LOGIC FOR THE NEW AI

Bruce MacLennan
Computer Science Department
Naval Postgraduate School
Monterey, CA 93943

The psyche never thinks without an image.
— Aristotle

I. INTRODUCTION

A. The New AI

There is growing recognition that the traditional methods of artificial intelligence (AI) are inadequate for many tasks requiring machine intelligence.[1] Space prevents more than a brief mention of the issues.

Research in connected speech recognition has shown that an incredible amount of computation is required to identify spoken words. This is because the contemporary approach begins by isolating and classifying phonemes by means of context-free features of the sound stream. Thus the stream must be reduced to acoustic atoms before classification can be accomplished. On the other hand, for people the context determines the phoneme, much as the melody determines the note.[2] This is why a phoneme can vary so greatly in its absolute (i.e., context-free) features and still be recognized. People recognize a gestalt, such as a word or phrase, and identify the phonetic features later if there is some reason to do so (which there's usually not).

---

1. A critique from a (predominantly Heideggerian) phenomenological viewpoint of current AI technology can be found in H. Dreyfus (1979) and H. Dreyfus and S. Dreyfus (1986). In his (1982, pp. 3-27) Hubert Dreyfus claims that current AI technology is making the same fundamental error that Husserl made in his approach to phenomenology, and that it is facing the same "infinite task." Gardner (1985) provides a good overview of the strengths and limitations of cognitive science; much of this applies to contemporary AI. Haugeland (1985) likewise shows AI and cognitive science in their historical context. His (1981) collects important papers pro and con traditional AI methods.

2. See Köhler (1947), p. 118. Dreyfus (1979) quotes Oettinger (1972): "Perhaps ... in perception as well as in conscious scholarly analysis, the phoneme comes after the fact, namely ... it is constructed, if at all, as a *consequence* of perception not as a step in the process of perception itself."

In general, much of the lack of progress in contemporary pattern recognition can be attributed to the attempt to classify by context-free features. If the chosen features are coarse-grained, the result is a "brittleness" in classification and nonrobustness in the face of novelty.[3] On the other hand, if the features are fine-grained, then the system is in danger of being swamped by the computation required for classification.[4] People (and other animals) do not seem to face this dilemma. We recognize the whole, and focus on the part only when necessary for the purpose at hand.[5] The logical atomism of contemporary AI precludes this approach, since wholes can be identified only in terms of their constituents. But, as we've seen, the constituents are determined by the whole. Thus, with contemporary AI technology, pattern recognition faces a fundamental circularity.

A similar problem occurs in robotics. Robotic devices need to coordinate their actions through kinesthetic, visual and other forms of feedback. Contemporary systems attempt to accomplish this by building explicit models of the world in which the system must function. Coordination and planning are accomplished by symbolic manipulation (basically deduction) on the knowledge base. This is subject to the same limitations we've already seen. If the model is simple, then the robot will be unprepared for many exigencies. If the model is extremely detailed, then the system will be swamped by computation — and still face the possibility of unforeseen circumstances. How is it that people and animals avoid this predicament?

Heidegger has shown that much of human behavior exhibits a ready-to-hand understanding of our world that is not easily expressed in propositional form. "When we use a piece of equipment,

---

3. For example, a speech recognition system may have to be trained to a specific voice, and may work poorly if the speaker's voice changes (e.g., when he's agitated).

4. Peirce, a pioneer of symbolic logic, saw that logic machines would be "minutely analytical, breaking up inference into the maximum number of steps." (Goudge, 1969, p. 61). For a calculus of reasoning, efficiency is critical, so we should "seek to reduce the number of individual steps to a minimum." (Goudge, 1969, p. 61)

5. As Wittgenstein (1958) says, "The question 'Is what you see composite?' makes good sense if it is already established what kind of complexity — that is, which particular use of the word — is in question." (§ 47, p. 22)

Heidegger claims, we actualize a bodily skill (which cannot be represented in the mind) in the context of a socially organized nexus of equipment, purposes, and human roles (which cannot be represented as a set of facts)."[6] Such skill is acquired through our successful use of our animal bodies to cope with the physical and social worlds in which we find ourselves. Our use of this knowledge is *unconscious* in that we do not think in terms of propositional rules that are then applied to the situation at hand. "The peculiarity of what is proximally ready-to-hand is that, in its readiness to hand, it must, as it were, withdraw in order to be ready to hand quite authentically."[7]

In contrast, all computer knowledge, at least with current AI technology, is rule based. The knowledge is stored in the form of general rules and schemata, and the computer's "thinking" proceeds by the application of these general rules to particular situations. Furthermore, since the computer has no body with which to interact with the world and since it does not develop in a culture from which it can learn norms of behavior, its knowledge must be acquired either in the form of decontextualized general rules, or by mechanized generalization processes from decontextualized data.[8] As Papert and Minsky[9] have said, "Many problems arise in experiments on machine intelligence because things obvious to any person are not represented in any program."

Combinatorial explosion threatens many other applications of AI technology. For example, automatic theorem provers proceed by blind enumeration of possibilities, possibly guided by

---

6. Dreyfus (1982), p. 21.

7. Heidegger (1962), p. 99. He continues: "That with which our everyday dealings proximally dwell is not the tools themselves. On the contrary, that with which we concern ourselves primarily is the work — that which is to be produced at the time; and this is accordingly ready-to-hand too. The work bears with it that referential totality within which the equipment is encountered." The latter observation is especially important with regard to the issues of context dependence and intentionality, discussed below.

8. "This context and our everyday ways of coping in it are not something we *know* but, as part of our socialization, form the way we *are*." (Dreyfus, 1982, p. 21)

9. M.I.T. Artificial Intelligence Laboratory Memo No. 299 (September 1973), p. 77; quoted in Dreyfus (1979), p. 34.

context-free heuristics. On the other hand, human theorem provers are guided by a contextual sense of relevance, and a sense of similarity to previously accomplished proofs. With current AI technology, automatic deduction cannot take advantage of past experience to guide its search.

The same problem occurs in automatic *induction.* This is relatively simple *if* the system is told in advance the relevant variables. That is, given measurements of a dozen different variables, it's not so hard to find which are related and to conjecture a relationship between them. Unfortunately, scientists face a much harder problem, since the number of possible variables is unlimited, as is the number of their relationships. How then are scientific laws ever discovered? First, prior experience gives scientists a sense of relevance (in the context of their investigations); this guides their search. In addition, human cognition permits scientists to *first* recognize similarities and patterns, and *then* to identify the common features (if any) upon which these similarities and patterns are based.

