s are involved) status using this isomorphism technique (or isomorphism with restricted domain). We can, for example, find that two sub-CRKS’s are isomorphic, analogical, and “predict” from one to the other by attempting to expand the domain of the isomorphism. (These predictions must, as usual, be “tested” in some way.) This helps to build, to teach, the range CRKS if one is familiar with the domain CRKS, i.e. if it has already been taught.

This use of isomorphism to compare structures for full or partial analogical status is the analogical mode of reasoning or, because it is founded upon inductive abstraction, the induction mode of reasoning. The “principle of induction” involves comparing two recursively defined structures.

To close this section we must make a pertinent point. Deductive reasoning is also sometimes termed “vertical reasoning”. Inductive reasoning is also sometimes called “lateral reasoning”. In learning we cannot confine ourselves to any one form of reasoning: we need to be proficient in all five. Males have a tendency to “think vertically”, while in females the tendency is to “think laterally”. This can cause misunderstanding in communication. All too often study material is structured, and presented, “vertically”, particularly in science and mathematics, (if it is structured at all – unstructured study material, out of its structural context, almost always forces rote learning with little motivation other than “it is so”). This results in silly claims such as “girls cannot master science and mathematics”, which is complete nonsense. Anyone can “study” anything at all if it is placed in context in a “taught”, or “discovered”, structure.

8. An Exercise in Speculative Reasoning

This will consist of some claims with some explanatory comment. The general claim is: In any turbulent, polluted fluid we get “accretions” of the pollutants. They get together in “lumps” of various sizes and compositions, each of which will eventually break up.

We have the pattern “simple to complex to simple”. This is an observable, an invariant, inductively abstracted from many individual cases. The process is unpredictable, by any analytic means, in all but the most simple cases. We cannot predict what will happen, or when a given accretion may happen: probability is meaningless, and the process is not “random”, i.e. unstructured. It is unpredictable! (This is the basis of the “best” current hypothesis for the origin of living organisms, and for natural selection.)

Consider now a collection of living organisms such as a swarm of bees, a shoal of fish, a nest of ants, a flock of birds, a herd (not a mob) of animals, even a collection of bacteria. The individuals will form a collection if natural selection leads to a collection that will survive longer than the individuals that belong to it. (They have an “instinct” to survive as individuals, and a “duty” to “behave” in such a way as to preserve the collection.) As another such collection, consider the cells in a brain.

Now in each of these collections, the “communication interactions” between individuals are simple, elementary, unstructured, “mindless”. For a bird in a flock, for example, it is simply "do what the birds next to you do”. For a brain cell it is “if you are on, switch on some of the cells that you are connected to”. The “accretions” of these interactions are unpredictable, but lead to relatively complex behaviour, partially predictable, of the collection as a whole. The collection will eventually decay into individuals again in the sense that there will be too few, or they will be “communication isolated”, or even that there are none left.

The simple individual communications accrete, unpredictably, to complex and (partially) predictable collection characteristics, which will eventually “decay”: simple to complex to simple.

The challenge is now to construct, for each collection mentioned, the statements made about it, and more if you wish, design at least one cluster for each statement, and plot the association graph. From it construct a CRKS. From these CRKS’s we then want to abstract, by induction, an invariant

KH so that we see which parts of each CRKS are pairwise analogical. In doing this we may need to change some of these CRKS's, these changes of course involve re-wording statements of relationship, but not changes of relationships.

We can then use "analogy" to teach one by reference to another. We can check, analyse and present the KH or any one of its interpretations using intuitive, inductive and deductive reasoning. Hopefully the KH will tell us at least "simple to complex to simple" as the invariant structure!

We have shown, in [VS2007] and [VS2008], how a simple speculative model of a network of neurons can be trained/taught/propagandised by control, or at least partial control, of the feedback through its environment. The interactions of the neurons are basically unpredictable, so one must permit unpredictable intermittent "drifting" invariances. Thus there is unpredictable noise together with unpredictable temporary invariances, so such a network can exhibit unpredictable behaviour. These intermittent invariances could be regarded as providing individuality of the network behaviour, and it will affect learning. It provides the fleeting invariances, the hypotheses, the new ideas, which can become more permanent if the appropriate feedback exits, failing which they will decay into the general noise.

