Real Time Machine Deduction and AGI

Peter G. TRIPODES
pgtripodes@cs.com

Abstract: Consistent with the ultimate goals of AGI, we can expect that deductive consequences of large and grammatically varied text bases would not be generated by sequential application of inference rules but would instead be recognized in a single massively parallel pattern matching operation on their semantic structures which executes near instantaneously. We describe an approach to realizing such a capability in graphical terms whereby semantic structures are depicted as certain graphical arrays, and deductive relationships are determined by requisite pattern connections within and among those arrays.

Keywords: Real time deduction, massively parallel pattern matching, semantic structure as graphical array

1. Realizing Natural Language Deduction Machines

1.1 Natural Language Deduction Machines

One goal of AGI is to realize natural language deduction machines, by which I mean machines that can near instantaneously identify deductive consequences of general text like newspapers, regardless of the number or grammatical complexity of the sentences involved, and do so without the assistance of a human. The difficulties that lie in the path to realizing such machines seem insuperable and the prospect of realizing them to any degree of generality has been virtually abandoned. Yet such machines appear central to the realization of the wider AGI goals involving machine intelligence and reasoning in the handling of natural language[1].

1.2. Two Problems in Realizing Natural Language Deduction Machines

Broadly speaking, there are two problems that need to be solved in order to realize such machines. One is the *massively parallel deduction problem*, which is the problem of defining sentence representations and massively parallel deductive operations to operate on them applicable to texts of arbitrary size and grammatical complexity. The second problem is the *conversion problem*, which is the problem of formulating procedures for on-line inter-conversion between sentences and their representations.

1.3. Scope of Paper

In this paper, I suggest an approach to the massively parallel deduction problem which defines sentence representations as certain graphical arrays, and treats deduction as a massively parallel and simultaneously executed pattern matching operation which executes on them, thereby supplanting sequential forms of deductive inferencing [1] [2] [3] [4]. We refer to this proposed operation as *immediate deductive recognition* (IDR). The conversion problem is not treated here inasmuch as it involves complex pragmatic and grammatical issues that are beyond the scope of this paper. Some of these issues are addressed in an unpublished paper by the author entitled "A Theory of Readings [5]."[2]

1.4. Immediate Deductive Recognition (IDR)

While the theory underlying IDR is essentially that of model theoretic semantics, we describe IDR wholly in graphical terms which much more clearly exhibit the kinds of patterns that are to be matched

[6] [7] [8] [9] [10]. Accordingly, we define a *local graph* of a sentence (or of a set of sentences) as a linked array of node-and-arc graphs which collectively graphically depict the denotation of a sentence (or of a set of sentences) relative to a given "permissible" (model theoretic) interpretation which validates it, that is, an interpretation under which that sentence (every sentence in the set) is true. Permissible interpretations are interpretations restricted in certain ways to render local graphs finite and comparable (see section 6.1). And we define the *global graph* of a sentence (or set of sentences) as a linked array of its local graphs relative to all its permissible validating interpretations. IDR executes as a single-step massively parallel operation which simultaneously compares every local graph of a given set of sentences against every local graph of a given sentence and, simultaneously, against every local graph of its negation.[3] If every local graph of the given set of sentences is compatible with some local graph of the given sentence, or if every local graph of the given set of sentences is incompatible with every local graph of the negation of that given sentence, the given sentence is thereby determined to be a deductive consequence of that given set of sentences. When this determination is made by machine we will refer to the operation which makes it as *machine IDR*. Compatibility (incompatibility) of a local graph of a set of sentences with a local graph of a given sentence is defined in graphical terms and corresponds to the model theoretic circumstance that there exists (does not exist) a permissible interpretation which validates both the set of sentences and the given sentence and which is such that the local graphs in question respectively depict their denotations relative to that interpretation.

1.5. Human IDR.

The motivation for our approach to machine IDR derives from an apparent human capability, variously noted in the literature[3], to carry out near-instantaneous deduction (NID) on natural language sentences, particularly when they involve small numbers of simple sentences, and to do so without apparent intervening conscious thought or calculation. *We hypothesize that the cognitive mechanism underlying NID may be a form of pattern matching.*[4] Accordingly, we will refer to the operation of this hypothesized pattern matching mechanism in humans as *human IDR*.

1.6. Hypothesized Underlying Cognitive Mechanism of Human IDR[5].

The underlying mechanism of human IDR may be akin to an individual's near-instantaneously recognizing the face of a friend from a photograph by matching (remembered) selected structural characteristics of the face of the friend (such as the set of the eyes or shape of the nose) against those observed in the photograph. That is, an individual's near-instantaneously recognizing simple deductive consequences of given sentences may be carried out by matching selected structural characteristics of the given sentences against selected structural characteristics of those deductive consequence. We will argue below that these structural characteristics may be intuitively perceived patterns of semantic connections among the underlying logical morphemes of those sentences.

