CHAPTER 6

# Information Processing in Dynamical Systems: Foundations of Harmony Theory

P. SMOLENSKY

## INTRODUCTION

### The Theory of Information Processing

At this early stage in the development of cognitive science, methodological issues are both open and central. There may have been times when developments in neuroscience, artificial intelligence, or cognitive psychology seduced researchers into believing that their discipline was on the verge of discovering the secret of intelligence. But a humbling history of hopes disappointed has produced the realization that understanding the mind will challenge the power of all these methodologies combined.

The work reported in this chapter rests on the conviction that a methodology that has a crucial role to play in the development of cognitive science is *mathematical analysis*. The success of cognitive science, like that of many other sciences, will, I believe, depend upon the construction of a solid body of theoretical results: results that express in a mathematical language the conceptual insights of the field; results that squeeze all possible implications out of those insights by exploiting powerful mathematical techniques.

This body of results, which I will call the *theory of information processing*, exists because information is a concept that lends itself to mathematical formalization. One part of the theory of information processing is already well-developed. The classical theory of computation provides powerful and elegant results about the notion of *effective*

*procedure*, including languages for precisely expressing them and theoretical machines for realizing them. This body of theory grew out of mathematical logic, and in turn contributed to computer science, physical computing systems, and the theoretical paradigm in cognitive science often called *the (von Neumann) computer metaphor.*[1]

In his paper "Physical Symbol Systems," Allen Newell (1980) articulated the role of the mathematical theory of symbolic computation in cognitive science and furnished a manifesto for what I will call *the symbolic paradigm.* The present book offers an alternative paradigm for cognitive science, the *subsymbolic paradigm*, in which the most powerful level of description of cognitive systems is hypothesized to be lower than the level that is naturally described by symbol manipulation.

The fundamental insights into cognition explored by the subsymbolic paradigm do not involve effective procedures and symbol manipulation. Instead they involve the "spread of activation," relaxation, and statistical correlation. The mathematical language in which these concepts are naturally expressed are probability theory and the theory of dynamical systems. By dynamical systems theory I mean the study of sets of numerical variables (e.g., activation levels) that evolve in time in parallel and interact through differential equations. The classical theory of dynamical systems includes the study of natural physical systems (e.g., mathematical physics) and artificially designed systems (e.g., control theory). Mathematical characterizations of dynamical systems that formalize the insights of the subsymbolic paradigm would be most helpful in developing the paradigm.

This chapter introduces *harmony theory*, a mathematical framework for studying a class of dynamical systems that perform cognitive tasks according to the account of the subsymbolic paradigm. These dynamical systems can serve as models of human cognition or as designs for artificial cognitive systems. The ultimate goal of the enterprise is to develop a body of mathematical results for the theory of information processing that complements the results of the classical theory of (symbolic) computation. These results would serve as the basis for a manifesto for the subsymbolic paradigm comparable to Newell's manifesto for the symbolic paradigm. The promise offered by this goal will, I hope, be suggested by the results of this chapter, despite their very limited scope.

---

[1] Mathematical logic has recently given rise to another approach to formalizing information: *situation semantics* (Barwise & Perry, 1983). This is related to Shannon's (1948/1963) measure of information through the work of Dretske (1981). The approach of this chapter is more faithful to the probabilistic formulation of Shannon than is the symbolic approach of situation semantics. (This results from Dretske's move of identifying information with conditional probabilities of 1.)

It should be noted that harmony theory is a "theory" in the *mathematical* sense, not the *scientific* sense. By a "mathematical theory"—e.g., number theory, group theory, probability theory, the theory of computation—I mean a body of knowledge about a part of the ideal mathematical world; a set of definitions, axioms, theorems, and analytic techniques that are tightly interrelated. Such mathematical theories are distinct from scientific theories, which are of course bodies of knowledge about a part of the "real" world. Mathematical theories provide a language for expressing scientific theories; a given mathematical theory can be used to express a large class of scientific theories. Group theory, for example, provides a language for expressing many competing theories of elementary particles. Similarly, harmony theory can be used to express many alternative theories about various cognitive phenomena. The point is that without the concepts and techniques of the mathematical language of group theory, the formulation of *any* of the current scientific theories of elementary particles would be essentially impossible.

The goal of harmony theory is to provide a powerful language for expressing cognitive theories in the subsymbolic paradigm, a language that complements the existing languages for symbol manipulation. Since harmony theory is conceived as a language for using the subsymbolic paradigm to describe cognition, it embodies the fundamental scientific claims of that paradigm. But on many important issues, such as how knowledge is represented in detail for particular cases, harmony theory does not itself make commitments. Rather, it provides a language for stating alternative hypotheses and techniques for studying their consequences.

## A Top-Down Theoretical Strategy

How can mathematical analysis be used to study the processing mechanisms underlying the performance of some cognitive task?

One strategy, often associated with David Marr (1982), is to characterize the task in a way that allows mathematical *derivation* of mechanisms that perform it. This *top-down* theoretical strategy is pursued in harmony theory. My claim is not that the strategy leads to descriptions that are *necessarily* applicable to all cognitive systems, but rather that the strategy leads to new insights, mathematical results, computer architectures, and computer models that fill in the relatively unexplored conceptual world of parallel, massively distributed systems that perform cognitive tasks. Filling in this conceptual world is a necessary subtask, I believe, for understanding how brains and minds are capable of intelligence and for assessing whether computers with novel architectures might share this capability.

## The Centrality of Perceptual Processing

The cognitive task I will study in this chapter is an abstraction of the task of perception. This abstraction includes many cognitive tasks that are customarily regarded as much "higher level" than perception (e.g., intuiting answers to physics problems). A few comments on the role of perceptual processing in the subsymbolic paradigm are useful at this point.

The vast majority of cognitive processing lies between the highest cognitive levels of explicit logical reasoning and the lowest levels of sensory processing. Descriptions of processing at the extremes are relatively well-informed—on the high end by formal logic and on the low end by natural science. In the middle lies a conceptual abyss. How are we to conceptualize cognitive processing in this abyss?