It has long been recognized that people rarely use language as a logical calculus. As Wittgenstein says, "in philosophy we often *compare* the use of words with games and calculi which have fixed rules, but cannot say that someone who is using language *must* be playing such a game."[10] Rather than being fixed by formal definition, the meanings of words expand and retract as required by context and the particulars of the speech situation. If computers are to be able to understand natural language "as she is spoke," then they too must be able to treat meaning in this context and situation dependent manner — without a combinatorial explosion.

There is evidence[11] that expert behavior is better characterized as automatized *knowledge-how* rather than explicit *knowledge-that.*[12] As we've seen, even in predominantly symbolic activities such as mathematics a primary determinant of the *skill* of the expert is his sense of relevance and

10. Wittgenstein (1958), § 81, p. 38.

11. See H. Dreyfus and S. Dreyfus (1986).

12. For this distinction, see Ryle (1949), chapter II (3).

his ‘‘nose’’ for the right attack on a problem. These automatized responses guide his behavior at each step of the process and avoid the combinatorial explosion characteristic of rule-based systems. What is missing from current AI is an explanation of the *vectors* of gestalt psychology: ‘‘When one grasps a problem situation, its structural features and requirements set up certain strains, stresses, tensions in the thinker. What happens in real thinking is that these strains and stresses are followed up, yield vectors in the direction of improvement of the situation, and change it accordingly.’’[13] As the Dreyfuses note, rule-based behavior with explicit heuristics is more characteristic of advanced beginners than of experts. But — are there alternatives to logical atomism and rule-based behavior?

With current computer and AI technology, it seems unlikely that computers can be made to exhibit the sort of ready-to-hand understanding that people do. Thus contemporary AI emphasizes present-at-hand[14] knowledge that can be expressed in terms of decontextualized concepts and verbal structures. With our present techniques, for the computer to know *how,* it is necessary for it to know *that.*

Traditional logic, of which modern logic and conventional AI technologies are developments, is an idealization of certain cognitive activities that may be loosely characterized as *verbal.* On the other hand, many of the tasks for which we would like to use computers are nonverbal.[15] Seen in this way, it is no surprise that idealized verbal reasoning is inadequate for these tasks — the

---

13. Wertheimer (1959), p. 239.

14. Heidegger distinguishes ready-to-hand and present-at-hand as follows: ‘‘Original familiarity with beings lies in *dealing with them* appropriately. ... The *whatness* of the beings confronting us every day is defined by their equipmental character. The *way* a being with this essential character, equipment, is, we call *being handy* or handiness, which we distinguish from being extant, [present] at hand.’’ (Heidegger, 1982, p. 304) Nature itself can be considered either as ready-to-hand — the characteristic stance of technology — or as present-at-hand — the characteristic stance of science. ‘‘If its kind of Being as ready-to-hand is disregarded, this ‘Nature’ itself can be discovered and defined simply in its pure presence-at-hand.’’ (Heidegger, 1962, p. 100)

15. There is now ample evidence that nonverbal reasoning is an essential part of human (and animal) cognitive activity. See for example Arnheim (1971), Wertheimer (1959), Kosslyn (1980), Shepard (1971, 1975), and Gardner (1985). On the importance of nonverbal thinking in scientific creativity, see Miller (1986).

idealization is too far from the fact. I suggest that AI is being driven by the needs and limitations of its current methods into a new phase which confronts directly the issues of nonverbal reasoning. This new phase, which I refer to as the *new* AI, will broaden AI technology to encompass nonverbal as well as verbal reasoning.

The preceding suggests that the new AI will require a *new* logic to accomplish its goals. This logic will have to be an idealization of nonverbal thinking in much the same way that conventional logic is an idealization of verbal thinking. In this paper I outline the requirements for such a logic and sketch the design of one possible logic that satisfies the requirements.

But isn't verbal thinking inherent in the very word *logic* (*λογική* < *λόγος*)? Can there be such a thing as a nonverbal logic? In the next section I justify my use of the term *logic* to refer to an idealization of nonverbal thinking.

B. Why a New Logic?

1. The Three Roles of Logic

Based on ideas of Peirce[16] I distinguish three different *roles* fulfilled by logic and related subjects (epistemology, mathematics, philosophy of science).

Empirical logic is a nomological account, in the form of *descriptive* laws, of what people actually

---

16. In Peirce's scheme, the mathematics of logic is a subclass of mathematics, which is a science of discovery. Logic is an order within normative science, which is a subclass of philosophy, which is a science of discovery. History of science is a family within history, which is a suborder of descriptive psychics, which is an order in the psychical sciences, which is a subclass of idioscopy, which is a science of discovery. Psychology is also an order of the psychical sciences. "Mathematics studies what is and what is not logically possible, without making itself responsible for its actual existence." One of its branches is the mathematics of logic. "Nomological psychics [psychology] discovers the general elements and laws of mental phenomena." One of the divisions of normative science is logic. "Logic is the theory of self-controlled, or deliberate, thought; and as such, must appeal to ethics for its principles," since ethics "is the theory of self-controlled, or deliberate, conduct." (from *A Syllabus of Certain Topics of Logic,* 1.180-192, quoted in Buchler (1955), p. 60-62). All of the above are sciences of discovery. Within the theoretical sciences of discovery there are three classes: mathematics, philosophy (including normative science) and idioscopy (a term Peirce borrowed from Bentham). The latter is divided into physiognosy (the physical sciences) and psychognosy (the psychical sciences). The differentiae of the latter two are relevant: "Physiognosy sets forth the workings of efficient causation, psychognosy of final causation." (1.239-42, Buchler (1955), p. 67) We might say, physiognosy deals with mechanical laws, psychognosy with intentional laws.

do when they reason (cf., ‘‘epistemology naturalized’’). Thus it can be considered a specialized discipline within psychology or sociology. As such it must account for those patterns of reasoning that are not formally valid, as well as those that are.

Mathematical logic, which is a subdiscipline of mathematics, provides, by means of *formal* laws, an idealized model of the reasoning process. There is no presumption that people do in fact reason this way all the time. Indeed, there's ample evidence that they don't.[17] Of course, for it to be an interesting model of reasoning, it must bear some relationship to actual reasoning. For example, mathematical logic should explain why the actual reasoning processes used by people work when they do. Similarly, mathematical theories of induction should explain why confirmation of unlikely outcomes is more valuable than the confirmation of likely ones, etc.[18] Thus, empirical logic provides the motivation for mathematical logic, and in turn mathematical logic suggests theories that guide the descriptive activities of empirical logicians.