1.7. Common Examples of NID

Examples of near instantaneous deduction (NID) appear quite common, and would include cases such as near instantaneously comprehending - without specific practice or familiarization with the specific sentences involved – that the sentences "John loves Mary" and "Mary is loved by John" necessarily follow from each other, and that the sentence "John is dating a waitress" necessarily follows from the sentence, "John is dating many waitresses," as well as from the pair of sentences, "John is dating Mary" and "Mary is a waitress." Instances of NID, while common, appear to be limited for most individuals to small sets of simple sentences, such as the sentences in these examples. This is not to say that humans could not make deductive determinations on larger sets of sentences and of greater complexity, but only that they would not be able (in most cases) to do so without conscious thought and calculation, and certainly could not do so near instantaneously. Near instantaneous deductions of these simple kinds have

been widely noted by other researchers, under various names, such as “immediate inference.”[11] The difference between the usual treatment of NID and the account given here is that we hypothesize that the underlying mechanism of NID is a form of massively parallel pattern recognition which we refer to as human IDR[5]. The question is how would such a mechanism work? We examine this next.

2. How is Human IDR Possible?

2.1. The Logic of the Language May Be "Hardwired" in the Brain

Human IDR may have its roots in some sort of neural representation of what might be called "the logic of the language," that is, in structures within the brain that represent the semantic structures of logical morphemes and their interconnections which determine deductive relationships[10]. We would regard these structures as specific to certain languages or language groups and hypothesize that an individual’s understanding of their logical function develops in the course of his or her "learning” a particular language. The logical morphemes in the above (English) examples would be those expressed there by the word-strings (i.e., morphs), "many," "is," (in the sense of both predication and identity), "es," "ing," "ed," "a," "and," and "by." (Logical morphemes are to be distinguished from so-called non-logical or “lexical” morphemes such as those expressed in the above examples by the word-strings, "John," "Mary," and "waitress.") The fact that human IDR is apparently limited to simple sentences may be due to inherent ambiguities in complex sentences deriving from multiple possible functional roles for the logical morphemes occurring in them which obscure their deductive consequences. Simple sentences tend to be less ambiguous in this regard. And the fact that human IDR is apparently limited to small sets of premise sentences may be due to neuro-physiological limits in representing the myriad interconnections among logical morphemes occurring in large numbers of sentences. Machines would, of course, suffer no such limitations.

2.2. Memory: In the Brain and Machines

I use the term “memory” in the sense of language-based (or declarative) memory, and intend it to apply equally to the brain and machines. By language-based memory I mean the store of information held in the brain or in a machine which has, by one means or other, been "entered into it" in the form of sentences. In the case of the brain, these would be sentences which one has either read or heard. In the case of machines, these would be sentences entered into machine memory by any of the customary sentence-input means for machines.

2.3. A Massive Parallelism Assumption Underlying Human IDR

It seems reasonable to assume that the brain executes IDR by a massively parallel mechanism of some kind that is, by a mechanism which executes all subtasks involved in carrying out IDR in parallel and, moreover, executes them globally and simultaneously across the whole of memory. Some version of this assumption is widely shared by neuroscientists and others as necessary to explain how the brain accomplishes certain reasoning tasks so rapidly [1] [2] [3] [10].

2.4. How Memory Is Represented Is Key

The way that sentences are represented in memory wholly conditions the manner in which and the extent to which their deductive consequences can be accessed by massively parallel means. In order to facilitate human IDR, therefore, memory would need to be represented in the brain in such a way that the structure of its deductive consequences is reflected in its own structure to a degree sufficient to permit the massively parallel recognition of at least the simpler of those deductive consequences. And this requires,

in turn, that the structure of memory be "global" in the sense that all connections among the components of memory relevant to deduction be included in its representation. These requirements strongly suggest that that representation should be "semantic" rather than "syntactic," that is, that it should represent the semantic structures of sentences rather than their syntactic forms: first, because semantic structures are better preserved under deduction than are syntactic forms; and second, because connections among memory components relevant to deduction are better depicted as semantic structures than as syntactic forms. When memory is represented syntactically, its deductive consequences cannot be directly recognized from that representation, but can only be generated from it and into forms which look very different from those they were derived from. On the other hand, semantic structures of memory can be defined in such a way that their structure is preserved under deductive inference. We next inquire into the kind of structures of memory would expedite human IDR.