9. Suggested Definition of a Plan

By the term “method” in this section we will mean the broad requirements for constructing a plan. What, then, is a plan?

A general plan, the format of the plan, can be visualized in the form of a KH, while a specific plan can be seen as one of its interpretations. It is more simple to explain a specific plan: to speak of a CRKS.

The title is “to achieve the following goals”: this specifies the problem in a broad form. We must collect the “evidence”. This takes the form of the primaries and the primary statements of relationship, i.e. those that produce a tuple in which every entry, but the last, is of a primary. From there we build the CRKS to the required goal, the title, using, in the statements of relationship, only derived concept-names (but possibly for the last one mentioned). Each statement can be regarded as a rule of inference; as a statement of the form “if these occurrences of concept-names are given, in this order, then this concept-name follows”. A limited access cascade will check the structure for CRKS form. The KH abstracted from this CRKS stipulates the form of the plan.

Now if we re-word statements so as to reverse the direction of every arrow then we view the CRKS “top-down”, from title, i.e. the goal, to the primaries, i.e. the required data. This view completely specifies the problem. If we leave the CRKS as is then we have the “bottom-up” view, i.e. from primaries, the input data, to goal, the “title”. This view completely specifies the algorithm for solving the problem.

Move from statements to members of relations and we abstract the KH from this specific CRKS. All the interpretations of this KH are specific plans with the same form. (Previously we have referred, briefly, to this approach as the “action diagram” approach. [VGS2004, GVS1999].)

To be more specific, we have the following proposal. A hypernet K is called a plan iff

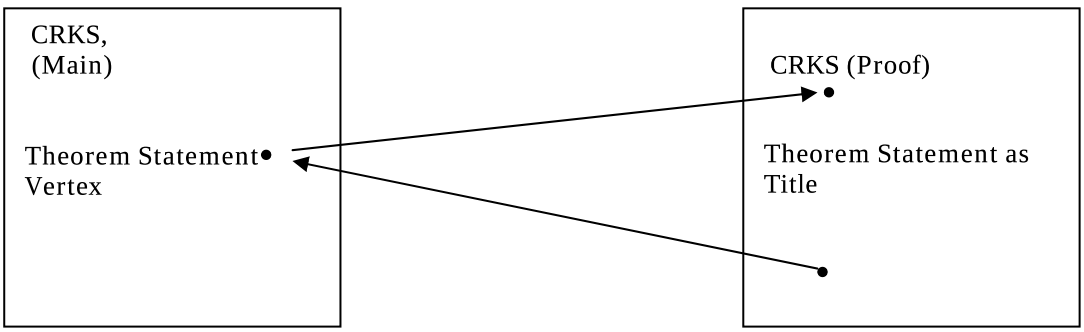
- K is a KH, <V, E>
- V and E are both finite
- K is minimally weakly connected, i.e. there is a derivation semi-path between every given pair of vertices of K, and deletion of any edge from K leaves K no longer weakly

connected. (Connected [VGS2004] implies weakly connected, but the converse is not generally true.)

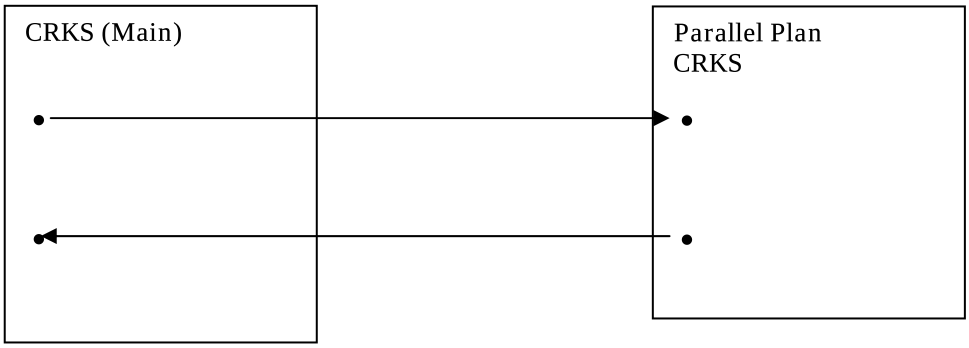
- K has a single root, when viewed top-down; the goal when viewed bottom-up.