Empirical logic tells us how people in fact reason; mathematical logic explains the validity of various reasoning processes. Neither tells us how we *ought* to reason. This is the role of *normative logic,* which is a subdiscipline of ethics.[19] Whereas empirical logic is formulated in terms of descriptive laws, and mathematical logic in terms of formal laws, normative logic is formulated in terms of *prescriptive* laws.

Normative logic must, of course, draw results from mathematical logic, since the latter explains

---

17. See papers by Johnson-Laird (1970), Wason (1966, 1972), Kahneman, Slovik, and Tversky (1982, 1984).

18. See, for example, G. Polya's (1968).

19. In Peirce's classification there are also Practical Sciences, which presumably include a science of practical logic. In his ‘‘Minute Logic’’ (1.239, quoted by Buchler (1955), p. 66), Peirce distinguishes two branches of science, ‘‘Theoretical, whose purpose is simply and solely knowledge of God's truth; and Practical, for the uses of life.’’ (1.239) In particular, to the three normative sciences, logic, ethics and esthetics, there are three corresponding practical sciences, or arts: the art of reasoning, the conduct of life and fine art. The normative sciences, like the practical, ‘‘study what ought to be, i.e., ideals,’’ but ‘‘they are the very most purely theoretical of purely theoretical sciences.’’ (1.278-82, Buchler (1955), p. 69-70) For Peirce, esthetics is the primary normative science, for it is ‘‘the science of ideals, or of that which is objectively admirable without any ulterior reason.’’ (1.191, Goudge (1969), p. 48) I deviate from Peirce's scheme in that I include the practical sciences under the normative. Thus I do not distinguish between logic as a normative science and logic as a practical science.

why some reasoning processes are valid and others aren't. It must also draw from empirical logic for insights into the psychology of knowing and the practical limitations of human reason. Normative logic may thus prescribe rules that are not mathematically necessary, but that are psychologically or sociologically desirable.

Finally, to the extent to which its norms are followed, normative logic influences the way people actually reason, and hence future descriptive logic. And, to the extent to which mathematical logic models actual reasoning, the normative science also affects the mathematical science.

Logic is important in all three of its roles, but it is its normative role that is ultimately relevant to AI: we want to use logic as a guide for programming intelligent machines. On the other hand, the mathematical role is central, since it forms the core of the normative principles and a standard for empirical studies. Therefore a mathematical logic of nonverbal reasoning must be our first goal, and on this I concentrate in the rest of this paper.

2. Conventional Logic Inadequate for the New AI

Conventional logic — by which we mean any of the well known idealizations of verbal reasoning — is inadequate as a logic for the new AI. I summarize the reasons.

There is now ample evidence[20] that conventional logic is inadequate as an empirical description of the way people actually think. People apparently use a mixture of verbal and nonverbal reasoning that tends to combine the advantages of both. For example, Miller describes the alternation in the uses of imagery and mathematics that led to the development of quantum mechanics.[21] He also describes the nonverbal processes used by Einstein in creating relativity theory; an alternate account of this process is given by Wertheimer.[22]

---

20. See Johnson-Laird (1970), Wason (1966, 1972), and Kahneman, Slovik and Tversky (1982, 1984).

21. See Miller (1986), pp. 125-183.

22. See Wertheimer (1959), Chapter 10.

Gardner summarizes recent empirical studies of the thought processes actually used by people in problem solving situations.[23] They show that the conventional logic is much too idealized. For example, the research of Peter Wason and Philip Johnson-Laird suggests that people are much more likely to reason correctly when the problem has relevance to practical action, than when it is merely abstract. Gardner's work and the references he cites contain additional examples.

Conventional logic is also inadequate as a normative discipline, since it provides standards for verbal reasoning but not for nonverbal reasoning. The costs and benefits and therefore the tradeoffs involved in nonverbal reasoning are different from those of verbal reasoning. Hence the practical guidance provided by the two logics will differ. The "old AI" has been following the norms of conventional logic — and has found their limitations.

Like any mathematical theory, the conventional logic is an idealization. There is nothing wrong with such idealizations, so long as they are appropriate. Unfortunately, conventional logic's idealization of verbal reasoning is often inappropriate to nonverbal problems. It often leads to a discrete, atomistic approach that results in a combinatorial explosion of possibilities.

3. Potential Value of a New Logic

There are many potential benefits that we may expect from a logic for the new AI. A mathematical theory would provide idealizations of the processes involved in nonverbal thinking. As such it would provide a basis for structuring empirical investigations of nonverbal thinking, and a standard upon which to base the norms of nonverbal thinking. The new AI will benefit directly from the normative science, since it is this science that will supply the guidelines for the design of machines that "think" nonverbally.[24] Thus AI should be helped to pass beyond its current difficulties and achieve some of the goals that have eluded the "old" AI.

---

23. See Gardner (1985), Chapter 13.

24. It seems likely that the new AI will bring with it a rebirth of interest in analog computation. Current research on analog, molecular and hybrid optical computers is perhaps a harbinger.

II. REQUIREMENTS FOR A NEW LOGIC

In this section I review some of the processes of (human and animal) thought that, although poorly modeled by conventional logic, are essential to many present and future applications of AI. The new AI must provide a logic geared to the description and analysis of these processes. They are the basis of the criteria by which our own proposal for a new logic should be evaluated.

A. Indefinite Classification

Most, perhaps all, of the concepts that we use in everyday life have *indefinite boundaries.* That is, the presence of borderline cases is the rule rather than the exception. There are several reasons that we should expect this. First, many of the phenomena of nature that are important for survival are continuous. Thus it is natural to expect animal life to have evolved cognitive means for dealing with continuously variable qualities. Second, perception is subject to noise and error, caused both by imperfections in the sense organs and by circumstances in the environment. Survival requires that perception be robust in the face of a wide variety of disturbances. Third, indefinite boundaries avoid the ‘‘brittleness’’ associated with definite boundaries. What is ‘‘brittleness’’? Suppose I have a rule: ‘‘Flee from predators bigger than myself.’’ It seems unreasonable — that is, anti-survival — to treat a predator one millimeter shorter than myself as though this rule is inapplicable. Animal cognition avoids ‘‘brittleness’’ — a thing doesn’t cease being a threat just because it’s one millimeter too short.

It seems that indefinite classification will be as important for intelligent machines as it is for intelligent life. This does not imply that there is no need for definite boundaries. But for many purposes, especially the ones normally classified as ‘‘everyday’’ and hence typical of the new AI, indefinite classes will be required.

