On the Role of Affect in Artificial Intelligence and Music

Author(s): Robert Duisberg

Source: *Perspectives of New Music*, Vol. 23, No. 1 (Autumn - Winter, 1984), pp. 6-35
Published by: Perspectives of New Music
Stable URL: http://www.jstor.org/stable/832899
Accessed: 01/01/2010 19:27

On the Role of Affect in Artificial Intelligence and Music

Robert Duisberg

Le coeur *a ses raisons,*
que la raison ne connait point.
—Blaise Pascal

1. Introduction

1.1 why music in AI?

Artificial Intelligence is a field rich in potential for both science and industry. But the very name of the field, whatever one may mean by it, engenders controversy of a profound and metaphysical sort, often associated in philosophical circles with the Mind-Body Problem or the Physicalist/Spiritualist debate (Searle 1980; Dennet 1978). The problem apparently revolves around the personal character we associate with our conscious experience of our own minds, and certain apprehensions we have regarding the embodiment of such things, again whatever they may be, in physical structures other than organic ones like ourselves. The intent here is not to debate the Mind-Body Problem, nor to discuss the locus of the soul. It is rather to consider the implications of taking seriously the apparent goal of AI, to construct the closest thing we can to an artificially intelligent machine.

Many useful insights into how a mind could be modeled in AI may be gained from a knowledge of our own minds, and thus AI and Cognitive Psychology interact increasingly. Most of the notable work in the field has, reasonably, concentrated on the aspects of thought amenable to logical analysis, subject to a clean syntax.

However, it is clear that much of our conscious intellectual life, indeed its fundamental motivations, are based in urges, drives and emotions that are less appropriately considered under such a logical regimen: "only the surface of reason is rational" (Minsky 1981, 28). The very act of knowing involves the ordering of sensory input according to conceptual frameworks, but the development and application of appropriate concept networks is necessarily a creative and intuitive act. "Aesthetic sensibility plays the part of the delicate sieve" (Wechsler 1978, 1) in appraising the efficacy of such application.

It is an aberration of recent western culture that has so polarized the concepts of Science versus Art, rational versus aesthetic, so as to render them mutually incommensurable, when they are clearly and necessarily interacting parts of a complex whole.

> . . . The prevailing conception of science, based on the disjunction of subjectivity and objectivity, seeks—and must seek at all costs—to eliminate from science such passionate, personal, human appraisals of theories, or at least to minimize their function . . . [But] the act of knowing includes an appraisal; and this personal coefficient, which shapes all factual knowledge, bridges in doing so the disjunction between subjectivity and objectivity (Polanyi 1958, 15–17).

An epistemology which incorporates such passionate appraisal into the heart of factual knowledge suggests itself as an important feature of any realistic mind model that would avoid a specious rational/intuitive dichotomy.

As artists and scientists, if we honestly profess to study the mind, we must try to understand the structure of feeling, for it is only by and with such feelings as wonder and love that we grow to create within our own individual mental lives the world as we interpret it, and it is through such structures that we apprehend our physical and social surroundings.

First one must gain an understanding of how concepts come to be formed in the mind, and model how sensory patterns come to be ordered into structured concepts.

> First, facts are not data. They are mental artifacts, selected by human concerns and abstracted from experience by filtering through a screen of schemata. Second, this screen is necessarily tacit; we infer its nature only from observing its operations, but our inferences can never be complete or up to date. Third, the screen is itself a product of the process which it mediates and, though tacit, can be developed by deliberately exposing it to what we want to influence it. (This is the essence of education.) (Wechsler 1978, 152)

Now music is a fertile field for such modeling, for to understand any piece of music involves recognizing its salient structures, always in its own terms, in order to appreciate how they are manipulated. Minsky likens a sonata to a "teaching machine" in which material is presented in the exposition and its meanings subsequently explored (indeed established) by displaying the material in varied forms, contexts, and elaborations. The suggestion is clear that if one could model how a listener decides what is important or salient or interesting in a piece of music, where there is no predetermined "vocabulary," then one might have a reasonable model of how minds go about apprehending significant structure in superficially chaotic environments. He suggests the possibility of "future simulations that grow artificial musical semantic networks, perhaps by 'raising' simulated infants in traditional musical cultures" (Minsky 1981, 35).

Furthermore music seems to span an interesting middle ground, highly structured formally and numerically on the one hand and indisputably evocative of emotion on the other. Its structure lends it to quantitative study and yet its semantic content is reflective of our emotional intuitions. An effort to understand the deep and semantic structure of music, the structure of the emotional meanings so precisely expressible in music, may pay off in a better understanding of the mind with respect to its motivational underpinnings in feeling. This broad issue will be discussed in the context of the central theme of this paper, that affect may in some useful sense be considered the *driver* of concept formation. This intuition is supported by Papert's suggestion of "an applied genetic epistemology beyond Piaget's cognitive emphasis to include a concern with the affective" in child development (Papert 1980, vii). Higher level concept formation may also be so conceived in view of numerous corroborative statements by the great mathematicians and physicists of being led to their theories by aesthetic intuitions and emotions, the rigorous proofs only coming later (Hardy 1967; Poincaré 1963; Pais 1983). "I maintain that attitudes do really precede propositions, feelings come before facts. This seems strange only because we cannot remember what we knew in infancy" (Minsky 1979, 5).

1.2 EXPRESSION AND DENOTATION

Music communicates meaningfully, without question. But there are at least two modes in which such communication is commonly appreciated, which may as well be called Apollonian and Dionysian. This is again, perhaps, a needless dichotomy but it is historically manifest in abundant aesthetic debates such as that of Formal versus Program Music or Brahms versus Wagner (Hanslick 1957), Classicism versus Romanticism, etc. The first mode has to do with the recognition of structure, and is akin in this respect to the apprehension of mathematical beauty.

> We bear in mind the feeling of mathematical beauty, of the harmony of numbers and forms and of geometric elegance. It's a real aesthetic feeling that all mathematicians recognize, and this is truly sensibility . . . The useful combinations are precisely the most beautiful . . . (Poincaré 1958, 15).

The forms and objects that partake in the structures so perceived in music are essentially syntactic in character. In this mode it therefore makes fair sense to speak of "Form without Content" or "Symbol without Referent," and we are motivated to bring to bear all the technology of phrase structure, generative and systemic grammars used in parsing of the structured languages used in computing (Roads 1979; Lerdahl 1983). This mode of appreciation lends itself to formal and computational models of which Terry Winograd's work discussed in the next section is a classic example.

But the Apollonian mode cannot be the whole story any more than mere syntax can be adequate to encompass natural language. Syntax tells little about meaning.

The second mode of appreciation requires consideration of music not so much as one would consider a mathematical theorem but more as one would consider expressions in a kind of language. Musical events, be they rhythmic, harmonic, melodic, or whatever, can scarcely be said to denote in the way words do in a verbal language. But musical symbols may properly be said to *express* things other than themselves, and thus it becomes important to clarify the distinction between denotation and expression.

The logical structure of denotation is relatively straightforward. Computationally the denotation of a referent by a symbol may be thought of as a pointer from the symbol to the referent. Thus words have pointers to their denotata and text can be "understood" by performing table-lookups of definitions. Furthermore such association is typically established through convention as in verbal language.

Now consider a slow rubato passage in the minor mode played by 'cello solo expressing great sadness. This description gives information of two kinds, saying something about the properties the passage possesses as well as what feelings it expresses. The first kind is clear; the passage is a concrete instance of slow 'cello sound in minor mode and free rhythm. But instead of saying that the melody expresses sadness one might say simply it is a sad melody. Is it sad, then, in the same way that it is slow? Whereas the passage literally possesses slowness and it literally belongs to the class of 'cello sounds, only figuratively does it possess sadness or belong to the class of sad things.