The strategy of the symbolic paradigm is to conceptualize processing in the intermediate levels as symbol manipulation. Other kinds of processing are viewed as limited to extremely low levels of sensory and motor processing. Thus symbolic theorists climb *down* into the abyss, clutching a rope of symbolic logic anchored at the top, hoping it will stretch all the way to the bottom of the abyss.

The subsymbolic paradigm takes the opposite view, that intermediate processing mechanisms are of the same kind as perceptual processing mechanisms. Logic and symbol manipulation are viewed as appropriate descriptions only of the few cognitive processes that explicitly involve logical reasoning. Subsymbolic theorists climb *up* into the abyss on a perceptual ladder anchored at the bottom, hoping it will extend all the way to the top of the abyss.[2]

---

[2] There is no contradiction between working from lower level, perceptual processes up towards higher processes, and pursuing a top-down theoretical strategy. It is important to distinguish levels of *processing entities* from levels of *theoretical entities*. Higher level *processes* involve *computational* entities that are computationally distant from the peripheral, sensorimotor entities that comprise the "lowest level" of processing. These processing levels *taken together* form the processing system as a whole; they causally interact with each other through bottom-up and top-down *processing*. Higher level *theories* involve *descriptive* entities that are descriptively distant from entities that are directly part of an actual processing mechanism; these comprise the "lowest level" description. Each theoretical level *individually* describes the processing system as a whole; the interaction of descriptive levels is not *causal*, but *definitional*. (For example, changes in individual neural firing rates at the retina *cause* changes in individual firing rates in visual cortex after a delay related to causal information propagation. The same changes in individual retinal neuron firing rates *by definition* change the *average firing rates of pools* of retinal neurons; these higher level descriptive entities change instantly, without any causal information propagation from the lower level description.) Thus in harmony theory, models of higher level *processes* are derived from models of lower level, perceptual, processes, while lower level *descriptions* of these models are derived from higher level descriptions.

In this chapter, I will analyze an abstraction of the task of perception that encompasses many tasks, from low, through intermediate, to high cognitive levels. The analysis leads to a general kind of "perceptual" processing mechanism that is a powerful potential component of an information processing system. The abstract task I analyze captures a common part of the tasks of passing from an intensity pattern to a set of objects in three-dimensional space, from a sound pattern to a sequence of words, from a sequence of words to a semantic description, from a set of patient symptoms to a set of disease states, from a set of givens in a physics problem to a set of unknowns. Each of these processes is viewed as *completing an internal representation of a static state of an external world.* By suitably abstracting the task of interpreting a static *sensory* input, we can arrive at a theory of interpretation of static input *generally*, a theory of the *completion task* that applies to many cognitive phenomena in the gulf between perception and logical reasoning. An application that will be described in some detail is qualitative problem solving in circuit analysis.[3]

The central idea of the top-down theoretical strategy is that properties of the task are powerfully constraining on mechanisms. This idea can be well exploited within a perceptual approach to cognition, where the constraints on the perceptual task are characterized through the constraints operative in the external environment from which the inputs come. This permits an analysis of how internal representation of these constraints within the cognitive system itself allows it to perform its task. These kinds of considerations have been emphasized in the psychological literature prominently by Gibson and Shepard (see Shepard, 1984); they are fundamental to harmony theory.

## Structure of the Chapter

The goal of harmony theory is to develop a mathematical theory of information processing in the subsymbolic paradigm. However, the theory grows out of ideas that can be stated with little or no mathematics. The organization of this chapter reflects an attempt to ensure that the central concepts are not obscured by mathematical opacity. The analysis will be presented in three parts, each part increasing in the level of formality and detail. My hope is that the slight redundancy

[3] Many cognitive tasks involve interpreting or controlling events that unfold over an extended period of time. To deal properly with such tasks, harmony theory must be extended from the interpretation of *static* environments to the interpretation of *dynamic* environments.

introduced by this expository organization will be repaid by greater accessibility.

Section 1 is a top-down presentation of how the perceptual perspective on cognition leads to the basic features of harmony theory. This presentation starts with a particular perceptual model, the letter-perception model of McClelland and Rumelhart (1981), and abstracts from it general features that can apply to modeling of higher cognitive processes. Crucial to the development is a particular formulation of aspects of schema theory, along the lines of Rumelhart (1980).

Section 2, the majority of the chapter, is a bottom-up presentation of harmony theory that starts with the primitives of the knowledge representation. Theorems are informally described that provide a competence theory for a cognitive system that performs the completion task, a machine that realizes this theory, and a learning procedure through which the machine can absorb the necessary information from its environment. Then an application of the general theory is described: a model of intuitive, qualitative problem-solving in elementary electric circuits. This model illustrates several points about the relation between symbolic and subsymbolic descriptions of cognitive phenomena; for example, it furnishes a sharp contrast between the description at these two levels of the nature and acquisition of expertise.

The final part of the chapter is an Appendix containing a concise but self-contained formal presentation of the definitions and theorems.

# SECTION 1: SCHEMA THEORY AND SELF-CONSISTENCY

## THE LOGICAL STRUCTURE OF HARMONY THEORY

The logical structure of harmony theory is shown schematically in Figure 1. The box labeled *Mathematical Theory* represents the use of mathematical analysis and computer simulation for drawing out the implications of the fundamental principles. These principles comprise a mathematical characterization of computational requirements of a cognitive system that performs the completion task. From these principles