Read top-to-bottom the root is the only primary of K, and K completely specifies the (potentially infinite) set of applications for which it is the inductively abstracted invariant structure. Read bottom-to-top the root of the plan K is the only goal, the primaries specify the input required to solve the problem, and K is then seen as an algorithm for solving problems – a computer algorithm if the problems are computable and therefore implementable on a digital computer.

In [GVS 1999] and [VGS 2004] we briefly discussed theorem proofs as "attached plans', i.e.plans associated with the "main" CRKS. There are two simple associations – more complex ones can occur -, both essentially parallel associations. For theorem proofs, we proposed setting the proof aside as indicated below



The next most simple form of association is that of a parallel plan as indicated below



where we may have more than one top-down goal on the right.

A parallel plan may, in the process of activation, need to draw bits of data, and even statements, from the main CRKS, and vice versa, a case of closer association of the two CRKS's. Of course one can have cases in which two CRKS's are "chained" in the sense that some, or all, of the goals of one are some, or all, of the primaries of the other.

By a specific plan we mean a "real world" application, an interpretation, of a plan. The vertices of the plan are used to denote "real world" concept-names/words, turning the general plan into a specific plan.

All specific plans must be proved to achieve the "desired result" in a finite number of steps, i.e. using a finite number of edges, where the "desired result" may be thought of as the goal of the plan<V, E> when read bottom-to-top. This can be done either by implementing the plan with every relevant input, or by presenting an abstract argument. (Suppose our input is "any pair of counting numbers $n, m \in \{0, 1, 2, \ldots\}$"!)

Now a KH plan can be regarded as recursively generated. How does this work? The KH is generated by running a limited access cascade from its primaries. This situation is one of recursive definition of the KH: we start with the primaries and some "generating relations". Each member of each such relation is an edge in the KH, so the KH can be regarded as being recursively defined from its primaries using the generating relations; the rules of inference or rules of derivation. Since an interpretation of our KH, in the realization of the KH, is a CRKS, "generating relation" becomes "relationship", and "member of relation" becomes "statement of relationship". Since a KH can be seen as recursively defined, there is an induction principle associated with it, as is automatic for any finite or countably infinite recursively defined structure.

Briefly here, if a relation net is a formal schema then all its primaries are at shortest path length zero from the primaries and are derived vertices. If now all vertices at shortest path length n from the primaries are assumed to be derived vertices, and all edges incident from those vertices can be shown to be derivation edges so that the end vertex of each is a derived vertex at shortest path length n + 1, then that relation net is a CRKS.

Given that a KH is an abstraction from its realization in terms of CRKS interpretations, called applications of it, we see it as the invariant structure inductively abstracted from the set of all its applications. Each CRKS in the realization of the KH is said to be represented by the KH, and that CRKS is arrived at by giving “values’, concept-names and statements of relationship, to vertices and edges of the KH which then denote the concept-name vertices and edge statements of relationship of that particular CRKS. Notice that we regard the vertices of K as referent free symbols while those denoted in a specific CRKS interpretation/application of it are specific word concept-names.

(Recall that a relationship can be stated in a potentially infinite number of ways!). We have now reached the level, then, of scientific method: of observables (primaries) and relationship statements involving them (primary edges), constituting a model, and a derived theory which is supposed to derive, predict, new observables, and relationships among them and previous ones, by generation via derivation. Now we have empirical test against which to gauge incompleteness, falsification, and domain of functionality.

Finally, it is possible to see this whole trend in science and mathematics in terms of sequence of noticeably similar specific plans: the same method applies! A very similar situation applies in the fundamental specific plans of teaching and learning: again we have the same method.