> What is expressed is metaphorically exemplified. What expresses sadness is metaphorically sad. And what is metaphorically sad is actually but not literally sad, i.e., comes under a transferred application of some label coextensive with "sad." Thus what is expressed is possessed (Goodman 1976, 85).

If expression is seen to be the possession of a set of attributes, then determination of what is expressed is like a recognition problem achieved by measuring a match or fit between the strictly musical properties (i.e., pitch structure, rhythmic structure, etc.) of the expressive object and the properties of the emotion which the object possesses and expresses.

1.3 AFFECTIVE MEANING, MEMORY AND HEURISTICS

This suggests that cognitive appreciation of musical semantics requires *a priori* knowledge of the emotional world, that is, a known set of expressible emotional properties, in the same way that understanding stories and text in natural language requires knowledge of the world in the form of scripts or some other template (Schank 1978). Of the Yale centered group doing research in natural language processing, Wendy Lehnert's work on narrative comprehension and summarization succeeds remarkably well by being driven by a system of "affect state" relations. The emotional states and motivations of characters are the structural nodes in a network which comes to represent the meaning of the text. From her striking empirical demonstration of how cleanly the concepts imparted by a narrative text can be organized by simple affect structures it is a short step to envisioning such affective structures as actual generators of concepts. Her affect-driven knowledge representation is highly suggestive and could be extended to an application to music.

The notion that the meaning of a narrative can be embodied in a graphic network of connected affect nodes, and that useful summarization and question answering can be achieved by means of traversal of such a graph is similar to Minsky's idea that meaning arises out of a kind of network of associations. "Something has a 'meaning' only when it has a few; if we understood something just one way we would not understand it at all. That is why the seekers of the 'real' meanings never find them" (Minsky 1981, 29). Minsky is one of very few authors willing to address the question directly, and comes to describe the networked character of a mental state by the metaphor of a "society" of mental "agents," with memories described as partial traces of the network (K-lines) (Minsky 1979). Agents are seen to exist in a hierarchy being driven by requests from above and responses from below. *"The recognition of what events should be considered "memorable" . . . must usually be made by some other agency that has engaged the present one for a purpose"* (Minsky's italics). The "other agency" could be a "feeling agent" expressing our disposition toward a given configuration of the network, that is, whether a situation is worth remembering.

Affect would function, then, as a system of heuristics regarding the appropriateness of executing storage and abstraction operations. In a graphic and geometric visualization, one might think of an affect state as a temporal and "spatial" pattern (where "spatial" dimensions may correspond to appropriate parameters such as "tension," "activity," "direction," or "weight") whose form could be recognized, in the same way that the emotional referents expressed by music might be recognized. This structure corresponds in some clear way to the configuration of the network representing the overall mental state, and in a crude first model could be derived by some simple transform. The abstraction involved in such a transform arises almost automatically out of Minsky's "Level-band Principle" and is related to the function of metaphorical transference as described in the process of musical expression.

Once abstracted from the network representing the mental state and recognized, the feeling state, functioning like a "hunch" or a "gut feeling," figures prominently into the preconditions of any heuristic to determine appropriate subsequent action, in particular actions pertaining to laying down memory traces or forming new concepts. So it seems natural in conclusion to consider how such "feeling heuristics" apply to machine learning and concept formation. Doug Lenat's work on the nature of heuristics (Lenat 1983a; Lenat 1983b) yields remarkably successful learning behavior as demonstrated by the undeniably creative concept formation performed by his EURISKO and AM programs. His work will be discussed in the final section of this paper with an eye to what insights might be gained from acknowledging the prominence of emotions in guiding human heuristic behavior and how structural modeling of emotions, guided by what we learn from musical structure, relates to his "heuristics about heuristics."

2. Winograd: Syntactic and Semantic Interaction

Terry Winograd's program to analyze tonal harmony has the virtue of restricting its domain to a tractable subproblem within music theory while demonstrating the power of a methodology with much broader applicability. Indeed, much of a vastly more complete and detailed current theory of music is acknowledged to have analogs in such methodology (Lerdahl and Jackendoff 1983).

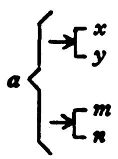
2.1 A terse outline of systemic grammar

The type of grammar chosen to represent and analyze harmony is known as *systemic* and is apparently not common in computer science applications. Systemic grammar differs from phrase-structure and transformational grammar both in the history of its development and its intent, which is reflected in its formal structure. It is an outgrowth of concerns in anthropology and sociology (Halliday 1981; Halliday 1978; Kress 1976; deJoia and Stenton 1980) rather than mathematics or formal logic. Its motivations lie in consideration of language as a social activity, to which concerns about the nature of grammaticality are adjunct.

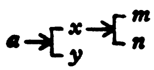
Systemic grammar is based on the notion of choice embodied in the concept of a *system*, which is a set of choices of *features* together with a condition of entry, and is notated graphically in two dimensions by a *system network* (see figure 2-1). Any sentence or item in the language may be described by the set of choices that have been selected in its realization. A set of *realization rules* is associated with each feature choice, pertaining to what functions and categories must be present when a given feature is chosen as well as requirements of agreement and ordering of constituents. These rules are less precise and formal than the generative rules of immediate constituent and transformational grammars, being more of the character of consistency checks than actual effective production rules. The features chosen are said to be realized as structures represented by the linear sequence of constituents.

> Since systemic grammar is not centered on a concern with formal rules, the general attitude is that it is better to say something less precise about an important aspect of language than to ignore it completely because it does not yield to the available formal tools. It is possible to provide descriptions that are structured (i.e., they include formal representations like system networks, not just descriptive text) but that are not generative in the strong sense of providing rigorous rules. Much of systemic grammar follows this course (Winograd 1983, 277–78).

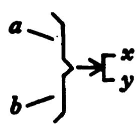
there is a system *x*/*y* with entry condition *a* (if *a*, then either *x* or *y*)



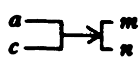
there are two simultaneous systems *x*/*y* and *m*/*n*, both having entry condition *a* (if *a*, then both either *x* or *y* and, independently, either *m* or *n*)



there are two systems *x*/*y* and *m*/*n*, ordered in dependence so that *m*/*n* has entry condition *x* and *x*/*y* has entry condition *a* (if *a* then either *x* or *y*, and if *x*, then either *m* or *n*)



there is a system *x*/*y* with compound entry condition, conjunction of *a* and *b* (if both *a* and *b* then either *x* or *y*)



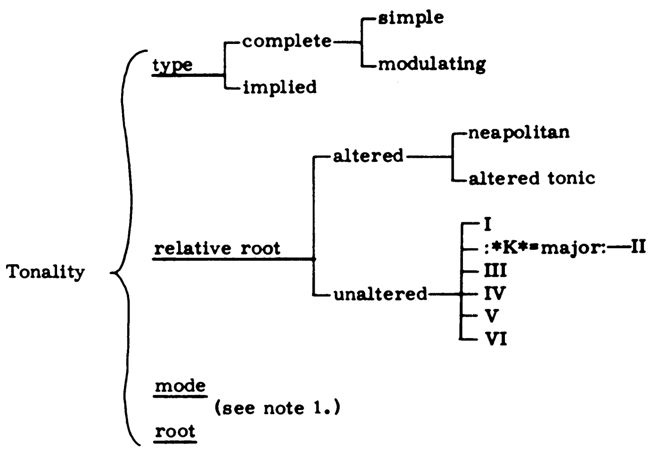
there is a system *m*/*n* with two possible entry conditions, union of *a* and *c* (if either *a* or *c*, or both, then either *m* or *n*)

FIGURE 2-1: NOTATIONAL CONVENTIONS IN A SYSTEM NETWORK (KRESS 1976)

The power of systemic grammar lies in its recognition that the form of a sentence may be the result of choices made within several systems operating simultaneously in orthogonal dimensions, as contrasted to the usual generative grammar in which a constituent is associated with exactly one path in the derivation tree. (Even though in a transformational grammar the role of a constituent may change under transformation, still at any point in the derivation all relationships are taken to be inferable from the single constituent tree structure at that point.) The advantage of this type of grammar is in the close connection between the syntax and semantics, as some choices along certain dimensions can correspond closely to the meaning of the sentence.