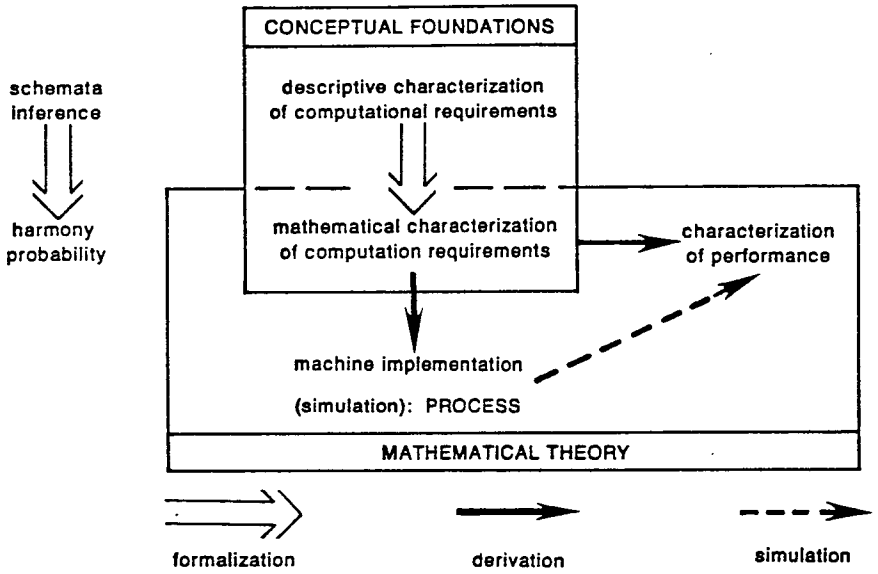


FIGURE 1. The logical structure of harmony theory.

it is possible to mathematically analyze aspects of the resulting performance as well as rigorously *derive* the rules for a machine implementing the computational requirements. The rules defining this machine have a different status from those defining most other computer models of cognition: They are not ad hoc, or post hoc; rather they are logically derived from a set of computational requirements. This is one sense in which harmony theory has a top-down theoretical development.

Where do the "mathematically characterized computational requirements" of Figure 1 come from? They are a formalization of a descriptive characterization of cognitive processing, a simple form of *schema theory*. In Section 1 of this chapter, I will give a description of this form of schema theory and show how to transform the descriptive characterization into a mathematical one—how to get from the *conceptual* box of Figure 1 into the *mathematical* box. Once we are in the formal world, mathematical analysis and computer simulation can be put to work.

Throughout Section 1, the main points of the development will be explicitly enumerated.

*Point 1. The mathematics of harmony theory is founded on familiar concepts of cognitive science: inference through activation of schemata.*

## DYNAMIC CONSTRUCTION OF SCHEMATA

The basic problem can be posed à la Schank (1980). While eating at a fancy restaurant, you get a headache. Without effort, you ask the waitress if she could possibly get you an aspirin. How is this plan created? You have never had a headache in a restaurant before. Ordinarily, when you get a headache your plan is to go to your medicine cabinet and get yourself some aspirin. In the current situation, this plan must be modified by the knowledge that in good restaurants, the management is willing to expend effort to please its customers, and that the waitress is a liaison to that management.

The cognitive demands of this situation are schematically illustrated in Figure 2. Ordinarily, the restaurant context calls for a "restaurant script" which supports the planning and inferencing required to reach the usual goal of getting a meal. Ordinarily, the headache context calls for a "headache script" which supports the planning required to get aspirin in the usual context of home. The completely novel context of a headache in a restaurant calls for a special-purpose script integrating the knowledge that ordinarily manifests itself in two separate scripts.

What kind of cognitive system is capable of this degree of flexibility? Suppose that the knowledge base of the system does *not* consist of a set of scripts like the restaurant script and the headache script. Suppose

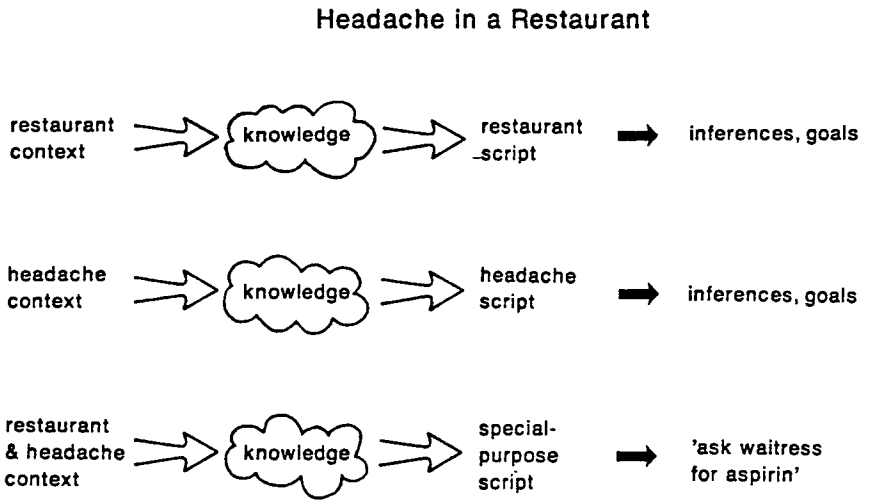


FIGURE 2. In three different contexts, the knowledge base must produce three different scripts.

instead that the knowledge base is a set of *knowledge atoms* that configure themselves dynamically in each context to form tailor-made scripts. This is the fundamental idea formalized in harmony theory.[4]

The degree of flexibility demanded of scripts is equaled by that demanded of all conceptual structures.[5] For example, metaphor is an extreme example of the flexibility demanded of word meanings; even so-called literal meaning on closer inspection actually relies on extreme flexibility of knowledge application (Rumelhart, 1979). In this chapter I will consider knowledge structures that embody our knowledge of objects, words, and other concepts of comparable complexity; these I will refer to as *schemata*. The defining properties of schemata are that they have conceptual interpretations and that they *support inference*.

For lack of a better term, I will use *knowledge atoms* to refer to the elementary constituents of which I assume schemata to be composed.[6] These atoms will shortly be given a precise description; they will be interpreted as a particular instantiation of the idea of *memory trace*.

*Point 2. At the time of inference, stored knowledge atoms are dynamically assembled into context-sensitive schemata.*

This view of schemata was explicitly articulated in Feldman (1981). It is in part embodied in the McClelland and Rumelhart (1981) letter-perception model (see Chapter 1). One of the observed phenomena accounted for by this model is the facilitation of the perception of letters that are embedded in words. Viewing the perception of a letter as the result of a perceptual inference process, we can say that this inference is supported by a *word schema* that appears in the model as a single processing unit that encodes the knowledge of the spelling of that word. This is *not* an instantiation of the view of schemata as dynamically created entities.