3. Back to the Brain: The Wiring Hypothesis[5]

3.1. Is The Brain "Wired" For IDR?

The apparent capability in humans to carry out deductions on simple sentences near instantaneously and which appears to require no conscious thought or calculation when doing so suggests that the brain may be "wired" for it. Let us suppose that there are neural configurations in the brain that physically encode sentence information as memory. We can refer to these configurations as "sentence representations" (ignoring for the moment whether these representations are more appropriately of sentence forms or of sentence meanings). By the hypothesis that the brain is "wired" for IDR I mean that there may exist pathways in the brain inter-linking stored sentence representations in such a way as to facilitate the immediate recognition of their collective deductive consequences. These pathways, i.e., "wiring", may develop (through an evolutionarily conditioned disposition to do so) in the course of learning a language. Their development may consist in growing new pathways or in reinforcing the capacity of existing pathways to transmit signals, and would progress as the language and the inferences framed in it are practiced. One might regard this hypothesized process as a physical encoding of the "logic of the language" within the brain.

3.2. How Might The Brain Be Wired For IDR? A Proposal

Assuming that the brain is "wired for IDR" in the above sense, what sort of wiring design would be most efficient for its execution? Putting the question teleologically, what would be the most efficient wiring design for the brain to have evolved in order to maximally facilitate the immediate recognition of simple deductive consequences from memory? ("Efficiency", for the brain - as for machines - means compact storage and rapid execution.) To give this question a more determinate form: Let us suppose that a given set X of simple sentences is represented in memory as a pattern X* of electro-chemical signals, and that some possible (simple) deductive consequence Y of X (being considered) has its negation not-Y represented in memory also as a pattern, say [not-Y]*, of electro-chemical signals. Let us suppose further that the brain would seek to determine whether [not-Y*] is inconsistent with X* in the most efficient way possible. Our question then becomes: How might the structures of X* and [not-Y]* be related in order to provide the most efficient mechanism for determining that these structures were incompatible, and hence that Y was a deductive consequence of X? We note that it would not be efficient to have X* sequentially generate patterns in search of some that were inconsistent with [not-Y*], that is, by sequentially generating a contradiction from X* and [not-Y]*, for such a "generation procedure" would appear to take much more time than the brain apparently devotes to extract deductive inferences "essentially instantaneously." *The most efficient mechanism would thus appear to be one which required no more storage than that already devoted to memory - to represent X* - and required now to represent [not-Y*] as well.*

3.3. A "Most Efficient" Mechanism Would Be Semantic

The requirement that a deductive recognition mechanism be efficient entails that no additional structures be introduced than those already constituting memory and, therefore, that the deductive consequences of memory must already be present and accessible within the structure of memory. We had earlier stated (Section 2.4) that a semantic representation of memory for IDR would tend to be superior to syntactic representations in this regard. However, not all proposed semantic memory structures are equal in this regard: some tend to exhibit their deductive consequences to a greater degree than others. In the following sections we will describe a notion of semantic structure for memory that is specifically designed to be maximally preserved under deductive inference.

4. Back to Machines

4.1. Analogue of Machine IDR in Elementary Algebra

4.1.1. Properties of Algebraic Graphs

Machine IDR for natural language generalizes three familiar properties of elementary classroom algebra of the plane. *The first property of algebraic graphs* is that individual solutions of equations, inequalities, and systems of equations and inequalities can be graphically depicted as points on the plane, which are said to "correspond to" those solutions, and their solution sets can be graphically depicted as corresponding figures, i.e., as "graphs" on the plane (in the usual sense). *The second property of algebraic graphs* is that a given equation or inequality is a deductive consequence of a given system of equations and inequalities if and only if every point graphically depicting a solution of the system is either identical with some point graphically depicting a solution of the given equation or inequality or is distinct from every point graphically depicting a solution of the negation of that equation or inequality. *The third property of algebraic graphs* is that the determination of whether given points on the plane are distinct or identical can be approximated by a massively parallel comparison operation simultaneously executed on their pixel approximations (i.e., on the pixels containing/enclosing them) that assesses those pixels as distinct or identical, and to thereby make a correspondingly approximate determination that the given equation or inequality is or is not a deductive consequence of the system[4].

4.1.2. Generalizing Algebraic Terminology to Natural Language

In generalizing to machine IDR for natural language deduction we generalize "equation or inequality" to sentence; "system of equations and inequalities" to sets of sentences; "solution of an equation or inequality" to "denotation relative to an interpretation under which that equation or inequality is true," "solution of a system of equations and inequalities" to "denotation of a set of sentences relative to an interpretation under which all sentences of the set are true, "solution set" to "set of denotations," "point corresponding to solution of" to "local graph of," "graph corresponding to solution set of an equation, inequality, or system" to "global graph of that equation, inequality, or system"; and the relations of "identity" and "distinctness" among points to the relations of "graphical compatibility" and "graphical incompatibility" among local graphs.