2.2 A SYSTEMIC GRAMMAR FOR TONAL HARMONY

The grammar consists of five ranks of system networks: Composition, Tonality, Chord Group, Chord, and Note. The system networks and realization rules for Tonality and Chord are reproduced in figures 2-2 and 2-3 by way of



Note 1. The choice of features in the mode and root systems are dependent in a complex way on the choice of relative root and on *K*. This dependence is most easily expressed in the form of a table, Table I, but the status in the grammar is the same as dependencies indicated with lines and brackets.

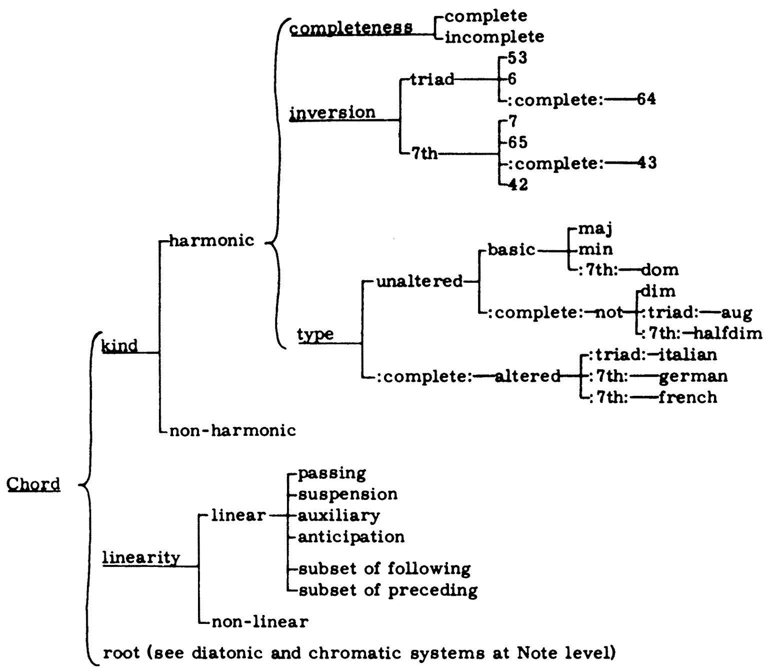
Realizations:

Tonality : : +dominant; +(sec). . .(sec); (sec)⁀dominant
complete : : +tonic; dominant⁀tonic
simple : : tonic⁀‡
modulating : : +sec_2⁀(sec_2)⁀. . .⁀(sec_2); tonic⁀sec_2
implied : : +domprep; domprep dominant; dominant ≡Chord Group direct
relative root, mode, root : : The realizations of the relative root system act through the connections with the mode and root systems to produce *K* as indicated in Table I.

Constituents:

dominant : : $\text{Tonality}_{\text{simple, V}}$ or $\text{Chord Group}_{\text{V or VII}}$
sec : : Tonality or Chord Group
tonic : : $\text{Chord Group}_{\text{I, direct}}$
sec_2 : : $\text{Tonality}_{\text{simple or implied}}$ or Chord Group
domprep : : $\text{Chord Group}_{\text{II or IV or VI}}$

FIGURE 2-2: SYSTEM NETWORK AND REALIZATION RULES FOR A TONALITY (WINOGRAD 1968)



Realizations:

completeness, inversion, type : : These three systems are realized jointly in the specification of notes which serve in the constituent structure. Completeness and inversion function together to determine the diatonic values of these notes. The realization is a list of diatonic intervals above the bass, where each note must form one of these intervals, and there must be at least one representative for each interval. This pair of systems also specifies which note has the subscript root, by specifying its interval above the bass. The details of these facts are included in Table III. All three systems jointly determine the set of chromatic intervals in the chord in the same way. The facts are included in Table IV, where empty entries appear wherever the particular combination of features is precluded by the system network.

non-harmonic : : This is a default term, realized by the presence of a structure which cannot be produced by any combination of features depending on harmonic.

linear : : This is a context-sensitive realization, which limits the notes according to the notes of the following (FOL) and preceding (PREC) chords. Details are in Table V.

non-linear : : A default term.

root : : The selection of features is passed down to the constituent $\mathrm{Note}_{\mathrm{root}}$ as a restriction.

Constituents:

N : : Note

FIGURE 2-3: SYSTEM NETWORK AND REALIZATION RULES FOR A CHORD (WINOGRAD 1968) example. Within the grammar though, it is clear that the structures so described are ambiguous at every level, and it is this that makes the problem interesting. For example a chord sequence vi $\rightarrow$ V^6 $\rightarrow$ I in C-major might also be ii $\rightarrow$ I^6 $\rightarrow$ IV in G-major or I $\rightarrow$ III in a-minor with a passing lowered-VII^6. The only way to resolve such multiplicities of interpretation is in the context of the entire piece without which the full relative function of the chords cannot be known. A music theorist in rendering a "reading" of a composition that he feels expresses its basic logic is guided by considerations of clarity and economy of the parse (where an economic parsing would have a minimum average depth of nesting of secondary tonalities throughout the composition). For the purposes of this implementation, Winograd combined all such considerations into a very rough numerical value associated with a given parsing he calls a "plausibility value." An integer value is assigned to each possible sequence of two adjacent harmonies according to the following arbitrary scheme.

0: authentic cadence, opening progression, dominant preparation, dominant substitution

1: fifth progression, plagal cadence, deceptive cadence, triadic outline

3: for any chord interpreted as a linear or passing chord

5: confusion of tonal structure, that is, anything other than V or IV going to I

4: anything not covered above.

The rules are ordered so that if anything higher in the list applies the others are ignored. Clearly the lower values imply a more "plausible" parse and the intuition is that the progressions associated with values 0 and 1 are the progressions with clear structural functions that tend to establish a tonal center.

As the parsing proceeds, a separate "path" is constructed for each of the different possible parsings of the string encountered up to that point. That is, in the spirit of the systemic grammar a number of different tonal systems are considered in parallel. A disadvantage is that the number of paths considered can grow exponentially in the length of the string. In order to limit this growth, advantage is taken of the fact that modulatory progressions are rare compared to complete cadences, so the string is scanned in reverse order from right to left, and preparations are preferentially considered in terms of where they arrive. Further reductions are achieved by eliminating parse trees that are redundant in the sense that they arrive at the same analysis via different paths, and trees which seem to be "doing badly" in that they are coming up with plausibility values that are too high.