---

[4] Schank (1980) describes a *symbolic* implementation of the idea of dynamic script construction; harmony theory constitutes a *subsymbolic* formalization.

[5] Hofstadter has long been making the case for the inadequacy of traditional symbolic descriptions to cope with the power and flexibility of concepts. For his most recent argument, see Hofstadter (1985). He argues for the need to admit the approximate nature of symbolic descriptions, and to explicitly consider processes that are *subcognitive*. In Hofstadter (1979, p. 324ff), this same case was phrased in terms of the need for "active symbols," of which the "schemata" described here can be viewed as instances.

[6] A physicist might call these particles *gnosons* or *sophons*, but these terms seem quite uneuphonious. An acronym for *Units for Constructing Schemata Dynamically* might serve, but would perhaps be taken as an advertising gimmick. So I have stuck with "knowledge atoms."

However, the model also accounts for the observed facilitation of letter perception within orthographically regular nonwords or *pseudowords* like *MAVE*. When the model processes this stimulus, several word units become and stay quite active, including *MAKE, WAVE, HAVE*, and other words orthographically similar to *MAVE*. In this case, the perception of a letter in the stimulus is the result of an inference process that is supported by the *collection* of activated units. This collection is a *dynamically created pseudoword schema*.

When an orthographically irregular nonword is processed by the model, letter perception is slowest. As in the case of pseudowords, many word units become active. However, none become very active, and very many are equally active, and these words have very little similarity to each other, so they do not support inference about the letters effectively. Thus the knowledge base is incapable of creating schemata for irregular nonwords.

*Point 3. Schemata are coherent assemblies of knowledge atoms; only these can support inference.*

Note that schemata are created *simply by activating the appropriate atoms*. This brings us to what was labeled in Figure 1 the "descriptively characterized computational requirements" for harmony theory:

*Point 4: The harmony principle. The cognitive system is an engine for activating coherent assemblies of atoms and drawing inferences that are consistent with the knowledge represented by the activated atoms.*

*Subassemblies* of activated atoms that tend to recur exactly or approximately are the schemata.

This principle focuses attention on the notion of *coherency* or *consistency*. This concept will be formalized under the name of *harmony*, and its centrality is acknowledged by the name of the theory.

## MICRO- AND MACROLEVELS

It is important to realize that harmony theory, like all subsymbolic accounts of cognition, exists on two distinct levels of description: a microlevel involving knowledge atoms and a macrolevel involving schemata (see Chapter 14). These levels of description are completely analogous to other micro- and macrotheories, for example, in physics. The microtheory, quantum physics, is assumed to be universally valid. Part of its job as a theory is to explain why the approximate macrotheory, classical physics, works when it does and why it breaks

down when it does. Understanding of physics requires understanding *both* levels of theory *and* the relation between them.

In the subsymbolic paradigm in cognitive science, it is equally important to understand the two levels and their relationship. In harmony theory, the microtheory prescribes the nature of the atoms, their interaction, and their development through experience. This description is assumed to be a universally valid description of cognition. It is also assumed (although this has yet to be explicitly worked out) that in performing certain cognitive tasks (e.g., logical reasoning), a higher level description is a valid approximation. This macrotheory describes schemata, their interaction, and their development through experience.

One of the features of the formalism of harmony theory that distinguishes it from most subsymbolic accounts of cognition is that it exploits a formal isomorphism with statistical physics. Since the main goal of statistical physics is to relate the microscopic description of matter to its macroscopic properties, harmony theory can bring the power of statistical physics concepts and techniques to bear on the problem of understanding the relation between the micro- and macroaccounts of cognition.

## THE NATURE OF KNOWLEDGE

In the previous section, the letter-perception model was used to illustrate the dynamic construction of schemata from constituent atoms. However, it is only pseudowords that correspond to composite schemata; word schemata are single atoms. We can also represent words as composite schemata by using digraph units at the upper level instead of four-letter word units. A portion of this modified letter-perception model is shown in Figure 3. Now the processing of a four-letter word involves the activation of a set of digraph units, which are the knowledge atoms of this model. Omitted from the figure are the

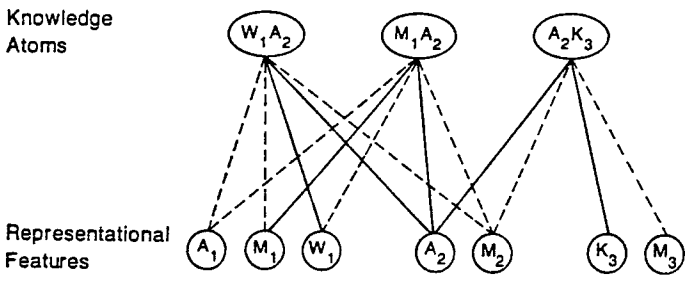


FIGURE 3. A portion of a modified reading model.

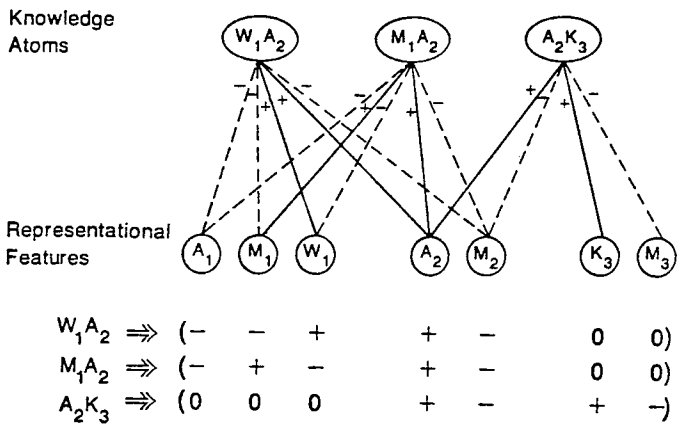


FIGURE 4. Each knowledge atom is a vector of +, −, and 0 values of the representational feature nodes.

line-segment units, which are like those in the original letter-perception model.