To reiterate, indefinite classes are often preferable to definite classes. Recall Wittgenstein’s game example: ‘‘One might say that the concept ‘game’ is a concept with blurred edges. — ‘But is a blurred concept a concept at all?' — Is an indistinct photograph a picture at all? Is it even an advantage to replace an indistinct picture by a sharp one? Isn't the indistinct one often exactly what we need?"[25]

Indefiniteness should not be considered a defect. As Wittgenstein points out, "one pace" is a perfectly useful measurement, despite the absence of a formal definition. "But this is not ignorance. We do not know the boundaries because none have been drawn. To repeat, we can draw a boundary for a special purpose."[26] We expect as much from our new logic — it should be capable of operating with or without boundaries, as the situation requires.

The reader will no doubt think of Zadeh's *fuzzy set theory.*[27] Although this is a step in the right direction, I do not think Zadeh's proposal goes far enough. Some of its limitations for the present purpose will become clear below.

B. Context Sensitivity

The indefiniteness of the boundaries of a concept is dependent on the context in which it is being used. Thus 'pure water' means one thing when I am thirsty in the woods and another when I'm serving my guests — or working in a chemistry lab. Many of our rules — heuristic or otherwise — are couched in context-dependent words and phrases: 'too near', 'dangerous', 'acceptable', 'untrustworthy', etc. etc.

Human (and animal) use of context-sensitive abstractions gives flexibility to the rules that use them. Context sensitivity seems a prerequisite of the intelligent use of rules. Indeed, isn't the blind following of rules — independent of context — the principal example of *stupidity?*

Here we can see the limitations of a fuzzy set theory that attaches a fixed membership distribution

---

25. Wittgenstein (1958), § 71, p. 34.

26. Wittgenstein (1958), § 69, p. 33.

27. See for example Zadeh (1965, 1975, 1983). Kickert (1978) has a good summary of the postulates of fuzzy set theory. See also Goguen (1969).

to a class. To achieve the flexibility characteristic of animal behavior, it's necessary to have this distribution adjust in a manner appropriate to the context.[28]

Context is not something that can be added onto an otherwise context-free concept. Rather, all our concepts are context-dependent; the notion of a decontextualized concept is an idealization formed by abstraction from contextual concepts in a wide variety of contexts. Unfortunately, when it comes to modeling commonsense intelligence the idealization is too far from reality, "because everything in this world presents itself in context and is modulated by that context."[29] In Heideggerian terms, we are "always already in a situation."

The above observations also apply to activities that hold definite-boundaried, context-free abstractions as an ideal. The prime example is mathematics; here the abstractions are all defined formally. Observe, however, that many of the concepts — such as rigor — that guide mathematical behavior have just the indefinite, contextual character that I've described. Is this not the reason that computers, the paragons of formal symbol manipulators, are so poor at doing mathematics?

C. Logical Holism

The use of conventional logic drives AI to a kind of logical atomism. That is, all classification is done on the basis of a number of "atomic" features that are specified in advance. Whether these features have definite boundaries or are "fuzzy" is not the issue. Rather, the issue is whether it is possible to specify in advance (i.e., independent of context) the essential properties of a universal.

Wittgenstein, in *Philosophical Investigations* (1958), has criticized the notion of essential attributes and the basis for logical atomism. He observes that "these phenomena have no one thing in common which makes us use the same word for all,— but that they are *related* to one

---

28. See the 'pure water' example above. Note also that the context itself is indefinite boundaried. What if I'm serving guests at my campsite?

29. Arnheim (1971), p. 37.

another in many different ways."[30] Further, he instructs us to *"look and see* whether there is anything common to all. — For if you look at them you will not see something that is common to *all,* but similarities, relationships, and a whole series of them at that."[31] This is what gives flexibility and adaptability to human classification; we are not dependent on a particular set of "essential" attributes. If we come upon a sport lacking one of the essentials, it will still be recognized, if it's sufficiently similar. "Is it not the case that I have, so to speak, a whole series of props in readiness, and am ready to lean on one if another should be taken from under me and vice versa?"[32]

What is the alternative to classification by essentials? "How then is it possible to perform an abstraction without extracting common elements, identically contained in all particular instances? It can be done when certain aspects of the particulars are perceived as deviations from, or deformations of, an underlying structure that is visible within them."[33] This is in accord with Eleanor Rosch's research, in which she concludes, "Many experiments have shown that categories appear to be coded in the mind neither by means of lists of each individual member of the category, nor by means of a list of formal criteria necessary and sufficient for category membership, but, rather, in terms of a prototype of a typical category member. The most cognitively economical code for a category is, in fact, a *concrete image* of an average category member."[34] Kuhn sees this as the usual pattern of science: "The practice of normal science depends on the ability, acquired from exemplars, to group objects and situations into similarity sets which are primitive in the sense that the grouping is done without an answer to the question, 'Similar with respect to what?' "[35] In contrast to logical atomism, he emphasizes this is "a

---

30. Wittgenstein (1958), § 65, p. 31.

31. Wittgenstein (1958), § 66, p. 31.

32. Wittgenstein (1958), § 79, p. 37.

33. Arnheim (1971), p.49.

34. Rosch (1977), p. 30. See also Rosch (1978), and Armstrong, Gleitman and Gleitman (1983).

manner of knowing which is misconstrued if reconstructed in terms of rules that are first abstracted from exemplars and thereafter function in their stead."[36]

Classification by family resemblance avoids two problems characteristic of classification by context-free features. First, classification by context-free features can be inflexible, since the number of features used is by necessity limited. Classification by family resemblance is more flexible because of the open-ended set of features upon which the classification is based. On the other hand, if we attempt to improve the flexibility of context-free classification by classifying on the basis of more context-free attributes, the efficiency of the process is much degraded, since the computer must consider all the attributes. Classification by family resemblance naturally focuses on those attributes likely to be relevant in the life of the person. This improves the efficiency of human cognitive processing.

Classification by family resemblance may in part account for animals' ability to adapt to novel situations. An object may be judged as belonging to a certain class in spite of the fact that it lacks certain "essential" characteristics, provided that it satisfies the overall gestalt. That is, in a given context certain attributes may be more relevant than others.

D. Intentionality

Another characteristic of human cognition that should be accounted for by our logic is *intentionality,* the directedness of consciousness towards its objects.[37] The effect of intentionality is to restrict awareness to just those aspects of the environment that are likely to be relevant to the problem at hand. It shifts some things into the foreground so that the rest can be left in the background. If we think of the foreground as having a high probability of being considered and

---

35. Kuhn (1970), p. 200.

36. Kuhn (1970), p. 192.

37. We use this term in Brentano's and Husserl's sense, i.e., "the unique peculiarity of experiences 'to be the consciousness *of* something.' " (Husserl, 1962, §84, p. 223)

the background as having a low probability, then the effect of this focusing process is to decrease the entropy of the probability distribution. Indeed, the functions that Peirce attributes to consciousness are self-control and improving the efficiency of habit formation.[38] This suggests that we can get the beneficial effect of intentionality by any process that on the average skews the probability of processing in favor of the more relevant information.