Teaching and learning requires planning. Plans are often incomplete, and always eventually false as we extend them. Modification and replacement of plans forces further learning: an ongoing process of never-ending learning. Teaching/learning starts with searching for invariants via the primary process of procedure: trial-and-error. It must find relationships among those invariants, so that it can construct a model. It must extend that model, for the given topic, by reasoning, to generate further invariants and invariant relationships among them and previously generated ones. It must test these predictions by problem solving using specific plans with the same method. It must test the generated structure in “real life”, for example in examinations, and these are empirical tests. Does this sound familiar? We can view it in terms of a KH and its CRKS interpretations/applications in each topic. The plans for teaching/learning are those of mathematics and of science: all three use the same method.

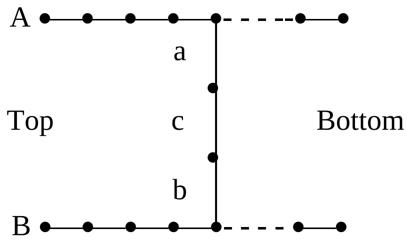
To close this section we mention a few points. In general the primaries are axioms, observables, invariants. The paths from a primary are derivations, deductions, sentences, formal proofs, arguments, and the word, concept-name, derived is a theorem. If we look at the set of interpretations, the realization, of a specific KH, then they are pairwise isomorphic via that KH.

Such an isomorphism is called a translation. Two interpretations, I_1 and I_2, of two different KH's K_1 and K_2, are not isomorphic unless K_1 and K_2 are, but we may find sub-hypernets H_1 of K_1 and H_2 of K_2 such that H_1 and H_2 are isomorphic, and then we will have corresponding isomorphic sub-hypernets of I_1 and I_2. If one is a KH then we can try to teach it and then extend the isomorphism inside its KH, "broadcasting" predicted data and information into the other KH. This is the essence of teaching the second KH, as far as possible, by using the first as a teaching model.

10. Example Of a Plan: An Information Ladder

DNA is complex: we must simplify as usual. The main features that we select are: it stores a plan, it must be able to change that plan, it must be able to copy that plan (reproduce itself), and it must “age” in reproduction. With this in mind, we build a KH plan. Bear in mind that this is just an example of a plan, not a model of a DNA molecule. The plan is called an information ladder.

Consider a “one-rung information ladder” KH. It has two struts composed of vertices linked by binary edges, and a single rung composed of four vertices and three non-binary edges as shown in the diagram below.



The edges in the struts A and B are all binary. The vertices in the two struts are of the same number. They are of two types:

- information vertices of A; a set I_A , and information vertices of B; a set I_B .
- binding vertices, of which there may or may not be the same number in struts A and B. These can appear anywhere in a strut.

The binding vertices partition into two sub-classes: ladder-to-ladder binders, or *l*-binders, which appear only at the top and the bottom of each strut, and the “age markers”, or t-binders, which appear anywhere in a strut, but of which there may not be the same number in each strut.

Edge a is labelled with all the vertices of I_B plus the two end vertices of the edge. Edge b is labelled with all the vertices of I_A plus the two end vertices of the edge. Edge c is labelled with all the t-binders from both struts and its end vertices.

The *l*-binders serve to join one rung ladders into multi-rung ladders. In a multi-rung ladder all of the information vertices in the extended strut A appear in each b-edge of each rung, and all the information vertices in the extended strut B appear in each a-edge of each rung. Every c-edge in every rung has in its label all the t-binders of the multi-rung ladder. In joining [VGS 2004] ladders which happens, for example, with the top two *l*-binders of one ladder and the bottom two *l*-binders of the other, with the appropriate binding edges coming into place, an accommodation that is assimilated if it produces a new ladder, the four *l*-binders involved now become neutral binders, a new class of binding vertices in a strut.

If we assume an "appropriate environment" – see the turbulent sea notion - in which the "pollutants" are components of ladders, we can get ladders constructed by accretion. We will get many accommodations of ladders, some of which are assimilated, i.e. produce new ladders. Thus ladders "grow", and different kinds of ladder KH's mean different plans. The primaries of a ladder could be taken to be its top two *l*-binders for example, in which case its goals are the bottom two *l*-binders. Derivation is immediate for a ladder. It stores information: it is a KH view of a plan if we choose a particular derivation direction for each step and add a dummy root.