4.1.3. Generalizing the Three Properties of Algebraic Graphs to Natural Language

The first property of algebraic graphs generalizes to natural language as follows: A denotation relative to an interpretation under which a given sentence (or every sentence in a given set of sentences) is true can be graphically depicted as a local graph (which that interpretation is said to determine), and the set of all such denotations can be graphically depicted as a linked array of local graphs, called a global graph of that sentence or set of sentences. *The second property of algebraic graphs* generalizes to natural language as follows: a given sentence is a deductive consequence of a given set of sentences if and only if every local graph (i.e., point corresponding to some solution) of the set is either compatible with (i.e., is identical with) some local graph (i.e., point corresponding to some solution) of the given sentence or is incompatible with (i.e., is distinct from) every local graph (i.e., point corresponding to some solution) of its negation. (Model theoretically, compatibility of local graphs of two sentences or of two sets of sentences means that there is some single interpretation which determines both graphs, that is, that there is some single interpretation under which the sentences or all the sentences in the sets in question are true; and incompatibility means that there is no such single interpretation.) *The third property of algebraic graphs generalizes* to natural language as follows: the determination of whether every local graph of a given set of sentences is compatible with some local graph a given sentence or is incompatible with every local graph of its negation, can be made by a massively parallel comparison operation simultaneously executed on all local graphs of the set of sentences, all local graphs of the given sentence, and on all local graphs of its negation, to assess in each case whether or not they are compatible, and to thereby make a determination that the given sentence is or is not a deductive consequence of the given set of sentences[6]. These determinations are approximate in the sense that the constitution of local and global graphs is dependent on the range of permissible interpretations used in their definition, their accuracy increasing generally as larger sets of permissible interpretations are introduced. While the situation is more complex for natural language than for the much simpler language of elementary algebra, it is clear that the underlying idea is the same.

5. Key Syntactic Notions for Natural Language

5.1. Relation-expressions.

A relation-expression is a syntactic representation of a character string which, in a given occurrence, is regarded as denoting (relative to a given interpretation) an m-place relation which holds among given m-tuples of elements of the universe of discourse.

5.2. Thing- expressions

A thing-expression is a syntactic representation of a character string which, in a given occurrence, is regarded as denoting a "thing" (relative to an interpretation), where a "thing" is a set of subsets of the universe of discourse.

5.3. Modifier-expression

A modifier-expression is a syntactic representation of a character string which, in a given occurrence, is regarded as denoting a function on the denotation of the expression to which it is applied, and which assigns to that denotation a "thing" or “relation” (relative to an interpretation), the structure of which is determined by the function denoted by that modifier. A thing or relation-expression to which a modifier is applied is said to be *governed* by that modifier, and what remains of that thing or relation expression when its governing modifiers have been removed is called *the base* of that thing or relation expression. Certain modifiers applied to thing expressions are syntactic representations of ordinary determiners such as “all,”, “some,” “many,” etc. Certain modifiers applied to relation expressions are syntactic representations of temporal operators.

5.4. Sentences

A sentence is a syntactic representation of a character string which, in a given occurrence, is regarded as denoting an "event" or "state of affairs," and is composed of (i.e., is formalized as): (1) an m-place relation- expression r^m, together with (2) m thing-expressions $a_1, \ldots, a_m$, each of which denotes a set of sets of elements of the domain of discourse, which elements stand in the relation denoted by r^m. The m thing-expressions $a_1, \ldots, a_m$ remarked in (2) are referred to as the m major thing expressions of the sentence, and the sentence composed of an m-place relation-expression r^m followed by the m major thing expressions $a_1, \ldots, a_m$ is schematically indicated as $r^m (a_1, \ldots, a_m)$.

6. Key Semantic Notions for Natural Language

6.1. Interpretations

An interpretation is a function which assigns a set to every meaningful expression, including individual sentences and sets of sentences, which are referred to as *the local denotation of that expression relative to that interpretation*. We restrict interpretations to *permissible* ones, that is, to interpretations that assign the same finite set to every thing expression and the same function to every modifier; so that the only way two permissible interpretations could differ would be in the structure of the relation they assign to the same relation expression. The reason for this restriction is to make semantic and graphical structures comparable and computationally coherent. An expression is a *constant* if all permissible interpretations assign the same denotation to that expression; otherwise that expression is said to be variable; accordingly, relation expressions are the only variable expressions.