We see that both classification by family resemblances and intentionality improve cognitive efficiency by focusing cognitive activity on factors likely to be relevant. Our goal is to program computers to have a sense of relevance.

E. Mixed-Mode Reasoning

The issues discussed so far suggest that the new logic be based on a continuous (or analog) versus discrete (or digital) computational metaphor. Yet the limitations of analog computation are well known. For example, errors can accumulate at each stage of an analog computational process to the extent to which all accuracy is lost.[39] This suggests that analog (continuous) reasoning cannot be as *deep* — support as long chains of inference — as digital (discrete) reasoning.

Introspection suggests a solution to this problem. Based on the current context and the measures of relevance it induces, we "digitize" much of our mental experience — we verbalize our mental images. Such verbalization permits longer chains of inference by preventing the accumulation of error. It is successful so long as the context is relatively stable. We expect a logic for the new AI to accommodate verbal as well as nonverbal reasoning, and to permit the optimal mix of the two to be determined.

Paivio's remarks on visual cognition apply as well to other forms of nonverbal reasoning: "Images and verbal processes are viewed as alternative coding systems, or modes of symbolic

---

38. Goudge (1969), p. 235; see also Tiercelin (1984).

39. I am grateful to R. W. Hamming for alerting me to this limitation of analog computation.

representation, which are developmentally linked to experiences with concrete objects and events as well as with language."[40] But, the two systems are not mutually exclusive: "Many situations likely involve an interaction of imaginal and verbal processes, however, and the latter would necessarily be involved at some stage whenever the stimuli or responses, or both, are verbal ...."[41]

The key point is that verbal thinking is really a special case of nonverbal, since "language is a set of perceptual shapes — auditory, kinesthetic, visual."[42] The relative definiteness of linguistic symbols stabilizes nonverbal thinking. "Purely verbal thinking is the prototype of thoughtless thinking, the automatic recourse to connections retrieved from storage. What makes language so valuable for thinking, then, cannot be thinking in words. It must be the help that words lend to thinking while it operates in a more appropriate medium, such as visual imagery."[43]

III. PRELIMINARY DEVELOPMENT OF NEW LOGIC

A. Approach

In this section we outline the general framework for our model for nonverbal reasoning. Recall however that it is not our intention to develop a psychological theory; that is, our goal is not descriptive. Rather, our goal is to develop an idealized theoretical model of certain functions of nonverbal mental activity, much as Boolean algebra and predicate logic are idealized models of verbal cognition. But there is also a normative goal; the resulting logic should be useful as a tool for designing and programming computers.

Consider all the neurons that comprise a nervous system.[44] We postulate that there are two distinct ways in which they can encode information. *Semipermanent* information is in some way

---

40. Paivio (1979), p. 8.

41. Paivio (1979), p. 9. See also Miller (1986), especially chapter 4, for a discussion of the interplay of verbal and nonverbal reasoning.

42. Arnheim (1971), p. 229.

43. Arnheim (1971), pp. 231-232.

encoded in the neural structure (e.g., in terms of strength of synaptic connection). Transient information is encoded in a way that does not alter the neural structure, (e.g., dynamic electrochemical activity). Transient information processing is involved in processes such as associative recall; alteration of semipermanent information is involved in processes such as learning.[45] We refer to transient information encoded in the electrochemical state of a neuron as the *state* of that neuron. Semipermanent information will be described in terms of *memory traces.*

Neurons can be divided into three categories on the basis of their connections to each other and to nonneural structures. We call neurons *afferent* if their state is determined *solely* by nonneural mechanisms, such as sense organs. We call neurons *efferent* if they have no effect on other neurons; that is, their state affects only nonneural mechanisms, such as motor organs. The remaining neurons we call *interneurons;* they both affect and can be affected by other neurons.[46]

We call the set of all afferent neurons the *afferent system.* Analogously we define the efferent and interneural systems. Since at any given time each neuron is in a state, at any given time each of these systems is in a state. Thus we can speak of the state of the afferent system, etc. To describe the possible states of these systems we define three spaces, A, I and E, the set of possible states of the afferent, interneural and efferent systems.

We will usually not need to distinguish the efferent neurons from other nonafferent neurons. Therefore we define $B = I \times E$, the space of all states of the nonafferent system.

---

44. The reader will observe that we use terminology inspired by neuropsychology. This should not be interpreted as implying that we are offering a theory of brain function. It is simply the case that we have found considerations of brain organization to be helpful in developing the theory.

45. Although, as will become apparent later, our model permits the possibility that *all* processes have some effect on semipermanent information.

46. Our definitions our inspired by, but, in the spirit of idealization, not identical with those common in neuropsychology. Kolb and Whishaw (1985), define afferent as ‘‘[c]onducting toward the central nervous system or toward its higher centers,’’ efferent as ‘‘[c]onducting away from higher centers in the central nervous system and toward a muscle or gland,’’ and interneuron as ‘‘[a]ny neuron lying between a sensory neuron and a motor neuron.’’

What properties can we expect of the spaces A and B? Taking the points in these spaces to represent neural states, it is reasonable to assume that there is some notion of "closeness" between these states. Therefore, for any two points $a, a' \in A$ we postulate a distance $\delta_A(a, a')$ such that (1) $\delta_A(a, a') \geq 0$, and (2) $\delta_A(a, a') = 0$ if and only if $a = a'$. That is, distance is a nonnegative number such that the distance between two neural states is zero if and only if the states are identical. Such a function is commonly called a *semimetric* on A. Similarly we postulate a semimetric $\delta_B(b, b')$ for $b, b' \in B$. Furthermore, we will drop the subscripts and write $\delta(a, a')$ and $\delta(b, b')$ when no confusion will result.

It seems reasonable that neural states cannot be infinitely different. Therefore we make an additional assumption, that the semimetrics are *bounded.* This means that there is some number Δ_A such that $\delta(a, a') \leq \Delta_A$ for all $a, a' \in A$. Similarly there is a bound Δ_B on distances in B. Without loss of generality take $\Delta_A = \Delta_B = 1$ (this only changes the distance scale). Thus

$$\delta_A \colon A^2 \to [0, 1], \quad \delta_B \colon B^2 \to [0, 1]$$

The above assumptions can be summarized by saying that A and B are bounded semimetric spaces.