How is all this information stored in the ladder? Consider for example, an A strut

a n d t c

•—•—•—•—•

where d is an information vertex, a and c are *l*-binders, n is neutral, and t is a t-binder. This is stored in the b-edge of a rung as the set label

{a, n, d , t, c},

from which the A-strut can be reconstructed. The example is unrealistic, but illustrates the point. Notice what happens if we delete the t from the label of this b-edge!

Any accommodation of a ladder is a "mutation" of that ladder. Some are "assimilated"; are ladder building ones. Each ladder is a join of cluster KH's (with the exception of the four *l*-binder vertices), one for each a-edge and one for each b-edge, each of the clusters containing the appropriate c-edge. In each cluster, some strut vertices may be redundant information vertices of the ladder. Some clusters may contain both struts, again with the exception of the four *l*-binder vertices. "Neutral" vertices may be a necessary part of such a cluster.

Consider now deletion of vertices and edges from a ladder. If we delete any binary edge we break up the ladder. If we delete any vertex we break up the ladder. If we delete an a-edge, or a b-edge, or a c-edge we break up the ladder. In each case of deletion we are left with "pollutants" in the form of relation nets. If we delete a single t-binder then an interesting possibility arises. For simplicity we consider the case of a one-rung ladder. The c-edge will disappear, and we will be left with strut A with edge a and strut B with edge b. Assuming that the t-binder is deleted from strut A, we now hypothesise that the interaction of strut A, with edge a, with the environment may

- allow strut A, with the “information” in edge a, to duplicate the strut B, edge b, side to produce a copy of the ladder, but without the deleted vertex.

This copy is called the parent ladder, and it can only reconstruct itself as many times as it has t-binders, so it is “one unit older” than the original ladder. Extending the hypothesis to the B strut, with edge b, it produces a duplicate of the strut A, edge a, side, so we get a copy of the original ladder, the daughter ladder, with all its t-binders! Dare one imagine a “cancer situation” in which the parent ladder picks up a duplicate of the deleted t-binder every time it duplicates itself, so that it never “ages”, and possibly its daughter ladder never “ages”? This reproduction process will slow when the parent ladder loses its last t-binder, thus being unable to copy itself and becoming “debris”, leaving a daughter ladder with perhaps only one t-binder. We are eventually left with relation net debris in the environment, but new ladders can accrete.

Accommodations, and mistakes in information copying, add to our simple to complex to simple scenario, making and breaking information ladders, taking out information from them and forcing redundant information into them. As ladders grow by joining with other ladders, we start with clusters, which store information, and join them. Some may constitute redundant information with respect to the resulting ladder, and one can expect that an “information complex” ladder will contain redundancy picked up from the environment as it “grows”. Is this “origin” and “natural selection”, evolution, in the ladder environment? If so then perhaps redundant parts of the ladder store previously useful information, previously useful plans, so that they trace the “history” of development of the ladder. “Redundant” parts may be telling a story!

Does the process sketched here conform to “method”?

In a given situation we have made observations of molecules and found that some play a key role: they are of a certain kind, picked out from the rest; inductively abstracted invariance. We name them DNA molecules. In many different such situations we find different DNA molecules. In all observations of ladder molecules of different kinds we have also inductively abstracted the invariants. Of these we chose vertices (molecules), “links”, “struts” and “rungs”. The observables are “molecular components”, the invariant primary relationships among them are “links”, “struts” and “rungs”. We construct a model representing these chosen invariants by a (referent free) KH, one KH for every kind of specific ladder molecule: our information ladder model. We apply reason, to the model, using what we know of KH’s and then hypothesising a conducive environment (also based on observations, but not part of the model). We make predictions from the model, such as “aging” and “reproduction” (copying). All this required “adjustments”; trial-and-error, and communication, i.e. procedural learning and declarative learning, and constitutes a “plan” to achieve predictions. Finally we “empirically test” the predictions in the “real world”; for example would we observe aging and copying in every ladder, do we observe redundancy, do we observe mutations, and do we observe incorrect copying? The model and reasoning, the theory, will be found incomplete or be falsified; i.e. it will have a limited domain of functionality or make at least one incorrect prediction. Of course we already knew about some of what should be predicted: that is part of setting up the model. The exciting thing is a model and reasoning, a theory, that predicts something completely new; unexpected and unsuspected. Then we are “in business”. Note that a theory is “the body” of a plan.