6.2. Set Theoretic Criterion for Deducibility

Every sentence or set of sentences, relative to a given interpretation, has a translation into set theory which expresses – in set theoretic terms - the relationship which must hold among the local denotations of its component expressions – relative to that interpretation - in order for that sentence or all sentences in the set of sentences to be true under that interpretation. This translation into set theory is referred to as the *truth condition* of that sentence or set of sentences relative to that interpretation. *The local denotation of a sentence* (or set of sentences) *relative to an interpretation f under which* that sentence (every sentence in the set) is true *is a set whose defining condition is the truth condition of that set (or set of sentences),* that local denotation is then said to be *determined by that interpretation.* The global denotation of a sentence (or set of sentences) is the set of its local denotations determined by all permissible interpretations under which it (every sentence in the set) is true. Local denotations of sentences (or of sets of sentences) are *compatible* if they are determined by a single interpretation, and are *incompatible* if there is no single interpretation which determines them. A given sentence is a *deductive consequence* of a given set of sentences if it is true under every interpretation under which all sentences of the set are true [12]. This is equivalent to the following: *A given sentence is a deductive consequence of a given set of sentences if every local denotation of the set relative to some interpretation under which its member sentences are true is incompatible with every local denotation of the negation of the given sentence relative to some interpretation under which it is true.*

6.3 Graphical Criterion for Deducibility

A *local graph* of a sentence or set of sentences relative to an interpretation under which it is true is a graphical entity which represents its local denotation relative to that interpretation in the sense that that graphical entity and the local denotation it represents are, in principle, inter-retrievable, and that the set theoretic relationships into which denotations enter are equivalently expressible as graphical relationships among the local graphs which represent them. (In the algebraic case, local denotations are solutions and local graphs that depict them are points on the plane.) The global graph of a sentence or set of sentences is a connected array of its local graphs, and depicts its global denotation. (In the algebraic case, global graphs are ordinary Cartesian graphs on the plane that depict solutions sets of equations, inequalities, and systems). *Two local graphs of sentences or of sets of sentences are compatible if they depict denotations that are compatible, and are incompatible if they depict denotations that are incompatible*. We will shortly give a graphical characterization of the relation of compatibility among graphs of sentences or sets of sentences. (In the algebraic case, points are compatible if and only if they are identical.) A given sentence is a deductive consequence of a given set of sentences if every local graph of the given set of sentences is incompatible with every local graph of the negation[3] of the given sentence. (In the algebraic case, the fact that compatibility of local denotations implies identity, whereas in the natural language case it does not, has several interesting consequences. One is that while in the algebraic case, the global graph of a system of equations and inequalities is the graphical intersection of the global graphs of its member equations and inequalities, this does not hold in the natural language case. A second consequence is that the following two statements (a) and (b) are equivalent in the algebraic case but in equivalent in the natural language case: (a) every local graph of the given system is incompatible with every local graph of the negation of a given equation or inequality; (b) every local graph of the system is also a local graph of the given equation or inequality.) In order to avoid attaching any sort of labels to graphs to indicate the semantic interconnections among their components, we use special arcs to join graphical components which represent the same denotation relative to a given interpretation. Accordingly, we refer to graphs which are joined in this way as "linked graphs."

7. Semantic Representation of Sentences

7.1. Positive and Negative Relational Profiles

If a is a thing-expression, let a^B be the base of a. Let f be an interpretation, let $r^m(a_1, \ldots, a_m)$ be a sentence, and let $CP(r^m(a_1, \ldots, a_m))$ be the Cartesian product $f(a_1{}^B) \times \ldots \times f(a_m{}^B)$. Finally, let $f(r^m)^c$ be the complement of the relation $f(r^m)$. Then we define the *positive relational profile of $r^m(a_1, \ldots, a_m)$ under f, which we write as POS_f ($r^m(a_1, \ldots, a_m)$), to be the intersection of the set $f(r^m)$ with $CP(r^m(a_1, \ldots, a_m))$*, and we define the *negative relational profile of $r^m(a_1, \ldots, a_m)$ under f, which we write as NEG_f ($r^m(a_1, \ldots, a_m)$), to be the intersection of the set $f(r^m)^c$ with $CP(r^m(a_1, \ldots, a_m))$.* [Analogy with elementary algebra of the plane with variables "x" and "y" is imperfect but instructive: Since there are no base expressions in any algebraic expression, $a_i{}^B$ is simply a_i , and the Cartesian product $f(a_1{}^B) \times f(a_2{}^B)$ becomes the pair $\langle f(a_1), f(a_2)\rangle$; the only relations r^m are the binary relations of equality and inequality and their negations, so that m = 2; $f(a_1)$ and $f(a_2)$ are real numbers built up out of the real numbers f("x") and f("y") using the customary algebraic operations of plus, times, exponentiation, etc.; the positive relational profile POS_f ($r^2(a_1, a_2)$) of $r^2(a_1, a_2)$ under f becomes the intersection of the set $f(r^2)$ with $CP(r^2(a_1, a_2))$, which is simply the pair <f("x"), f("y")> if the result of respectively replacing "x" and "y" by f('x") and f("y") in $r^2(a_1, a_2)$ is a true sentence of elementary algebra, and is the empty pair otherwise; and the negative relational profile NEG_f ($r^2(a_1, a_2)$) of $r^2(a_1, a_2)$ under f becomes the intersection of the set $f(r^2)^c$ with $CP(r^2(a_1, a_2))$, which is the pair <f('x"), f("y")> if the result of respectively replacing "x" and "y" by f('x") and f("y") in $(r^2)^c(a_1, a_2)$ is a true sentence of elementary algebra, and is the empty pair otherwise.]