The functions δ represent the *difference* between neural states. It will generally be more intuitive to work in terms of the *similarity* between states. Hence we define

$$\sigma(x, x') = 1 - \delta(x, x') \qquad -$$

A and B subscripts will be added as needed to make the space clear. Note that σ inherits from δ the following properties:

$$0 \leq \sigma(x, x') \leq 1, \qquad \sigma(x, x') = 1 \text{ if and only if } x = x' \qquad -$$

Thus $\sigma(x, x')$ ranges from 1, meaning that x and x' are identical, to 0, meaning that they're as different as they can be.

A final assumption that we make about the spaces A and B is that they are *dense.* That is, for every $\zeta < 1$ and $a \in A$, there is an $a' \in A$, $a \neq a'$, such that $\sigma(a, a') > \zeta$. That means, for every state of the afferent system, there is at least one different state that is arbitrarily similar. The same applies to the space B. These assumptions guarantee that the afferent and nonafferent systems can respond continuously, which seems reasonable, at least as an idealization.

We next define a number of functions that will be useful in the following development. A set of points in a bounded semimetric space can be characterized in terms of their minimum similarity:

$$\mu(S) = \min \{ \sigma(x, y) \mid x, y \in S \}$$

This can vary from 1 for a singleton set to 0 for a set containing at least two dissimilar elements.[47]

A useful quantity for the following derivations is $\sigma_S(x)$, the similarity of x to the other points in the set S. It is defined

$$\sigma_S(x) = \sum_{y \in S} \sigma(x, y)$$

Note that $0 \le \sigma_S(x) \le |S|$, where $|S|$ is the cardinality of S. Thus $\sigma_S(x) = 0$ if x is completely dissimilar from all the members of S.

Sometimes it is more meaningful to work in terms of the average similarity of a point to a set of points:

$$\hat{\sigma}_S(x) = \sigma_S(x) / |S|$$

Thus $0 \le \hat{\sigma}_S(x) \le 1$.

We will refer to the metrics δ_A and δ_B as the *physical metrics* on the afferent and nonafferent spaces because they are directly related to the neural structure. A major goal of the following theory is to show that the structure of the mental content does not depend strongly on the physical metrics. That is, we will attempt to show that the metrics induced by the mental content are "stronger" than the physical metrics.

Under the reasonable assumption that the neurons are the basis for cognitive processes, it makes sense to use the physical metrics as the basis of association. That is, things which cause the same

47. Note that the minimum similarity is just 1 minus the radius of the set (i.e., maximum distance in the set).

pattern of neurological stimulation are in fact indistinguishable. Further, we will make continuity assumptions: things which cause nearly the same pattern of stimulation should lead to nearly the same response. These will be introduced when needed, rather than here, so that it is more apparent which assumptions are necessary for which results.

In the following two sections we investigate logics based on the preceding postulates. The first logic is based on a finite number of memory traces formed at discrete instants of time. This is only a partial step to a logic that satisfies the requirements in Part II, but it is useful to build intuition. The second logic takes a further step in the required direction by postulating a single memory trace that evolves continuously in time. Both theories are very tentative, but they nevertheless should indicate my expectations for a logic for the new AI.

B. Discrete Time

1. Definitions

The first version of our logic will be based on a discrete time model. Thus the state transition process is discontinuous. We imagine the afferent system taking on a series of states (stimuli) s_1, s_2, s_3, ... under the control of external events. These stimuli drive the nonafferent system through a series of states (responses) r_1, r_2, r_3, ... That is, the response r, or new state of the nonafferent system, is a function of (1) the stimulus s, or current afferent state, and (2) the context c, or current nonafferent state.[48] We write $r = s : c$ to denote that r is a new nonafferent state resulting from the afferent state s and the nonafferent state c. Thus $s : c$ is the response resulting from the stimulus s in the context c.

An assumption we make here is that the semipermanent information is constant throughout this process (i.e., no learning takes place). We consider later the case where, in addition to a state

48. Actually, as will be explained shortly, it is a multiple-valued function and thus not, in the technical sense, a function.

transition, we have an alteration of memory traces. So long as the semipermanent information is fixed, the responses to the stimuli s_1, s_2, s_3, ... are

$$s_1 : c, \quad s_2 : (s_1 : c), \quad s_3 : [s_2 : (s_1 : c)], \ldots$$

That is, the response of each stimulus becomes the context for the next, $c_{i+1} = r_i = s_i : c_i$. Thus, excluding learning, we have a notation for the context-dependent interpretation of sensory data.

The new state is obviously dependent on the semipermanent information stored in the memory. Therefore we postulate that at any given time the memory M contains a finite number K of traces, (s_i, c_i, r_i), for $1 \le i \le K$.[49] These traces reflect situations in the past in which stimulus s_i in context c_i produced response r_i.

We expect that the response to a stimulus s in a context c will be dependent on the similarity of s and c to the stored s_i and c_i. Therefore, for fixed s and c we define $\alpha_i = \sigma(s, s_i)$ and $\beta_i = \sigma(c, c_i)$. Thus α_i is the similarity of the present stimulus to the ith stored stimulus, and β_i is the similarity of the present context to the ith stored context.

We expect the activation of memory traces to be related directly to the similarity of the present stimulus and context to the stimulus and context of the trace. On the other hand, we do not wish to make a commitment to a *particular* relationship. Thus we postulate a function $\Gamma(s_i, c_i)$, monotonically increasing in both its arguments, but otherwise unspecified. For fixed s and c we then define $\gamma_i = \Gamma(s_i, c_i)$. By monotonically increasing we mean that if $\alpha_i > \alpha_j$ and $\beta_i = \beta_j$, or if $\alpha_i = \alpha_j$ and $\beta_i > \beta_j$, then $\gamma_i \ge \gamma_j$. Thus γ_i measures the similarity of the current stimulus/context to the ith memory trace. Without loss of generality we take $\Gamma\colon A \times B \to [0, 1]$, that is, $0 \le \gamma_i \le 1$.

The monotonicity condition on Γ tells us that if s is more similar to s_i than to s_j, but c is equally similar to c_i and c_j, then $\gamma_i \ge \gamma_j$. Similarly, if c is more similar to c_i than to c_j, but s is equally similar to s_i and s_j, then $\gamma_i \ge \gamma_j$. Thus the similarity of a current state (s, c) to the stored states

49. The number K is not fixed, but increases as new traces are made in the memory. However, for the analysis of state (i.e. transient information) changes, it can be taken as constant.

(s_i, c_i) is a function of both the similarity of the stimuli and the similarity of the contexts. This will be the basis for context-sensitive classification.