This simple model illustrates several points about the nature of knowledge atoms in harmony theory. The digraph unit $W_1A_2$ represents a pattern of values over the letter units: $W_1$ and $A_2$ on, with all other letter units for positions 1 and 2 off. This pattern is shown in Figure 4, using the labels +, −, and 0 to denote *on*, *off*, and *irrelevant*. These indicate whether there is an excitatory connection, inhibitory connection, or no connection between the corresponding nodes.[7]

Figure 4 shows the basic structure of harmony models. There are atoms of knowledge, represented by nodes in an upper layer, and a lower layer of nodes that comprises a representation of the state of the perceptual or problem domain with which the system deals. Each node is a *feature* in the representation of the domain. We can now view "atoms of knowledge" like $W_1$ and $A_2$ in several ways. Mathematically, each atom is simply a *vector* of +, −, and 0 values, one for each node in the lower, representation layer. This pattern can also be viewed as a *fragment* of a percept: The 0 values mark those features omitted in the fragment. This fragment can in turn be interpreted as a *trace* left behind in memory by perceptual experience.

[7] Omitted are the knowledge atoms that relate the letter nodes to the line segment nodes. Both line segment and letter nodes are in the lower layer, and all knowledge atoms are in the upper layer. Hierarchies in harmony theory are imbedded within an architecture of only two layers of nodes, as will be discussed in Section 2.

*Point 5. Knowledge atoms are fragments of representations that accumulate with experience.*

## THE COMPLETION TASK

Having specified more precisely what the atoms of knowledge are, it is time to specify the task in which they are used.

Many cognitive tasks can be viewed as inference tasks. In problem solving, the role of inference is obvious; in perception and language comprehension, inference is less obvious but just as central. In harmony theory, a tightly prescribed but extremely general inferential task is studied: the *completion task*. In a problem-solving completion task, a partial description of a situation is given (for example, the initial state of a system); the problem is to complete the description to fill in the missing information (the final state, say). In a story understanding completion task, a partial description of some events and actors' goals is given; comprehension involves filling in the missing events and goals. In perception, the stimulus gives values for certain low-level features of the environmental state, and the perceptual system must fill in values for other features. In general, in the completion task some features of an environmental state are given as input, and the cognitive system must complete that input by assigning likely values to unspecified features.

A simple example of a completion task (Lindsay & Norman, 1972) is shown in Figure 5. The task is to fill in the features of the obscured portions of the stimulus and to decide what letters are present. This task can be performed by the model shown in Figure 3, as follows. The stimulus assigns values of *on* and *off* to the unobscured letter features. What happens is summarized in Table 1.

Note that which atoms are activated affects how the representation is

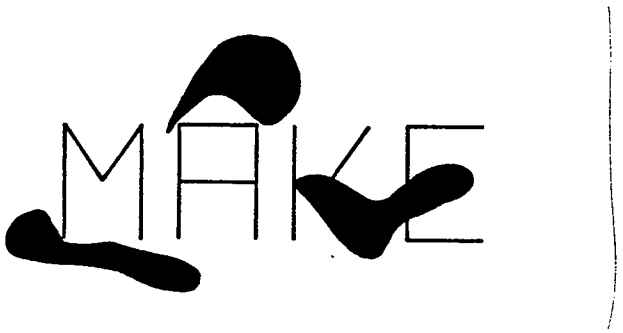


FIGURE 5. A perceptual completion task.

TABLE I

A PROCEDURE FOR PERFORMING THE COMPLETION TASK

| | |
|---|---|
| Input: | Assign values to some features in the representation |
| Activation: | Activate atoms that are *consistent* with the representation |
| Inference: | Assign values to unknown features of representation that are *consistent* with the active knowledge |

filled in, and how the representation is filled in affects which atoms are activated. The activation and inference processes mutually constrain each other; these processes must run in parallel. Note also that all the decisions come out of a striving for *consistency*.

*Point 6. Assembly of schemata (activation of atoms) and inference (completing missing parts of the representation) are both achieved by finding maximally self-consistent states of the system that are also consistent with the input.*

The completion of the stimulus shown in Figure 5 is shown in Figure 6. The consistency is high because wherever an active atom is

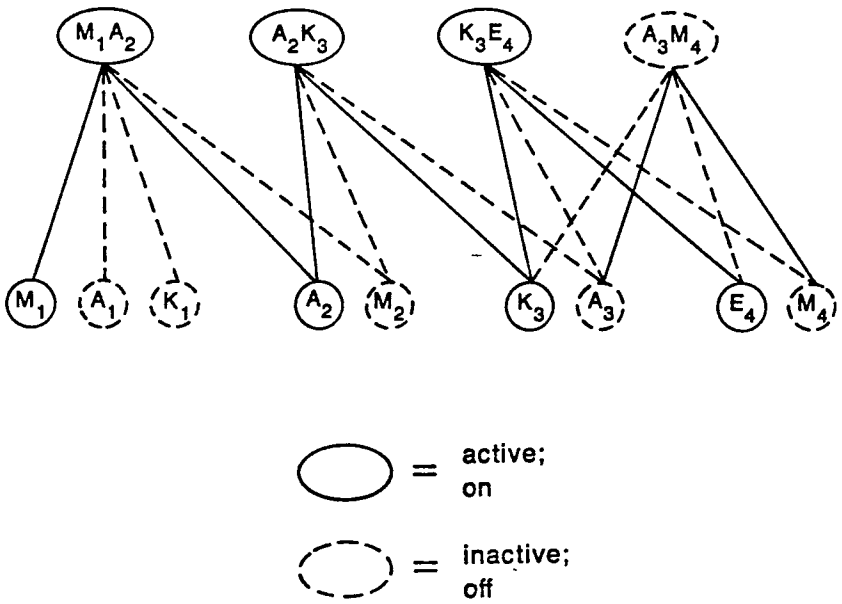


FIGURE 6. The state of the network in the completion of the stimulus shown in Figure 5.

connected to a representational feature by a + (respectively, −) connection, that feature has value *on* (respectively, *off*). In fact, we can define a very simple measure of the degree of self-consistency just by considering all active atoms, counting +1 for every agreement between one of its connections and the value of the corresponding feature, and counting −1 for every disagreement. (Here + with *on* or − with *off* constitutes agreement.) This is the simplest example of a *harmony function*—and brings us into the mathematical formulation.