7.2. Chain Functions and Traces:

Let f be an interpretation, and let $f(r^m)$, $f(a_1)$, …, $f(a_m)$, be denotations of r^m, a_1, …, a_m, respectively. We define a chain function through the sequence ($f(a_1)$, …, $f(a_m)$) as a function g which assigns, to every set $f(a_1)$, for $1 \leq i \leq m-1$, and for every element y belonging to $U(f(a_1))$, that is, belonging to any of the member sets of $f(a_1)$, a set g(i,y) belonging to one of the sets in $f(a_{i+1})$. We note that this definition is proper on any sequence ($f(a_1)$, …, $f(a_m)$) of denotations of thing-expressions a_1, …, a_m relative to f. Let g be a chain function through the sequence ($f(a_1)$, …, $f(a_m)$). Then we define the trace of g through ($f(a_1)$, …, $f(a_m)$) as the set: $\{(z_1, \ldots, z_m) \in D^m$ // for some $x_1 \in f(a_1)$, $z_1 \in x_1$, and $z_2 \in g(1, z_1)$, and $z_3 \in g(2, z_2)$ and … and $z_m \in g(m-1, z_{m-1})\}$. There are in general many possible chain functions through the sequence ($f(a_1)$, …, $f(a_m)$) of thing expressions of $r^m(a_1, \ldots, a_m)$ relative to f, but there is exactly one chain function whose trace is identical with the positive relational profile of $r^m(a_1, \ldots, a_m)$ relative to f if $r^m(a_1, \ldots, a_m)$ is true under f, and there is no such chain function if $r^m(a_1, \ldots, a_m)$ fails to be true under f. . [Continuing with the analogy with elementary algebra of the plane: Since $f(a_1)$, $f(a_2)$ are real numbers and, as such, have no member sets, and so the above definition of chain function is not proper on any sequence ($f(a_1)$, $f(a_2)$) of denotations of algebraic terms , so that the above definition of trace does not apply. On the other hand, we can define a trace as a degenerate notion whereby the trace of a sequence ($f(a_1)$, $f(a_2)$) of denotations of algebraic terms is simply that sequence.]