Overall the program implementing this grammar and parsing strategy gives sophisticated "readings" such as one might expect in an undergraduate harmony class. It is admittedly limited in being able to deal only with a direct sequence of chords as the input, and thus is applicable to music with the simplest rhythmic structure, and overlooks the structural effects of melodic and contrapuntal ideas, without reference to which harmonic structure can sometimes not be adequately explained. Moreover, little effort was spent on dealing with nonharmonic or linear chords. But these shortcomings are the result of deliberate restriction of the problem, and are not structural shortcomings of the methodology. The structures resulting from realizations in a systemic grammar need not be one-dimensional but may have several different dimensions, as the result of independently-operating systems. Also the resulting structure need not be segmented in one way but may have several constituent structures, even within the same dimension, and music abounds with such nonexclusive groupings, as in melodic groupings, phrase groupings, harmonic groupings, etc. There appears to be no limitation in principle why the systemic approach could not be extended to embrace such additional dimensions.

I would take issue with but one point. Winograd describes his system as being driven by "semantic heuristics," assigning a level of "meaningfulness" to any parsing, and thus it is able to choose the "best" parse of all considered. When it is revealed that the heart of the system's "semantics" is his ad hoc plausibility value, it appears to be a curious use of the term, if it is taken to mean "the study dealing with the relations between signs and what they refer to" (Webster's 3rd Int'l, 1961). These values associated with pairs of chords reflect the role of the progression in the context of the string of chords, but it is essentially a syntactic role, and has no bearing on what the chord string may *refer* to extramusically. Be it granted that the function of a chord is a higher-level syntactic category reflecting a larger context than its local lexical identification (spelling), still it confuses the issue to suggest that this is the meaning of the chord; one must look further for that.

Though beyond the scope of this paper, such a further look should lead to a "grammar of the emotions," and it seems worth digressing briefly to mention a detailed "Structural Theory of the Emotions," by Joseph deRivera. The theory postulates a "matrix of emotions" (see figure 2-4) describing "the particular emotions as movements in a three dimensional interpersonal space," which "suggests that any particular emotion is the outcome of a pattern of 'choices' that organize our relationship with the other" (deRivera 1977, 71). The structural similarity between this description and the form of systemic grammar is intriguing and suggests the plausible application of such a grammar to the "parsing" of emotional states based on "interpersonal motion." The theory describes emotions primarily in terms of their relations to each other and to externally observable interpersonal movement, rather than the internal structure and form of the emotion itself, so it is not directly applicable to modelling of subjective affect states. But an ability to parse a man-machine interaction into an appropriate affect state could lead to more sensible or to at least "friendly" machine responses.

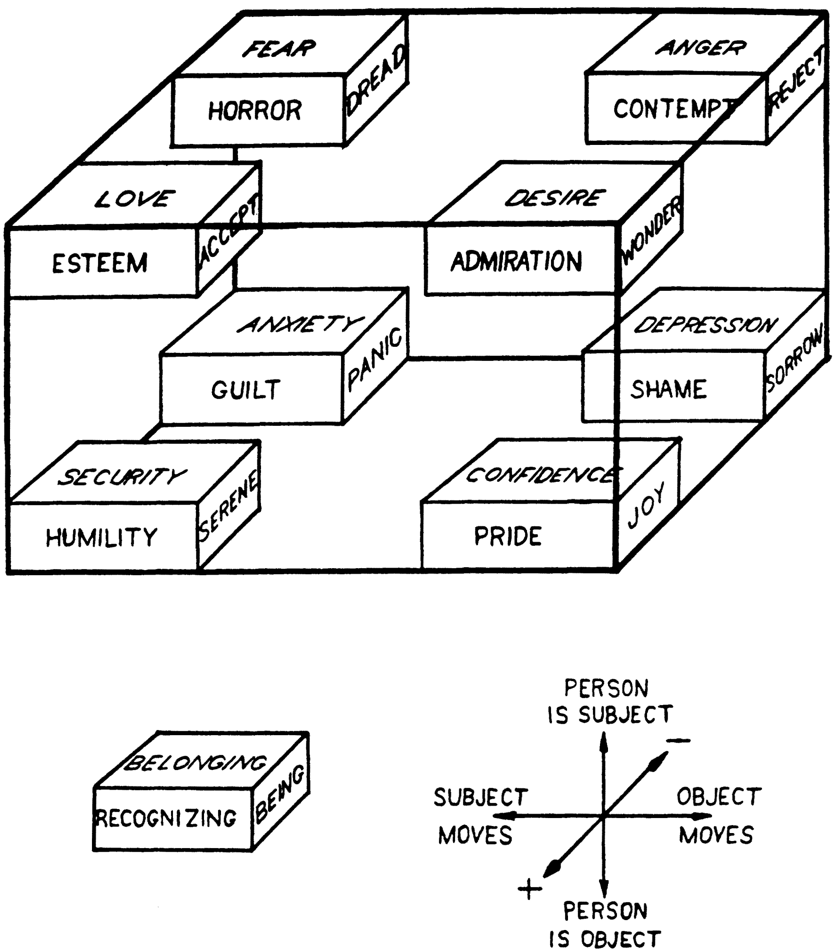


FIGURE 2-4: THE MATRIX OF EMOTIONS (DE RIVERA 1977)

3. Lehnert: Affect State Patterns

Wendy Lehnert's approach to the problem of comprehension of an utterance is typical of the "Yale School" of natural language processing in that there is little concern for the actual surface form of the utterance, the emphasis being instead on the nature of the structures created in memory in response to the text. The paradigm for exploring such memory representations is to ask the reader to summarize the narrative text, and the summarization behavior is taken to reveal what sorts of inferences, causal-chains, and integration of information into knowledge structures has occurred.

> When a reader is asked to summarize a story, vast amounts of information within the memory representation are selectively ignored in order to produce a distilled version of the original narrative. This process of simplification relies on a global structuring of memory that allows search procedures to concentrate on central elements of the story while ignoring peripheral details (Lehnert 1982, 376).

In her system the mechanism of such global structuring is built out of remarkably simple affect states which combine to form *plot units* of greater complexity. In fact, affect states of only three kinds are posited:

- events that please (notated + in diagrams)
- events that displease (–)
- mental states with neutral affect (M).

The neutral M-states are usually associated with intentions or goals. Patterns are in turn comprised of a number of such primitive states connected by arcs or "links." Links are of various types. Links between affect states involving a single character are called *causal links* since they relate to causes of mental states and intentionalities behind events. There are four labeled kinds of causal links:

- Motivational links (m) from an affective event to the mental state it causes
- Actualization links (a) from an intentional mental state to a resultant affective event
- Termination links (t) from one event or mental state back to another whose affective impact has been thereby supplanted or displaced
- Equivalence links (e) between distinct affective responses to the same event or mental state (i.e., resulting from a change in perspective).

Not all 36 pair-wise configurations are possible but legal configurations are constrained as follows:

- m-links must point to a mental state
- a-links must point from a mental state to an event
- t- and e-links must point:
  - from a mental state to a mental state, or
  - from an event to an event.

3.1 PLOT UNITS

The fifteen legal pair-wise combinations form the set of *primitive plot units* and each is referred to by name as in figure 3-1.

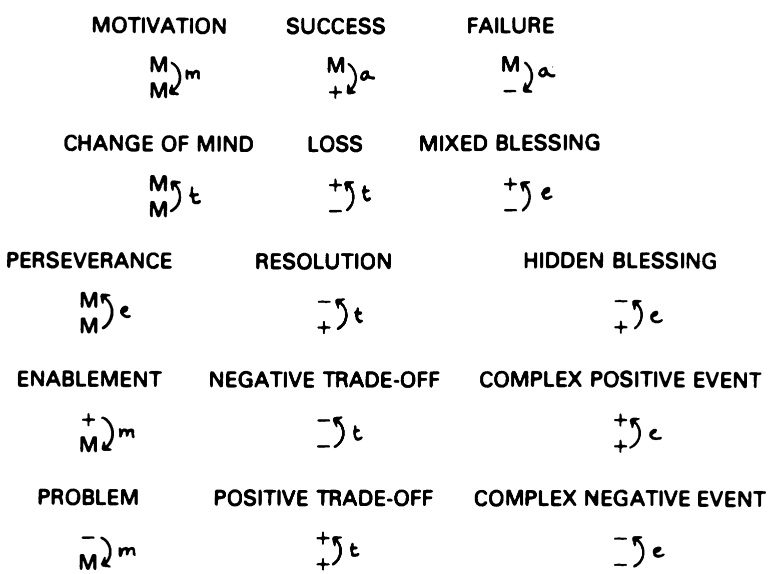


FIGURE 3-1: PRIMITIVE PLOT UNITS (LEHNERT 1982)

The graphic convention is that states involving a single character are aligned vertically and that time flows down the page. These primitive plot units can be combined into complex plot units such as in figure 3-2.