## THE HARMONY FUNCTION

Point 6 asserts that a central cognitive process is the construction of cognitive states that are "maximally self-consistent." To make this precise, we need only measure that self-consistency.

*Point 7. The self-consistency of a possible state of the cognitive system can be assigned a quantitative value by a harmony function, H.*

Figure 7 displays a harmony function that generalizes the simple example discussed in the preceding paragraph. A state of the system is defined by a set of atoms which are *active* and a vector of values for all representational features. The harmony of such a state is the sum of terms, one for each active atom, weighted by the *strength* of that atom. Each weight multiplies the self-consistency between that particular atom and the vector of representational feature values. That self-consistency is the similarity between the vector of features defining the atom (the vector of its connections) and the representational feature vector. In the simplest case discussed above, the function $h$ that measures this similarity is just the number of agreements between these vectors minus the number of disagreements. For reasons to be discussed, I have used a slightly more complicated version of $h$ in which the simpler form is first divided by the number of (nonzero) connections to the atom, and then a fixed value $\kappa$ is subtracted.

$$\text{harmony}_{\substack{\text{knowledge}\\\text{base}}}\left(\text{representational feature vector, activations}\right) =$$

$$\sum_{\substack{\text{atoms}\\\alpha}} \left(\begin{matrix}\text{strength of}\\\text{atom } \alpha\end{matrix}\right)\left(\begin{matrix}0 \text{ if atom}\\\alpha \text{ inactive;}\\1 \text{ if active}\end{matrix}\right) \text{similarity}\left(\begin{matrix}\text{feature vector}\\\text{of atom } \alpha\end{matrix}, \begin{matrix}\text{representational}\\\text{feature vector}\end{matrix}\right)$$

FIGURE 7. A schematic representation for a harmony function.

## A PROBABILISTIC FORMULATION OF SCHEMA THEORY

The next step in the theoretical development requires returning to the higher level, symbolic description of inference, and to a more detailed discussion of schemata.

Consider a typical inference process described with schemata. A child is reading a story about presents, party hats, and a cake with candles. When asked questions, the child says that the girl getting the presents is having a birthday. In the terminology of schema theory, while reading the story, the child's *birthday party schema* becomes active and allows many inferences to be made, filling in details of the scene that were not made explicit in the story.

The birthday party schema is presumed to be a knowledge structure that contains *variables* like *birthday cake, guest of honor, other guests, gifts, location*, and so forth. The schema contains information on how to assign values to these variables. For example, the schema may specify: *default values* to be assigned to variables in the absence of any counterindicating information; *value restrictions* limiting the kind of values that can be assigned to variables; and *dependency* information, specifying how assigning a particular value to one variable affects the values that can be assigned to another variable.

A convenient framework for concisely and uniformly expressing all this information is given by *probability theory*. The default value for a variable can be viewed as its most probable value: the mode of the marginal probability distribution for that variable. The value restrictions on a variable specify the values for which it has nonzero probability: the support of its marginal distribution. The dependencies between variables are expressed by their statistical correlations, or, more completely, by their joint probability distributions.

So the birthday party schema can be viewed as containing information about the probabilities that its variables will have various possible values. These are clearly statistical properties of the particular domain or *environment* in which the inference task is being carried out. In reading the story, the child is given a partial description of a scene from the everyday environment—the values of some of the features used to represent that scene—and to understand the story, the child must *complete* the description by filling in the values for the unknown features. These values are assigned in such a way that the resulting scene has the highest possible probability. The birthday party schema contains the probabilistic information needed to carry out these inferences.

In a typical cognitive task, many schemata become active at once and interact heavily during the inference process. Each schema contains probabilistic information for its own variables, which are only a fraction

of the complete set of variables involved in the task. To perform a completion, the most probable set of values must be assigned to the unknown variables, using the information in all the active schemata.

This probabilistic formulation of these aspects of schema theory can be simply summarized as follows.

*Point 8. Each schema encodes the statistical relations among a few representational features. During inference, the probabilistic information in many active schemata are dynamically folded together to find the most probable state of the environment.*

Thus the statistical knowledge encoded in all the schemata allow the estimation of the relative probabilities of possible states of the environment. How can this be done?

At the macrolevel of schemata and variables, coordinating the folding together of the information of many schemata is difficult to describe. The inability to devise procedures that capture the flexibility displayed in human use of schemata was in fact one of the primary historical reasons for turning to the microlevel description (see Chapter 1). We therefore return to the microdescription to address this difficult problem.

At the microlevel, the probabilistic knowledge in the birthday party schema is distributed over many knowledge atoms, each carrying a small bit of statistical information. Because these atoms all tend to match the representation of a birthday party scene, they can become active together; in some approximation, they tend to function collectively, and in that sense they comprise a schema. Now, when many schemata are active at once, that means the knowledge atoms that comprise them are simultaneously active. At the microlevel, there is no real difference between the decisions required to activate the appropriate atoms to instantiate many schemata simultaneously and the decisions required to activate the atoms to instantiate a single schema. A computational system that can dynamically create a schema when it is needed can also dynamically create many schemata when they are needed. When atoms, not schemata, are the elements of computation, the problem of coordinating many schemata becomes subsumed in the problem of activating the appropriate atoms. And this is the problem that the harmony function, the measure of self-consistency, was created to solve.