7.3 Denotations of Sentences and Their Graphical Representations

We define the denotation $f(r^m(a_1, \ldots, a_m))$ relative to the interpretation f, in symbols, $Den_f(r^m(a_1, \ldots, a_m)$, as the set:
$\{\{ <f(r^m), v> // v \in POS_f (r^m(a_1, \ldots, a_m))\}\}$ U $\{ <f(r^m)^c, v> // v \in NEG_f (r^m(a_1, \ldots, a_m)) >\}\}$, if there is a chain function g through the sequence ($f(a_1)$, …, $f(a_m)$) such that the trace of g through ($f(a_1)$, …, $f(a_m)$) is identical with $POS_f (r^m(a_1, \ldots, a_m))$); and is $\{\Theta\}$, otherwise. *The set $Den_f (r^m(a_1, \ldots, a_m)))$ can be completely graphically represented as a network of nodes and connecting arcs, where the un-negated connecting arcs graphically represent the relation and each m-tuple of the nodes they connect graphically represents m elements of the domain which stand in that relation, and where negated connecting arcs graphically represent the complement of the relation and each m-tuple of the nodes they connect graphically represent m elements of the domain which fail to stand in that relation.*[7,8]
[Continuing the analogy with elementary algebra of the plane, this definition reduces as follows: $Den_f (r^2(a_1, a_2)$ is the set: $\{\{ <f(r^2), v> //$ v is a pair $<f(\text{“x”}), f(\text{“y”})>$ such that, when the variables “x” and “y”, when respectively replaced by the values f(“x”), f(“y”) in $f(a_1)$ and $f(a_2)$, $f(a_1)$ and $f(a_2)$ stand in the relation $f(r^2)\}$ U $\{\{ <f((r^2)^c), v> //$ v is a pair $<f(\text{“x”}), f(\text{“y”})>$ such that, when the variables “x” and “y” when respectively replaced by the values f(“x”), f(“y”) in $f(a_1)$ and $f(a_2)$, $f(a_1)$ and $f(a_2)$ fail to stand in the relation $f(r^2)\}$. *This set can be completely graphically represented on the plane in the usual way as an array of points corresponding to the pairs $<f(a_1), f(a_2)>$ which stand in the relation $f(r^2)$. But this graphic representation can be extended to allow for the simultaneous depiction of its graphical complement, namely the depiction of the set of pairs $<f(a_1), f(a_2)>$ which stand in the relation $f(r^2)$. If we were to take this course, which is formally possible, it would require a graphical way of distinguishing the complementary points from the standard ones by some graphical device such as using a different color.*]

8. Structure of Local Graphs.

8.1. Nodes, Arcs, and Paths

Local graphs are composed of two types of basic graphical elements: nodes and arcs. Nodes represent elements of the underlying domain of discourse and arcs represent relations on those elements.

8.1.1. Simple and Compound Arcs

Simple arcs are arcs that join at most two entities. There are three types of simple arcs: (i) arrows, which joins at most two nodes and which can be barred or unbarred; an unbarred arrow represents a relation whose relata are elements of the underlying domain of discourse, and considered in the order indicated by the direction of the arrow, and a barred arrow represents the complement of that relation; (ii) dotted lines, which represent the identity relation when unbarred, and the non-identity relation when barred, and which join either two points to represent that they represent the same or different elements of the underlying domain of discourse, or two arrows to represent that they represent the same relation if both are barred or both are unbarred, or to represent complementary relations if one arrow is unbarred and the other is barred; and (iii) dashes, which represent the logical conjunction of the entities represented by the graphical entities it joins. *Compound Arcs* are arcs formed by joining two or more simple arcs with a graphical unit called a "brace," and represent many-place relations composed of those simple arcs. An arc that is not a constituent of a compound arc is said to be major.

8.1.2. Dot Paths, Arrow Paths, and Mixed Paths

A path is a simple or compound arc taken together with nodes it joins. If the constituent arcs of the path are all dotted lines, the path is called a *dot path*. If the constituent arcs of the path are all arrows, the path is called an *arrow path*, if the constituent arcs of the path are both dotted lines and arrows; the path is called a *mixed path*. An arc which is a constituent of a path is said to be a major constituent of that path if it is not itself a constituent of another constituent of that path. A path is said to be *barred or unbarred* according as its major constituent is barred or unbarred. A path represents that the elements of the domain of discourse respectively represented by the nodes of the path stand in the relation represented by the path. Arrow paths and mixed paths represent lexical relations, the place number of which corresponds to the number of nodes in the path. A single node placed at the origin or terminus of an arrow signifies respectively that the element represented by the node is in the domain or range of the relation represented by the arrow.

8.1.3. Similarity and Identity Linked Paths

Two paths are *similarity linked* if they differ only in the placement of bars on one or more of their constituent arcs, and all corresponding nodes and arcs joined by dotted lines. Two paths are *identity linked* if graphical depictions of the same denotation are joined by dotted lines.

8.2. Local and Global Graphs

8.2.1. Local Graphs

A *local graph* is an array of similarity linked paths. For definiteness, that array is organized in such a way that the corresponding nodes are displayed in vertical columns. We refer to a column of corresponding nodes as a node bank, and to the nth such column in a local graph as the nth node bank of the local graph.

8.2.2. Global Graphs as Similarity Linked Local Graphs
Two local graphs are *similarity linked* if they differ only in one or more constituent paths which are similarity linked, and all corresponding nodes and arcs are joined by dotted lines, and a *Global Graph* is an array of local graphs which are pair wise similarity linked.