Our theory of ladders has some support if it models DNA. But now we ask questions such as “what information is stored”; “how is that information used”, and “what is the effect of its use”? The situation is hugely complex! Our theory is clearly incomplete. Extending it will lead to its falsification. This enables us to determine the domain of functionality of our theory. It also forces us to replace it with another that has at least the same support in that domain of functionality, that extends the domain of functionality, or that makes us take a closer look at the “support” of the first theory. This whole procedure continues in a never ending outward spiral; and that is what is meant by “learning”.

The plan of this paper partially conformed to “method”. The paper will be “tested” and found wanting!

11. Conclusion

We have a referent free formal language L and its KH, K. K has a realization consisting of a potentially countably infinite set of CRKS's , each of which is an interpretation of K. Thus K can be regarded as the invariant (structure) inductively abstracted from all its interpretations, and these are pairwise isomorphic "via K".

Thus, in terms of KH's and CRKS's, we move

- from K to the formal language CRKS's , M_j, by interpretation
- some of these M_j are "real world" CRKS's, the I_k
- from the $I_{k,}$ to scientific method.

If we introduce a dummy vertex root into the KH, K, of L, then we have

- K – the master plan
- Partial interpretations produce the CRKS's M_j and I_k, the formal languages and the "real world" applications respectively

where two of these CRKS's that have the same KH abstraction differ only in notation; are mutually translatable, and two of these CRKS's that have different KH abstractions differ in primary/axiom sets and perhaps also in notation, but not in the notion of rules of inference, just in statements of them. In the case of an I_k, the model is the primary CRKS of I_k, and the theory is I_k. We are involved in the art of constructing languages used to describe situations.

In all this we use precisely stated "arguments" in a natural meta-language. Construction and analysis of arguments is dealt with in the proposition (or symbolic or statement) logic structure. These natural meta-languages, finite but not completely constrained, can "grow" by adopting words from other languages, and by defining new words, or new meanings of words, by letting the "new" word denote a precisely specified paragraph of one or more sentences. Definitions not only enhance the meta-language, but make it more precise and concise, enabling us to construct precise statements of relationship and precise arguments, thereby developing precise communication skills.

Using the notions of KH and CRKS, we believe that we have exposed some fundamental links between science, teaching and learning, and mathematics. Basically, all use the same method: invariance, trial-and-error, declaration (i.e. communication), reasoning to produce predictions, and testing of these – in mathematics by truth tables, by searching for counter examples, or by proving that a contradiction exists for example. (If we prove a legitimate word w, and we also prove ($\neg$ w), then the relevant structure – theory – is "falsified".) We have mentioned the role of each of these fields, and how "method" is implemented in each. The "pattern" is set in the KH of L, and preserved in its CRKS (partial) interpretations. Given two different applications, two different CRKS's, one may be "partially isomorphic" with the other: a sub-CRKS of one can be isomorphic with, i.e. an analogy of, some sub-CRKS of the other. (This can also hold for sub-relation-nets of the two CRKS's which is useful). Each of the "real world" CRKS's can be regarded as the basis of a plan, and every plan is inadequate because the chosen primaries and primary statements of relationship are. A theory involves the development of a plan!

Working in the realm outlined above are many specialists, from logicians to mathematicians and applied mathematicians, to the practical "real world" problem solvers. It is long past time that we brought all educators into it, perhaps by means of the simple notion of CRKS's which seems to run right through this realm and may provide one uniting theme for it.

Our work is probably properly classified as applied mathematics, so it is not coincidental that it has multidisciplinary implications, for example in the fields of computing , specifically knowledge representation, psychology, sociology, networking, education, the psychology of education, cognitive science, and discrete structural modelling in general, but also here in some facets of methodology. We seem to have exposed one possible technique for handling statements of relationship in a formal manner.

Bearing in mind the existence, implied by current (continuum) physics, of smallest units of space and of time, we should notice the following. Real and complex numbers are the basis of a model, in physics, that is more complicated, not more simple, than the real world modelled. There is detail in the model that cannot apply to the situations modelled, at least at small scale. The sole advantage is that of analytical solution of some of the relevant equations, but the approach does not seem to conform with the method of science. Is it not time for discrete structural modelling in physics?