## HARMONY THEORY

According to Points 2, 6, and 7, schemata are collections of knowledge atoms that become active in order to maximize harmony,

and inferences are also drawn to maximize harmony. This suggests that the probability of a possible state of the environment is estimated by computing its harmony: the higher the harmony, the greater the probability. In fact, from the mathematical properties of probability and harmony, in Section 2 we will show the following:

*Point 9. The relationship between the harmony function $H$ and estimated probabilities is of the form*

$$\text{probability} \propto e^{H/T}$$

*where $T$ is some constant that cannot be determined a priori.*

This relationship between probability and harmony is mathematically identical to the relationship between probability and (minus) energy in statistical physics: the Gibbs or Boltzmann law. This is the basis of the isomorphism between cognition and physics exploited by harmony theory. In statistical physics, $H$ is called the *Hamiltonian function*; it measures the energy of a state of a physical system. In physics, $T$ is the *temperature* of the system. In harmony theory, $T$ is called the *computational temperature* of the cognitive system. When the temperature is very high, completions with high harmony are assigned estimated probabilities that are only slightly higher than those assigned to low harmony completions; the environment is treated as *more random* in the sense that all completions are estimated to have roughly equal probability. When the temperature is very low, only the completions with highest harmony are given nonnegligible estimated probabilities.[8]

*Point 10. The lower the computational temperature, the more the estimated probabilities are weighted towards the completions of highest harmony.*

In particular, the very best completion can be found by lowering the temperature to zero. This process, *cooling*, is fundamental to harmony theory. Concepts and techniques from thermal physics can be used to understand and analyze decision-making processes in harmony theory.

A technique for performing Monte Carlo computer studies of thermal systems can be readily adapted to harmony theory.

*Point 11. A massively parallel stochastic machine can be designed that performs completions in accordance with Points 1-10.*

---

[8] Since harmony corresponds to *minus* energy, at low physical temperatures only the state with the *lowest* energy (the *ground state*) has nonnegligible probability.

For a given harmony model (e.g., that of Figure 4), this machine is constructed as follows. Every node in the network becomes a simple processor, and every link in the network becomes a communication link between two processors. The processors each have two possible values (+1 and −1 for the representational feature processors; 1 = *active* and 0 = *inactive* for the knowledge atom processors). The input to a completion problem is provided by fixing the values of some of the feature processors. Each of the other processors continually updates its value by making stochastic decisions based on the harmony associated at the current time with its two possible values. It is most likely to choose the value that corresponds to greater harmony; but with some probability—greater the higher is the computational temperature $T$—it will make the other choice. Each processor computes the harmony associated with its possible values by a numerical calculation that uses as input the numerical values of all the other processors to which it is connected. Alternately, all the atom processors update in parallel, and then all the feature processors update in parallel. The process repeats many times, implementing the procedure of Table 1. All the while, the temperature $T$ is lowered to zero, pursuant to Point 10. It can be proved that the machine will eventually "freeze" into a completion that maximizes the harmony.

I call this machine *harmonium* because, like the Selfridge and Neisser (1960) pattern recognition system *pandemonium*, it is a parallel distributed processing system in which many atoms of knowledge are simultaneously "shouting" out their little contributions to the inference process; but unlike pandemonium, there is an explicit method to the madness: the collective search for maximal harmony.[9]

The final point concerns the account of learning in harmony theory.

> *Point 12. There is a procedure for accumulating knowledge atoms through exposure to the environment so that the system will perform the completion task optimally.*

The precise meaning of "optimality" will be an important topic in the subsequent discussion.

This completes the descriptive account of the foundations of harmony theory. Section 2 fills in many of the steps and details omitted

---

[9] Harmonium is closely related to the *Boltzmann machine* discussed in Chapter 7. The basic dynamics of the machines are the same, although there are differences in most details. In the Appendix, it is shown that in a certain sense the Boltzmann machine is a special case of harmonium, in which knowledge atoms connected to more than two features are forbidden. In another sense, harmonium is a special case of the Boltzmann machine, in which the connections are restricted to go only between two layers.

above, and reports the results of some particular studies. The most formal matters are treated in the Appendix.

# SECTION 2: HARMONY THEORY

> *. . . the privileged unconscious phenomena, those susceptible of becoming conscious, are those which . . . affect most profoundly our emotional sensibility . . . Now, what are the mathematic entities to which we attribute this character of beauty and elegance . . . ? They are those whose elements are harmoniously disposed so that the mind without effort can embrace their totality while realizing the details. This harmony is at once a satisfaction of our esthetic needs and an aid to the mind, sustaining and guiding. . . . Figure the future elements of our combinations as something like the unhooked atoms of Epicurus. . . . They flash in every direction through the space . . . like the molecules of a gas in the kinematic theory of gases. Then their mutual impacts may produce new combinations.*
>
> Henri Poincaré (1913)
> Mathematical Creation[10]

In Section 1, a top-down analysis led from the demands of the completion task and a probabilistic formulation of schema theory to perceptual features, knowledge atoms, the central notion of harmony, and the role of harmony in estimating probabilities of environmental states. In Section 2, the presentation will be bottom-up, starting from the primitives.

## KNOWLEDGE REPRESENTATION

### Representation Vector

At the center of any harmony theoretic model of a particular cognitive process is a set of *representational features* $r_1, r_2, \cdots$. These

[10] I am indebted to Yves Chauvin for recently pointing out this remarkable passage by the great mathematician. See also Hofstadter (1985, pp. 655-656).

features constitute the cognitive system's representation of possible states of the environment with which it deals. In the environment of visual perception, these features might include pixels, edges, depths of surface elements, and identifications of objects. In medical diagnosis, features might be symptoms, outcomes of tests, diseases, prognoses, and treatments. In the domain of qualitative circuit analysis, the features might include *increase in current through resistor x* and *increase in voltage drop across resistor x*.

The representational features are variables that I will assume take on binary values that can be thought of as *present* and *absent* or *true* and *false*. Binary values contain a tremendous amount of representational power, so it is not a great sacrifice to accept the conceptual and technical simplification they afford. It will turn out to be convenient to denote *present* and *absent* respectively by $+1$ and $-1$, or, equivalently, $+$ and $-$. Other values could be used if corresponding modifications were made in the equations to follow. The use of continuous numerical feature variables, while introducing some additional technical complexity, would not affect the basic character of the theory.[11]

A *representational state* of the cognitive system is determined by a collection of values for all the representational variables $\{r_i\}$. This collection can be designated by a list or vector of $+$'s and $-$'s: the *representation vector* $\mathbf{r}$.