Endnotes

1. By "central to realizing wider AGI goals," I mean that without this sort of deductive capability, real time machine execution of other types of reasoning, such as those involved in inductive, probabilistic, or pre-suppositional inference, could probably not be achieved at the level envisioned in AGI. The reason is that deduction structures meaning interconnections within and among expressions which occur in these other types of reasoning as well as functions as a limiting special case for various of them (e.g., inductive and probabilistic inference). As an example where deduction functions as a limiting case for probabilistic reasoning, we note that deductive inference can be generalized to a certain kind of probabilistic inference whereby the probability that a given sentence is true given that all sentences in a given text base are true can be near-instantaneously calculated as the "weighted proportion" of local graphs of the text base which are incompatible with all local graphs of the negation of the given sentence. This same mechanism can be used to show the degree of consistency of the entire text base.
2. A sentence of a natural language is regarded here as a character string to which a specific syntactic and semantic structure, called a "reading," has been assigned, and which varies generally with the context of utterance. A character string, relative to a given reading, has a unique graphical representation which captures those aspects of its meaning that determine its deductive interconnections with other character strings relative to a their readings. We do not address the very difficult problem of developing effective procedures for determining readings of given character strings for given contexts of utterance; however, we have developed, though not included in this paper, effective procedures for obtaining suitable graphical representations of given character strings relative to given readings.
3. A sentence has as many negations as it has readings. When we refer to the negation of a given sentence we are assuming a specific reading of that sentence, and its negation is that sentence whose main relation is interpreted as the set-theoretical complement of the main relation of the given sentence.
4. Regarding the question of whether deduction is too complex or otherwise unsuitable to be treated as pattern matching, the point of this paper is that it can be so treated provided that the representations used are structured to enable it.
5. While the empirical evidence for the proposed mechanism for human IDR is at this point primarily anecdotal, it is still reasonable to speculate regarding the sorts of deductive mechanisms which could be consistent at least with that evidence, and which could serve as starting point for future empirical studies. The virtue of the proposed mechanism is that it is fully explicit and applies to a wide range of sentences.
6. While other sorts of inference mechanisms based on pattern matching have been proposed for inference forms other than deduction, such as abduction or analogy, these appear to be limited to a far narrower range of cases than those which we claim are addressed by the deductive mechanism proposed here.
7. The reason for defining the denotation of a sentence as a singleton-singleton set is that we want to have sentences qualify also as thing-expressions, and the denotations of thing expressions are always sets of sets of elements of the universe of discourse.
8. There are, in general, many possible chain functions on the sequence $(f(a_1), \ldots, f(a_m))$ of thing expressions of $r^m(a_1, \ldots, a_m)$ relative to f, but there is exactly one chain function whose trace is identical with the positive relational profile of $r^m(a_1, \ldots, a_m)$ relative to f if $r^m(a_1, \ldots, a_m)$ is true under f, and there is no such chain function if $r^m(a_1, \ldots, a_m)$ fails to be true under f.

References

[1] Hinton, G. E., J. L. McClelland, and D. E. Rumelhart. (1986) Distributed Representations. In D. E. Rumelhart and J. L. McClelland, Eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Cambridge, MA: MIT Press,

[2] Shastri, L., & Ajjanagadde, V. (1993). From simple association to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, **16**, 417 – 494.

[3] Shastri, Lokendra (1999) Advances in *Shruti* – A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony. *Applied Intelligence*, 11: 79 – 108.
[4] Johnson-Laird, P. N. & Byrne, R. M. J. (1991) *Deduction.* Hillsdale, NJ: Erlbaum
[5] Allwein, G, Barwise, J (1996) *Logical Reasoning with Diagrams*, OUP.
Evans, J. St.B.T, Newstead, S. E., &Byrne, R.M.J. 1993. *The Psychology of Deduction*. Hove, UK: Lawrence Erlbaum Associates Ltd
[6] Gardner, M. , (1958) *Logic Machines and Diagrams*, New York, NY, McGraw-Hill
[7] Hammer, E. M. (1995) *Logic and Visual Information*, CSLI Publications.
[8] Peirce, C. S. (1897-1906) Manuscripts on existential graphs. In *Collected Papers of Charles Sanders Peirce*, edited by Arthur W. Burks, vol. 4, (pp 320-410), Harvard University Press.
[9] Shin, S-J (1994) *The Logical Status of Diagrams.* CUP.
[10] Sowa, J. F. (1984) Conceptual Structures: Information processing in mind and machine.
[11] Hummel, J. E. & Choplin, J. M. (2000) Toward an integrated theory of reflexive and reflective reasoning. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society* (pp.232-237). Mahwah, NJ: Erlbaum.
[12] Tarski, A. 1956. The Concept of Truth in Formalized Languages. In A. Tarski. *Logic, semantics, and metamathematics: Papers from 1923-1928.* Oxford: Oxford University Press
[13] Tripodes, P. G. (Unpublishd Manuscript) A Theory of Readings.