We will need to be able to compare various responses in terms of their similarity to the stored responses, weighted by the similarity of the current stimulus/context to the corresponding stored stimulus/context. For this purpose we define $\varsigma(r)$, the weighted similarity of r to the responses in memory:

$$\varsigma(r) = \sum_{i=1}^{K} \gamma_i \, \sigma(r, r_i)$$

This is a kind of "score" for r, since $\varsigma(r)$ will tend to be high when r is close to the reponses of those traces that are most strongly activated by the current stimulus/context.

2. State Transition Function

Given the foregoing definitions it is easy to describe the state transition function. We are given a memory M consisting of the traces (s_i, c_i, r_i), with $1 \le i \le K$. Given the stimulus/state pair (s, c) we want the new state to be a r that is maximally similar to the r_i, but weighted according to the similarity of (s, c) to the corresponding (s_i, c_i). That is, we want $\varsigma(r)$ to be a maximum over the space B. Hence, we define the state transition operation:

$$s : c = \boldsymbol{\varepsilon} r[\varsigma(r) = \max \{\varsigma(r) \mid r \in B\}]$$

Here we have used Russell's *indefinite description* operation $\boldsymbol{\varepsilon}$. The definition can be read, "$s : c$ is any r such that the weighted similarity of r is the maximum of the weighted similarities of all the points in B."

Notice that, as implied by the use of the indefinite description operator, there may not be a unique point that maximizes ς. Thus there may be bifurcations in the state transition histories. They are in effect gestalt switches between equally good interpretations of the sensory input.

3. Activation of a Single Trace

Our first example is to show how a memory trace can be activated by a stimulus that's sufficiently close. In particular we assume[50] that the stimulus/state pair (s, c) is similar to the stored pair (s_1, c_1) but different from all the other pairs, (s_i, c_i), $i \neq 1$. We intend to show that the induced state $r = s{:}\,c$ is similar to r_1.

Since (s, c) is close to (s_1, c_1), take $\gamma_1 = \zeta \approx 1$. Since (s, c) is not similar to any of the other (s_i, c_i), take $\gamma_i < \varepsilon \approx 0$, for $i \neq 1$. Our goal is to show that $\sigma(r, r_1) \approx 1$.

Since r maximizes ς, we know $\varsigma(r_j) \leq \varsigma(r)$ for all responses r_j in memory. In particular $\varsigma(r_1) \leq \varsigma(r)$. Now note:

$$
\begin{aligned}
\varsigma(r_1) &= \sum_{i=1}^{K} \gamma_i\, \sigma(r_1, r_i) \\
&= \gamma_1\, \sigma(r_1, r_1) + \sum_{i=2}^{K} \gamma_i\, \sigma(r_1, r_i) \\
&= \zeta + \sum_{i=2}^{K} \gamma_i\, \sigma(r_1, r_i) \\
&\geq \zeta
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\varsigma(r) &= \sum_{i=1}^{K} \gamma_i\, \sigma(r, r_i) \\
&= \gamma_1\, \sigma(r, r_1) + \sum_{i=2}^{K} \gamma_i\, \sigma(r, r_i) \\
&= \zeta\, \sigma(r, r_1) + \sum_{i=2}^{K} \gamma_i\, \sigma(r, r_i) \\
&< \zeta\, \sigma(r, r_1) + \varepsilon \sum_{i=2}^{K} \sigma(r, r_i)
\end{aligned}
$$

The inequality follows from $\gamma_i < \varepsilon$, for $i \neq 1$. Now, since $\sigma(r, r_i) \leq 1$ always, we have $\varsigma(r) < \zeta\, \sigma(r, r_1) + \varepsilon(K - 1)$.

We know that the transition operation $s{:}\,c$ maximizes $\varsigma(r)$, so we know $\varsigma(r_1) \leq \varsigma(r)$. Hence, combining the three inequalities we have $\zeta\, \sigma(r, r_1) + \varepsilon(K - 1) > \zeta$. Solving for $\sigma(r, r_1)$ yields

---

50. There is no loss in generality in taking the index of the similar pair to be 1.

$\sigma(r, r_1) > 1 - (\varepsilon / \zeta)(K-1)$. Hence $\sigma(r, r_1)$ differs from 1 by at most the amount $(\varepsilon / \zeta)(K-1)$. This quantity reflects the extent by which the other memory traces interfere with perfect recall of r_1. Note that as $\varepsilon \to 0$ this quantity approaches zero. That is,

$$\lim_{\varepsilon \to 0} \sigma(r, r_1) = 1$$

Hence, as the interference approaches 0 the recall approaches perfect.

4. Activation of a Cluster of Traces

We now perform very much the same analysis in the situation in which (s, c) closely matches a number of traces in memory, all of which induce a similar response. Thus, suppose that H and $\bar{H}$ partition the indices $\{1, 2, \ldots, K\}$. Suppose that $\gamma_i > \zeta \approx 1$ for $i \in H$, and that $\gamma_i < \varepsilon \approx 0$ for $i \in \bar{H}$. Assuming all the matching traces induce a similar state transition implies that $\mu(H) = \eta \approx 1$. As before compute

$$\varsigma(r) \;=\; \sum_{i=1}^{K} \gamma_i\, \sigma(r, r_i) \;<\; \sum_{i \in H} \sigma(r, r_i) + \varepsilon \sum_{i \in \bar{H}} \sigma(r, r_i) \;\le\; \sigma_H(r) + \varepsilon|\bar{H}|$$

Hence $\varsigma(r) < \sigma_H(r) + \varepsilon|\bar{H}|$. For an arbitrary r_j we derive

$$\begin{aligned} \varsigma(r_j) &= \sum_{i \in H} \gamma_i\, \sigma(r_j, r_i) + \sum_{i \in \bar{H}} \gamma_i\, \sigma(r_j, r_i) \\ &> \zeta \sum_{i \in H} \sigma(r_j, r_i) + \sum_{i \in \bar{H}} \gamma_i\, \sigma(r_j, r_i) \\ &\ge \zeta \sum_{i \in H} \mu(H) \\ &= \zeta\, \eta\, |H| \end{aligned}$$

Also, since $r = s{:}c$ maximizes ς, We know that $\varsigma(r) \ge \varsigma(r_j)$, for all j, $1 \le j \le K$. Hence, combining the inequalities, $\sigma_H(r) + \varepsilon\,|\bar{H}| > \zeta\,\eta\,|H|$. Rearranging terms: $\sigma_H(r) > \zeta\,\eta\,|H| - \varepsilon\,|\bar{H}|$. Now, since $\hat{\sigma}_H(r) = |H|^{-1}\sigma_H(r)$,

$$\hat{\sigma}_H(r) > \zeta\,\eta - \varepsilon\,|\bar{H}| / |H|$$

Notice that as $\varepsilon \to 0$, $\varepsilon|\bar{H}|/|H| \to 0$. That is, the interference approaches zero as ε approaches zero.[51] Considering the limit as $\varepsilon \to 0$ and $\zeta \to 1$,

51. Note that $|\bar{H}| / |H|$ is an (inverse) measure of the number of activated traces.

$$\lim_{\substack{\varepsilon \to 0 \\ \zeta \to 1}} \hat{\sigma}_H(r) \quad \geq \quad \eta \quad \approx \quad 1$$

Hence, the recall is arbitrarily similar to the remembered responses, to the extent that the remembered responses are similar.