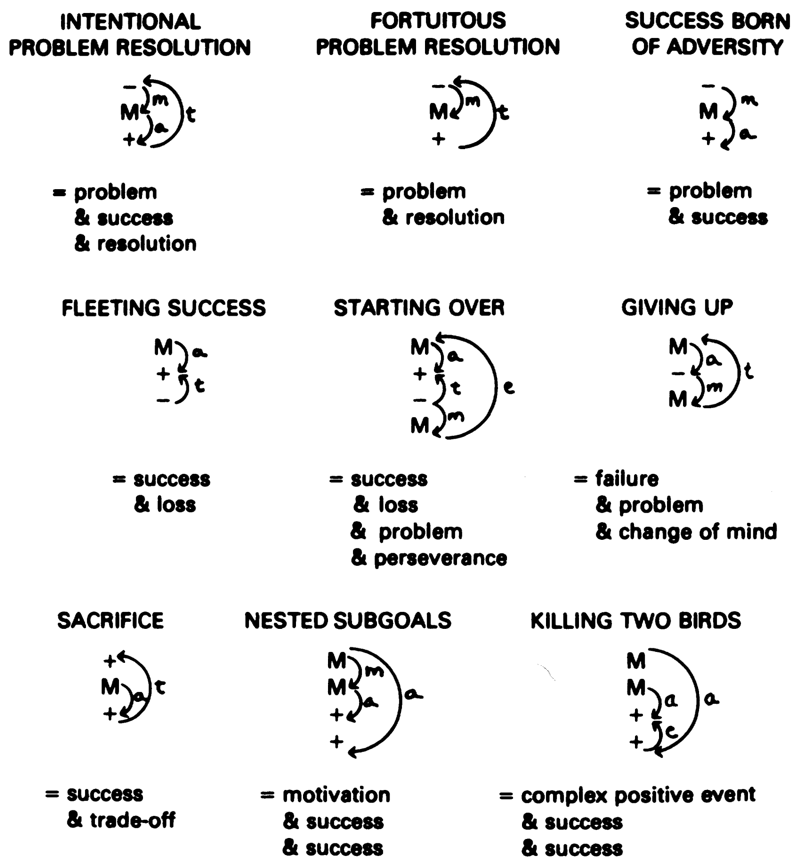


FIGURE 3-2: SOME COMPLEX PLOT UNITS (LEHNERT 1982)

In addition, narratives involving more than one character involve cross-character links, notated with diagonal segments between affect states where the higher state precedes the lower in time. Such links can occur between any pair of states and their interpretation is as in figure 3-3.

With these relatively simple elements then, many configurations of great complexity and subtlety arise, of which just a few are shown in figure 3-4.

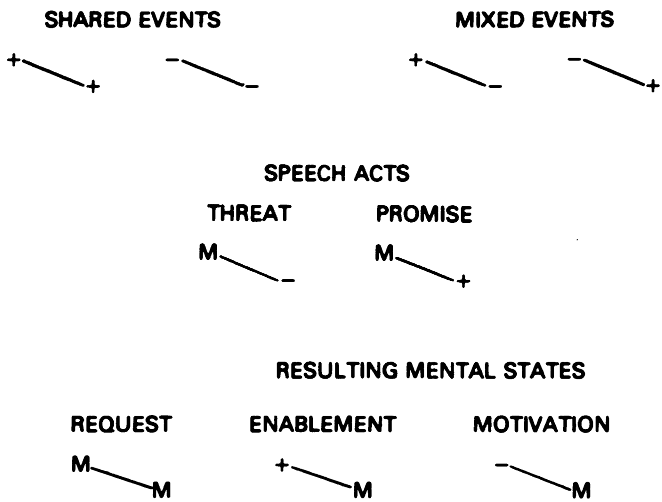


FIGURE 3-3: SOME CROSS-CHARACTER LINKS (LEHNERT 1982)

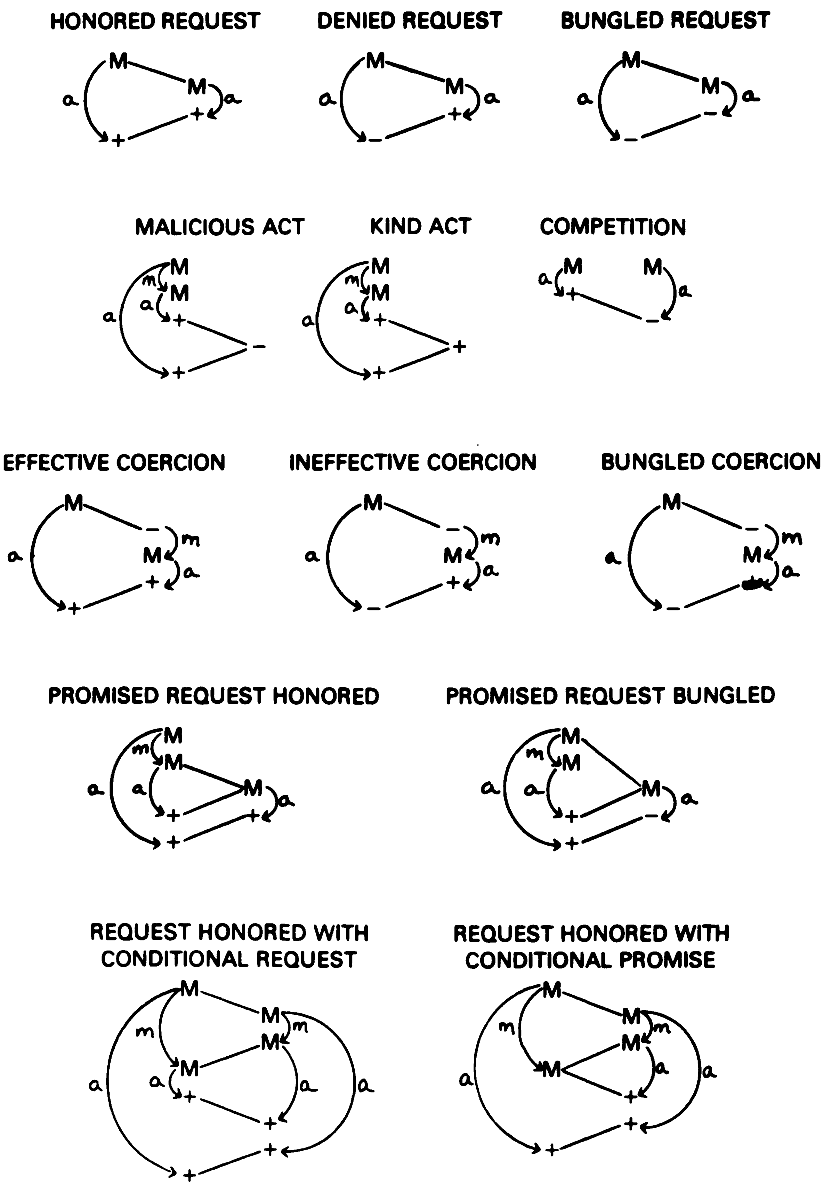


FIGURE 3-4: SOME COMPLEX PLOT UNITS INVOLVING TWO CHARACTERS (LEHNERT 1982)

3.2 CONNECTIVITY AND NARRATIVE COHESION

A set of precise definitions are given concerning graphical connectivity of plot units, among them:

- Two plot units A and B are *related* iff they share a common affect state.
- A is *connected* to B iff
  1. A = B.
  2. A is related to B.
  3. There exists a sequence of intervening plot units $U_1, \ldots, U_n$ such that A is related to U_1, U_i is related to U_{i+1}, and U_n is related to B.
- A *family* around A is the set of plot units related to A.
- F is a maximal family in K iff F is a family contained in K and the number of units in F is greater than or equal to the number in G for all families G contained in K.
- A is a *pivotal unit* in K iff the family around A is maximal in K.