Where do the features used in the representation vector come from? Are they "innate" or do they develop with experience? These crucial questions will be deferred until the last section of this chapter. The evaluation of various possible representations for a given environment and the study of the development of good representations through exposure to the environment is harmony theory's *raison d'être*. But a prerequisite for understanding the appropriateness of a representation is understanding how the representation supports performance on the task for which it used; that is the primary concern of this chapter. For now, we simply assume that somehow a set of representational features has already been set up: by a programmer, or experience, or evolution.

---

[11] While continuous values make the *analysis* more complex, they may well improve the performance of the simulation models. In simulations with discrete values, the system state jumps between corners of a hypercube; with continuous values, the system state crawls smoothly around inside the hypercube. It was observed in the work reported in Chapter 14 that "bad" corners corresponding to stable nonoptimal completions (local harmony maxima) were typically *not* visited by the smoothly moving continuous state; these corners typically *are* visited by the jumping discrete state and can only be escaped from through thermal stochasticity. Thus continuous values may sometimes eliminate the need for stochastic simulation.

## Activation Vector

The representational features serve as the blackboard on which the cognitive system carries out its computations. The *knowledge* that guides those computations is associated with the second set of entities, the *knowledge atoms*. Each such atom $\alpha$ is characterized by a *knowledge vector* $\mathbf{k}_\alpha$, which is a list of $+1$, $-1$, and 0 values, one for each representation variable $r_i$. This list encodes a piece of knowledge that specifies what value each $r_i$ should have: $+1$, $-1$, or unspecified (0).

Associated with knowledge atom $\alpha$ is its *activation variable*, $a_\alpha$. This variable will also be taken to be binary: 1 will denote active; 0, inactive. Because harmony theory is probabilistic, degrees of activation are represented by varying probability of being active rather than varying values for the activation variable. (Like continuous values for representation variables, continuous values for activation variables could be incorporated into the theory with little difficulty, but a need to do so has not yet arisen.) The list of $\{0,1\}$ values for the activations $\{a_\alpha\}$ comprises the *activation vector* $\mathbf{a}$.

Knowledge atoms encode subpatterns of feature values that occur in the environment. The different frequencies with which various such patterns occur is encoded in the set of *strengths*, $\{\sigma_\alpha\}$, of the atoms.

In the example of qualitative circuit analysis, each knowledge atom records a pattern of qualitative changes in some of the circuit features (currents, voltages, etc.). These patterns are the ones that are consistent with the laws of physics, which are the constraints characterizing the circuit environment. Knowledge of the laws of physics is encoded in the set of knowledge atoms. For example, the atom whose knowledge vector contains all zeroes except those features encoding the pattern $<$*current decreases, voltage decreases, resistance increases*$>$ is one of the atoms encoding qualitative knowledge of Ohm's law. Equally important is the *absence* of an atom like one encoding the pattern $<$*current increases, voltage decreases, resistance increases*$>$, which violates Ohm's law.

There is a very useful graphical representation for knowledge atoms; it was illustrated in Figure 4 and is repeated as Figure 8. The representational features are designated by nodes drawn in a lower layer; the activation variables are depicted by nodes drawn in an upper layer. The connections from an activation variable $a_\alpha$ to the representation variables $\{r_i\}$ show the knowledge vector $\mathbf{k}_\alpha$. When $\mathbf{k}_\alpha$ contains a $+$ or $-$ for $r_i$, the connection between $a_\alpha$ and $r_i$ is labeled with the appropriate sign; when $\mathbf{k}_\alpha$ contains a 0 for $r_i$, the connection between $a_\alpha$ and $r_i$ is omitted.

In Figure 8, all atoms are assumed to have unit strength. In general, different atoms will have different strengths; the strength of each atom

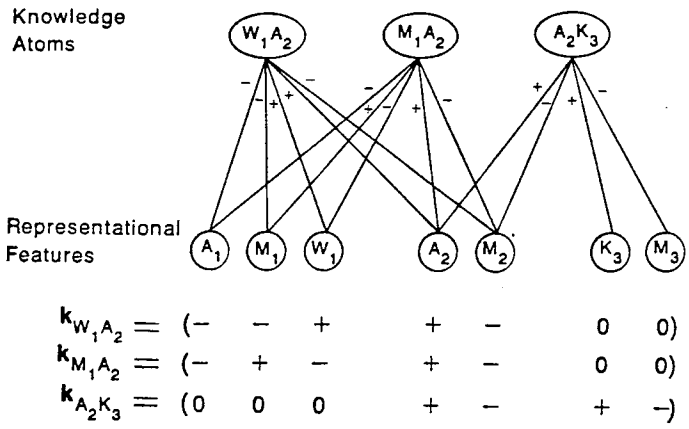


FIGURE 8. The graphical representation of a particular harmony model.

would them be indicated above the atom in the drawing. (For the completely general case, see Figure 13.)

## Hierarchies and the Architecture of Harmony Networks

One of the characteristics that distinguishes harmony models from other parallel network models is that the graph *always* contains two layers of nodes, with rather different semantics. As in many networks, the nodes in the upper layer correspond to patterns of values in the lower layer. In the letter-perception model of McClelland and Rumelhart, for example, the word nodes correspond to patterns over the letter nodes, and the letter nodes in turn correspond to patterns over the line-segment nodes. The letter-perception model is typical in its *hierarchical structure*: The nodes are stratified into a sequence of several layers, with nodes in one layer being connected only to nodes in adjacent layers. Harmony models use only two layers.

The formalism could be extended to many layers, but the use of two layers has a principled foundation in the semantics of these layers. The nodes in the representation layer *support representations of the environment at all levels of abstractness*. In the case of written words, this layer could support representation at the levels of line segments, letters, and words, as shown schematically in Figure 9. The upper, knowledge, layer encodes the patterns among these representations. If information is given about line segments, then some of the knowledge atoms