If there is a "mission" here, it can be summarized as follows. Observations lead to questions: question everything. Questions lead to the search, by trial-and-error, for invariant observables and invariant relationships among them. More questions lead to problems. Problems lead to speculation.

Speculation leads to suggested answers. Suggested answers lead to reasoning and plans. Plans yield predictions, and predictions must be tested. Tests either support a plan, or demand modification or replacement of the plan: they determine its domain of functionality. Changing and replacing plans is the essence of learning, and as we experience learning we develop communication skills.

We claim that basic universal education is essential. With the human population approaching 9 billion on a planet that can adequately support only a small fraction of that number, the future existence of the species depends upon individuals making informed decisions which benefit the species and therefore, in the long term, themselves. To reach this goal we must better understand learning, and we must all be teachers, so that we can approach an Age of Reason.

To close we mention that CRKS theory could constitute a model for the systematic construction, manipulation and management of a “semantic net” – see [VGS2004 pg 100 to 119 for example]. Without some foundation upon which to construct a net, and some fixed plans for its extension, the “building will develop flaws” and continual ad hoc “patching” will be required, or else we could have a net which steadily reduces to chaos. We could indeed identify the terms knowledge hypernet and semantic net.

The semantics and syntax of a given language are encapsulated in its KH, but both can be expressed in natural language in a variety of ways by way of the CRKS interpretations of that KH. A formal language is characterized by its fixed KH, and a natural language by changing KH’s, preferably with a constant core sub-KH in common. A KH specifies the semantic and syntactical form of a language while its (partial) interpretations specify the uses of that language.

Acknowledgements: We are indebted to the many researchers who have contributed to the fields addressed in this paper. We would like to acknowledge our debt to all the friends with whom we had enlightening conversations, and those with whom we worked constructively, during the course of many years: they all helped to crystallize these ideas, but we take full responsibility for the shortcomings of the ideas. Thanks to Prof. Dr. J. Heidema and Dr. A. E. Geldenhuys for comments on the drafts of this paper. That not all of these may have been satisfactorily addressed is the sole responsibility of the authors. Thanks are due to Sue Fenton for typing the draft, without whose willing help this would not have reached final form. In memory of Mike Sheridan, who said that for every problem one must have a plan. Finally, thanks to our families for patience and support.

References

[Gel1993] Geldenhuys, A.E., 1993: Concept-Relationship Knowledge Structures: Computer supported applications in teaching and learning. Ph.D. Thesis, Unisa.

[GVS1999] Geldenhuys, Aletta E., van Rooyen, Hendrik O., and Stetter, Franz: Knowledge Representation and Relation Nets. Kluwer Academic Publishers, 1999.

[VGS2004] Van Rooyen, Hendrik O., Geldenhuys, Aletta E., Stetter, Franz: Modelling Knowledge Systems using Relation Nets and Hypernets. Technical Report TR-2004-009, Dept. of Mathematics and Computer Science, Univ. of Mannheim, 2004. http://madoc.bib.uni-mannheim.de/madoc/volltexte/2004/850/pdf/TR200409.pdf

[VR1976]Van Rooyen, H. O., 1976. Binary networks in graph theory. Ph. D. Thesis, Unisa.

[VS2007] Van Rooyen, Hendrik O. and Stetter, Franz: Digraphs, Knowledge Hypernets, and Neurons. Technical Report TR-2007-009, Dept. of Mathematics and Computer Science, Univ. of Mannheim, 2007. http://madoc.bib.uni-mannheim.de/madoc/volltexte/2007/1662/pdf/Neurons.pdf

[VS2008] Van Rooyen, Hendrik O. and Stetter, Franz: Neurons, Knowledge Hypernets, and Information Flow. Technical Report TR-2008-001, Dept. of Mathematics and Computer Science, Univ. of Mannheim, 2008. http://madoc.bib.uni-mannheim.de/madoc/volltexte/2008/1910/pdf/Neuron2008_bearb.pdf