It is noted that stories that yield highly-connected plot-unit graphs correspond naturally to highly cohesive narratives, and it is claimed that narrative cohesion is an important factor for effective memory retention, and furthermore, the information expressed in a pivotal unit tends to be a central part of a general summary.

3.3 SUMMARIZATION

A summary is then generated by the following algorithm:

1. Extract all top-level plot units from narrative text. (This issue is not central to her current analysis, and is handled by mechanisms employed in an earlier story comprehension system, BORIS.)
2. Derive the plot unit graph structure.
3. Identify the pivotal unit P.
4. Generate a *base-line summary* from a frame for P.
5. Augment base-line summary with information from plot units related to P.

Step 4 is accomplished by associating with each plot unit a frame structure from which a natural language description of the content of the unit can be generated. For example, a general frame for a Retaliation unit might read: "Because X's [?] caused a [–] for Y, Y tried to [+] to cause a [–] for X."

The fairly elaborate formalism outlined here becomes convincing in view of abundant examples provided and empirical comparison of the kind of summarizations that people construct in response to a given story to those derived via the system.

Some of the limits of such analysis are clear and acknowledged. Since affect states reflect information about personal goals, plans and intentions, they are not applicable to texts that contain no information along these lines, for example, expository or purely descriptive writing. Still the most striking feature of the analysis is how even such an extremely coarse modeling of affect can so effectively organize information of this kind.

In thinking of an application of this paradigm to music the kind of summary one typically finds in concert program notes immediately comes to mind. The attempt to write a "music appreciation program" that would try to write such summaries in response to musical input is likely to yield insight into possible cognitive structures involved in music listening, in the same way that narrative summarization suggests strategies to structure stories. To pursue the analogy a little further one could imagine, corresponding to affect states or events, musical events, mapped probably not onto a single affective dimension (–, 0, +) but several. For example, within a particular musical style, descending minor appoggiaturas might map onto one end, and major-scale passage work onto the other end of a "happy/sad" dimension, presto/pizzicato and lento/legato onto two ends of an "activity" dimension, piccolo with string-harmonics and triangle on one end, and tuba with trombones and timpani one the other of a "weight" dimension, etc. One might then imagine primitive "plot units" like "Imitation" (with several distinct subtypes, i.e., direct, sequential, altered, inverted, etc.) or "Juxtaposition" (which might include contrasting and similar) or "Transformation" (including rhythmic and melodic augmentation and diminution, modulation, extension, dissection, concatenation, etc.) and others. Such smaller units then would fit together into sections and finally the standard large-scale musical forms.

4. Minsky: Memory, Metaphor and Meaning

Minsky's K-line theory offers an extremely simple graphic model which, in spite of its simplicity, nonetheless yields neat, almost procedural descriptions of such fuzzy phenomena as analogy and metaphor.

4.1 A BRIEF SUMMARY OF THE K-THEORY

A state of mind is envisioned as a great hierarchical lattice of *agents*, the partially autonomous individuals in a "society of mind," each embodying an active idea or concept. Information moves upward in the lattice so that any given agent receives input from a roughly pyramidal hierarchy of agents beneath him. Most agents are grouped into small "cross-exclusion" arrangements, each having inhibitory connections to others in the group so that only one in the group may be active at a time, and further resulting in a built-in short-term memory. This is rather like object-oriented data abstraction taken to its logical conclusion.

When a memory is formed a *K-line* is created for it, which is thought of as a "wire" with potential connections to any agent, and a *K-node* at the end whose activation in turn activates the connected agents. Memorizing a particular state of mind creates excitatory attachments to agents currently active in that state. The crucial issue is that actually only a subset of currently active agents can really afford to be reactivated in evocation of a memory. "It is *not* the goal of Memory to produce a perfect hallucination . . . *A memory should induce a state through which we see current reality as an instance of the remembered event*" (Minsky 1979, 8). To obtain the desired metaphorical function he proposes his *Level-Band Principle*. Simply stated, it says that when an agent A forms a memory, the K-line it creates should only attach to active agents in the current state of mind which lie "within an intermediate band of levels" somewhere in the hierarchy beneath A, not too high and not too low. This is so that in evocation, current goals which are likely to be represented by agents higher up will not be overwritten by the forced excitation of the activated K-line, and the particular contingencies of the current situation will not be lost by forcing the states of the lower-level perceptual agents.

The theory is elaborated by proposing the recursive notion of connecting K-lines to other K-nodes as well in accordance with the Level-Band Principle, which is just to say that new memories are clearly built on earlier remembrances. Graphically then K-nodes grow into a structure with connections echoing those in the agent hierarchy, except that information flows the other way: agents pass their information to their superiors, but K-nodes activate other K-nodes and agents below them.

It's interesting to see how abstraction happens almost automatically in such a network. Consider an agent encountering a problem in the current situation. If the agent activates a memory of a similar situation, but the mind-state of this memory differs in some detail from some current situation, then the agencies involved in the conflicting details will tend to become inactive because of the cross-exclusion connections. Thus the common, nonconflicting properties are automatically extracted, and this is the essence of abstraction and metaphor. If the resulting state of mind leads ultimately to resolution of the problem, this combination of earlier K-node and current activations could be remembered as a new K-node representing a new generalized concept.

4.2 LEARNING

Minsky is vague on the crucial subject of how learning is accomplished in his model except that he postulates the operation of a third agency network which controls the formation and linkage of K-nodes to other K-nodes and agents. He is forced to make this assumption, which is dangerously suggestive of infinite recursion (What agent controls the agent which controls . . . the formation of K-nodes?), since he rejects the behaviorist reinforcement theory of learning as being too simplistic. Reinforcement theories rely on "recency" to determine which strategies should be remembered when a goal is finally successfully achieved. This is considered inadequate since longer-range strategies are often involved and one would also like to forget more recent blind alleys and failures. Thus, he concludes that different scales of tactics, strategies and goals are segregated and kept track of in a separate web of agencies.

In order to break the recursion an agent at some level up in the control hierarchy will have to have enough information available to make a judgment on the "relevance" of active agents under its control. This judgment must be intuitive in character, as such judgments are in humans, that is, based on incomplete information and by means of metaphorical associations with other memories. Clearly we tend to "notice" percepts and states of mind associated with a strong affective response, and the link between emotion and memory is certainly profound; memories tend to carry with them their own emotional tone, and may be evoked by a similar emotion in the present. Thus an extension to the theory may be to think of K-lines as something more complex than a mere excitatory wire. Rather, think of a three-dimensional K-matrix, where entries correspond to agents in the intermediate Level-Band which may be active or not. Further, the entries may not necessarily be simple black or white, inactive or active, but may be "colored pixels" encoding a little more of the agent's state or "mood." The spatial form of this pattern may match, in ways previously described, a recognizable emotion. In this way one might imagine the control agent capable of recognizing an active agent configuration in some level-band under its jurisdiction, which corresponds to a strong emotion, which would then trigger the creation of a K-matrix for that configuration. The form and formation of that memory is then intimately connected to its emotional associations.

5. Lenat: Heuristics

Among the most interesting attempts to "raise an intelligence" is the EURISKO heuristic learning program, whose education now spans tens of thousands of hours of CPU time on a Xerox 1100. Types of learning are considered to lie on a spectrum from rote memorization in which the learner is relatively passive, through concept formation, theory formation and discovery, in which the learner is progressively more autonomous and responsible for the resultant representation of the knowledge. Doug Lenat's concern is for the extreme end of this latter region, motivated by the observation that knowledge acquisition from human experts is the bottleneck in building intelligent expert systems today, and "the neck of the bottle is narrow indeed for those fields in which there is as yet no human expert." And furthermore, "the world is too complex to be modelled deeply in any formal way, but a dynamically-growing body of heuristics might suffice" (Lenat 1983a, 36–37).

He proposes then an "Accretion Model of Theory Formation" in which a body of heuristics concerning a specific domain grows upon itself through observation and experimentation in the domain. The model is described by a set of not entirely sequential steps.

1. Given a set of definitions, objects, operations, and rules, gather data about them: find examples of definitions, apply rules and operations, etc.
2. As this proceeds look for patterns in the data gathered.
3. Build hypotheses upon patterns observed, and design and execute experiments to test these.
4. Make new definitions as the body of hypotheses develops, to make the statements of the most useful conjectures more concise, and go back to step 1.
5. From time to time, as the above loop proceeds, create some new heuristics, "by compiling the learner's hindsight."
6. Even more rarely, augment or shift the representation in which the domain knowledge is encoded, as necessary.

The pervading philosophy is that a dynamic collection of informal heuristics is sufficient to drive and direct the model, *even steps 5 and 6*!

Many assumptions are implicit in the model. The first step assumes that the learner is able to make direct observations, which immediately limits the accessible domains, essentially to the fields of mathematics, games and programming. The second rule implies the availability of a large store of known patterns to recognize, and the programs are provided with sets of general domain-independent low-level pattern-noticing rules. And it is emphasized that it is only a working assumption that such rules are indeed domain independent. The third step requires some rules in the initial body of heuristics that will allow the program to generalize meaningfully from the observed patterns, followed by specialization to new specific questions and test cases, and "deeply embedded into this point is a set of metaphysical assumptions about the world: most phomena [*sic*] should be explainable by a small set of simple laws or regularities, knowledge comes from rational inquiry, causality is inviolable, coincidences have meaning, etc." (Lenat 1983a, 38).

The fifth and sixth steps constitute the major design idea that distinguishes EURISKO from the more limited model used in AM. They assume that both heuristics and aspects of the representation language can be synthesized, modified, and evaluated just like any other domain object. The consequence for the implementation is that heuristics suggesting operations on heuristics and frame structures, which one might well be tempted to treat specially as metaheuristics, are in fact distinguished in no way at all.

Heuristics themselves are stated informally in a way similar to a typical production system rule: a set of preconditions followed by consequent action. But the fact that these are represented in a form identical to that used for domain concepts allows the program to learn about and alter its own knowledge base.

5.1 THE REPRESENTATION LANGUAGE FOR CONCEPTS

The basic representation in the system employs frame-like units with slots, and this is applied uniformly to domain concepts, heuristic rules, and system concepts as well. Each slot with its name may be considered as a unary function, which, when given that name returns the value in the slot, and thus rules need not distinguish slots from functions. Each slot also has its own unit describing the slot, including such information as the types of legal entries in the slot, what units such a slot is legally a part of, etc. (see for example figure 5-1). Using such information the program may alter its own representation; if it is noticed that certain slots and values are never used they may be eliminated or, more often, new slots created for values that have been computed repeatedly on the fly or, if entries on a given slot become too numerous, the program will search for specializations so that the slot can be split. Since concepts are related to each other through Generalizations, Specializations, IsA, and Examples slots, knowledge becomes organized in a huge generalization/specialization hierarchy. This is true of the heuristic rules as well, with the general so-called "Weak Methods" (generate and test, hill-climbing, etc.) at the top and hundreds of domain-specific judgments near the bottom. One of the principle stated goals of the research is to get a grasp on the structure of this "space of heuristics."

NAME: IsA, Isa, Is-a, ISA, IS-A
  Informally: is, element-of, is-in
DOMAIN/RANGE: ⟨Units → SetOfUnits⟩
IS-A: Sot
FilledWithA: Set
EachEntryMustBeA: Unit representing a set
Inverse: Examples
UsedByInheritanceModes: InheritAlongIsAs
MakesSenseFor: Anything
MyIsA: Eurisko unit
MySize: 500 words
MyCreator: D. Lenat
MyTimeOfCreation: 4/4/79 12:01
Generalizations: AKindOf
Specializations: MemberOf, ExtremumOf
Worth: 600
Cache: Always
English: The slot which tells which classes a unit belongs to.
ALGORITHMS:
  Nonrecursive Slow PossiblyLooping: λ (u) $\{c \in \text{Concepts} \mid c.\text{Defn}(u)\}$
DEFINITIONS:
  Nonrecursive Fast PossiblyLooping: λ (u,c) $c.\text{Defn}(u)$

FIGURE 5-1: FRAME UNIT REPRESENTING THE CONCEPT OF AN "ISA" SLOT (LENAT 1983B)

5.2 THE CONTROL STRUCTURE

The control algorithm as well is represented in EURISKO as a set of concepts with the intention that the program be able to meaningfully alter its own control code, but this has always resulted in bugs. Basically the structure consists of three nested Select-Execute-PostMortem loops: select and work on a topic; given a topic, select and work on a promising task; given a task, select and obey a relevant heuristic rule. Each topic in the system (e.g., NumberTheory, Games, DevicePhysics, Aesthetics) has a slot called Agenda containing a list of tasks pertaining to that topic and its specializations. When the execution of a heuristic suggests a new task involving, say PalindromicNumbers, EURISKO then searches up along Generalizations links (e.g., through

Numbers and SymmetricConstructs) until it finds topics (ultimately, NumberTheory and Aesthetics) and a pointer to the task is put on the agenda in each topic encountered. But note that these topics are represented, like any other concept, in a frame unit, and there are rules that can operate on them, to split and specialize a topic if it has too many tasks piling up on its agenda, or else when an agenda becomes too small, to merge it into all appropriate immediate generalizations' agendas. "In such cases, the general agendas should adopt (a little of) the small agenda's aesthetics, values, heuristics, reasons, goals, open problems, points of view" (Lenat 1983b, 68), though in the current implementation, only the tasks are inherited.

Task selection is based on ratings which are constantly updated during the post-mortem phase of actions. When a task fails it is likely to be put back on the agenda along with new tasks which might enable this one to succeed. If the subtasks in turn succeed, then their post-mortem should raise the rating of the original task.

Once a task is chosen a body of potentially relevant heuristic rules are collected that may help to satisfy it. There are in fact a number of different ways that a rule may be interpreted depending on space and time bounds, how many rules are applicable (whether they should be ordered first), whether the interpreter must check whether some new rules become relevant during the course of rule execution, and so on, and of course there are heuristics that govern these choices. The post-mortem phase of an individual rule execution involves bookkeeping information about time and space used, new units created, tasks placed on agendas, etc.

5.3 HUMAN INTERFACE

There is a high frequency with which EURISKO generates "new" concepts that are effectively equivalent to already existing ones. The problem of recognizing this equivalence is serious and difficult; and the program, as one might expect, has some heuristic tools to try to detect this. Each unit has slots which are specified as "criterial" and which define it. The description of each such slot, in turn, describes the way it makes sense to do matching. For example, for concepts that are defined logically in terms of others, one might recur on the boolean subexpressions, or if the concepts are of algorithms, one might evaluate the functions on test arguments to see if they yield the same results. But this is one area in which human supervision is most helpful in practice to keep things from bogging down. Lenat admits to considerable intervention and "hand smoothing" of newly-generated concepts.

There is also some elementary user modeling in order to accommodate the varied expectations of users from differing fields from whom the program is expected to learn. Each group of people and the set AllPeople get their own concept frames, which are arranged in a hierarchy with the concepts of individual users at the bottom. These concepts influence what kinds of topics and tasks are chosen to work on, and what is chosen to be explained in what depth. "EURISKO does learn simple models of each new user, but there are at present very few psychological and societal heuristics for building up (and testing!) such models. Based on our model of theory formation we are not surprised that only minimal sorts of learning were achieved without a deep model of the domain" (Lenat 1983b, 73). This is clearly an area where development of some emotional modeling could be important and useful.

5.4 CONCLUSIONS

Certain criteria emerge concerning appropriate domains for mechanized discovery, chiefly:

- The search space should be immense and largely unexplored by humans (as in three-dimensional VLSI design, or the Traveller Trillion Credit Squadron naval fleet design wargame).
- There must be ways to simulate or directly carry out experiments.
- The domain should be rich in heuristic structure with no good algorithms, as in a very complex domain where precise inference is unmanageable or impossible, raising the utility of plausible inexact reasoning.

Once one gets past the impressive success of the program in performing genuinely innovative discovery, one begins to recognize the "personality" of the program as that of a tireless monomaniac focused entirely on this task of search and synthesis. In a more general application, or in the context of a more general intelligence, it is clear that this kind of activity should be rather occasional, if for no other reason than that it is so extremely costly.

Thus one is brought back again to the question of what should initiate this kind of search. Within the present discussion, the answer is that it should be the recognition of an emotional state that initiates the search in order to displace the emotion (or augment it in the case of a positive state).

Consider, for example, a computer employed in a mundane task, say general accounting, in an environment of computer non-sophisticates, say in an insurance office, in which users can largely be expected to want to use the machine and access its data through a natural language interface. Now one can also expect a user to occasionally express his displeasure with the machine when it misinterprets some request in English and doesn't give him what he wants. It makes good sense then for the machine to respond to this expression appropriately by "parsing" it as a motion in the Emotional Matrix of Interpersonal Space (see figure 2-4) corresponding to a motion of the "object" (user) away from the "person" (machine) and the "person" (machine) is the object of the emotion, along the dimension of recognition. This position in the matrix reads out "shame." If this happens often the rating of the shame frame will increase until it demands the attention an intense emotion does, by achieving a high priority on a task agenda. The unit representing shame can then be expected to suggest appropriate actions, the first of which is probably to apologize, and subsequently to "think about it," try to form a concept of what the misunderstanding is and why it comes about, and to try to create some heuristics relevant to avoiding the misunderstanding in the future, augment the user models by noting that people like this get upset about such things, and finally to seek a little reassuring confirmation from the user that these surmises are reasonable, not to mention a little pat on the CRT.

The potential for Artificial Intelligence is certainly enormous, but already applications based exclusively on formal logical inference face bottlenecks and limitations. In our efforts to be practical and precise in our thinking, let us not forget the heart and "its reasons," because there are very good reasons indeed that our emotional lives are as rich and commanding as they are, for affect and aesthetics are the very basis for knowledge, even purely factual knowledge. Whatever our minds—or any minds we create—may achieve must owe much in the end to the influence of such reasons of the heart.

Bibliography

deJoia, Alex and Stenton, Adrian. 1980. *Terms in Systemic Linguistics: A Guide to Halliday*. London: Batsford Academic and Educational Inc.

Dennet, Daniel C. 1978. *Brainstorms: Philosophical Essays on Mind and Psychology*. Montgomery, Vt.: Bradford Books.

deRivera, Joseph. 1977. *A Structural Theory of the Emotions*. Psychological Issues, Monograph 40. New York: International Universities Press.

Goodman, Nelson. 1976. *Languages of Art, an Approach to a Theory of Symbols*. Indianapolis: Hackett.

Halliday, M.A.K. 1978. *Language as a Social Semiotic*. Baltimore, Md.: University Park Press.

Halliday, M.A.K. and Martin, J.R. eds. 1981. *Readings in Systemic Linguistics*. London: Batsford Academic and Educational Inc.

Hanslick, Eduard. 1957. *The Beautiful in Music*. New York: Liberal Arts Press.

Hardy, Godfrey Harold. 1967. *A Mathematician's Apology*. London: Cambridge University Press.

Kress, G.R. 1976. *Halliday: System and Function in Language*. London: Oxford University Press.

Lehnert, Wendy. 1982. "Plot Units: A Narrative Summarization Strategy." In *Strategies for Natural Language Processing*. Hillsdale N.J.: Erlbaum Associates.

Lenat, Douglas B. 1983. Theory Formation by Heuristic Search; The Nature of Heuristics, II: Background and Examples. *Arificial Intelligence*, 21(1): 31–59.

Lenat, Douglas B. 1983. EURISKO: A Program that Learns New Heuristics and Domain Concepts; The Nature of Heuristics, III: Program Design and Results. *Artificial Intelligence*, 21(1): 61–98.

Lerdahl, Fred and Jackendoff, Ray. 1983. *A Generative Theory of Tonal Music*. Cambridge, Mass.: MIT Press.

Minsky, Marvin. 1979. K-Lines: A Theory of Memory. *MIT AI Memo 516* (June, 1979): 1–35.

Minsky, Marvin. 1981. Music, Mind, and Meaning. *Computer Music Journal* 5(3): 28–44.

Pais, Abraham. 1982. *Subtle is the Lord. . .: The Science and Life of Albert Einstein.* London: Oxford University Press.

Papert, Seymour. 1980. *Mindstorms: Children, Computers, and Powerful Ideas.* New York: Basic Books, 1980.

Poincaré, Henri. Translated by Halstead, G.B. 1958. *Science and Method.* New York: Dover.

Poincaré, Henri. Translated by Bolduc, John W. 1963. *Mathematics and Science: Last Essays (Dernieres Pensees).* New York: Dover.

Polanyi, Michael. 1958. *Personal Knowledge, Towards a Post-Critical Philosophy.* Chicago: University of Chicago Press.

Roads, Curtis. 1979. Grammars as Representations for Music. *Computer Music Journal* 3(1): 48–55.

Schank, Roger C., and Abelson Robert P. 1978. *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures.* Hillsdale, N.J.: Erlbaum Assoc.

Searle, John R. 1980. "Minds, Brains, and Programs." In *The Behavioral and Brain Sciences.* London: Cambridge University Press.

Wechsler, Judith, ed. 1978. *On Aesthetics in Science.* Cambridge, Mass.: MIT Press.

Winograd, Terry. 1968. "Linguistics and the Computer Analysis of Tonal Harmony." *Journal of Music Theory* 12(1): 2–50.

Winograd, Terry. 1983. *Language as a Cognitive Process: Syntax.* Reading, Mass.: Addison-Wesley.