5. Family Resemblances

We expect a new logic to be able to describe the way in which universals are abstracted from concretes without an analysis in terms of features. Therefore we investigate how, in a fixed context c, a number of stimulus/response pairs define an abstraction. We assume that both positive examples E and negative examples $\bar{E}$ are provided, and that the response to the positive examples is r^+ and to the negative examples is r^-. The entire classification process is meaningless if the responses are the same, so we assume $\sigma(r^+, r^-) = \varepsilon \approx 0$. For convenience we let $\sigma^+ = \sigma(r, r^+)$ and $\sigma^- = \sigma(r, r^-)$; our goal is to show that $\sigma^+ \approx 1$ and $\sigma^- \approx 0$ for positive stimuli, and vice versa for negative.

We aim to show that the exemplars define an abstraction to which various stimuli belong or don't belong to the extent that they are similar to positive or negative examples. Therefore, every stimulus $s \in A$ can be characterized by its similarity to the positive and negative exemplars (in context). We define a parameter ρ that measures the similarity to positive exemplars relative to the similarity to negative exemplars: $\rho \;=\; \sum_{i \in E} \gamma_i \;\, / \,\; \sum_{i \in \bar{E}} \gamma_i$. We then relate to ρ the similarity of the response r to the trained responses r^+ and r^-.

First we compute the weighted similarity of the response to the stored responses:

$$\varsigma(r) \;=\; \sum_{i \in E} \gamma_i \sigma(r, r^+) + \sum_{i \in \bar{E}} \gamma_i \sigma(r, r^-) \;=\; \sigma^+ \sum_E \gamma_i + \sigma^- \sum_{\bar{E}} \gamma_i$$

Similarly, $\varsigma(r^+) \;=\; \sum_E \gamma_i + \varepsilon \sum_{\bar{E}} \gamma_i$. Combining via the inequality $\varsigma(r) \geq \varsigma(r^+)$ and solving for σ^+ yields:

$$\sigma^+ \;\geq\; 1 - (\sigma^- - \varepsilon) \sum_{\bar{E}} \gamma_i \,/ \sum_E \gamma_i \;\geq\; 1 - (1-\varepsilon)/\rho \;\geq\; 1 - 1/\rho$$

This then is our first result:

$$\sigma(r, r^+) \;\geq\; 1 - \rho^{-1}$$

Hence, the correctness of the response improves to the extent that the stimulus is more similar to the positive than to the negative exemplars. A symmetric analysis allows us determine the similarity of the response to r^-:

$$\sigma(r, r^-) \;\geq\; 1 - \rho$$

Hence, the reponse approaches r^- as the stimulus becomes more similar to the negative exemplars.

The preceding two results do not complete the analysis. They show that the response is similar to the appropriate correct response, but not that it is dissimilar to the appropriate incorrect response. Unfortunately, our current assumption, that $1 - \sigma(x, y)$ is a semimetric, does not guarantee this result. We could have the case that r^+ and r^- are very dissimilar, yet r is very similar to both; this is permitted by a semimetric. On the other hand, it certainly seems unintuitive to have a neural state that is simultaneously similar to two dissimilar neural states. Since the issue is unresolved, we introduce the additional assumption only here, where it is required.

A semimetric $\delta\colon S \times S \to R$ is a *metric* if is satisfies the *triangle inequality:*

$$\delta(x, z) \;\leq\; \delta(x, y) + \delta(y, z), \text{ for all } x, y, z \in S$$

From this we immediately derive a triangle inequality for similarity metrics:

$$\sigma(x, y) + \sigma(y, z) \;\leq\; 1 + \sigma(x, z)$$

We return to the analysis of family resemblance.

The triangle inequality tells us that

$$\sigma(r, r^+) + \sigma(r, r^-) \;\leq\; 1 + \sigma(r^+, r^-) \;=\; 1 + \varepsilon$$

Therefore,

$$\sigma(r, r^+) \le 1 + \varepsilon - \sigma(r, r^-) \le 1 + \varepsilon - (1 - \rho) = \varepsilon + \rho$$

since $\sigma(r, r^-) \ge 1 - \rho$. Similarly it is easy to show that $\sigma(r, r^-) \le \varepsilon + \rho^{-1}$. Thus, given the triangle inequality, we can derive the other two inequalities that characterize family resemblance:

$$\sigma(r, r^+) \le \varepsilon + \rho, \quad \sigma(r, r^-) \le \varepsilon + \rho^{-1}$$

Combining these with the previous two inequalities gives bounds on the similarity of the response to the two possible correct responses:

$$\sigma(r, r^+) \in [\varepsilon + \rho, 1 - \rho^{-1}], \quad \sigma(r, r^-) \in [\varepsilon + \rho^{-1}, 1 - \rho]$$

Hence, as expected, the response r reflects the similarity of the stimulus s to the positive and negative exemplars.

Note that ρ is defined in terms of the γ_i, which in turn depend on the similarities of the current stimulus s to the stored stimuli s_i and on the similarities of the current context c to the stored contexts c_i. Thus ρ is a function, $\rho\,(s, c)$, of both the current stimulus and the current context. In effect the positive and negative exemplars induce a *potential field* $\rho(s, c)$, which defines for each stimulus s in context c the extent to which it resembles the positive exemplars and is different from the negative exemplars (in their contexts). Thus ρ represents an abstraction from the examples, that is both indefinite boundaried and context-sensitive. Also note that the classification is not based on any specific features of the stimulus. On the contrary, since the state spaces are dense, we can "teach" an arbitrary abstraction by a suitable presentation of positive and negative cases.

We will say little else about learning at this time, except to note that present behavior can be habituated by adding each triple (s, c, r) to the memory at the end of the corresponding state transition.[52] Such a model is surely oversimplified; we investigate a slightly more sophisticated one below, in the context of continuous time.

---

52. Thus K increases with each state transition.

C. Continuous Time

1. Definitions

We now turn to the case in which time is taken to be continuous and the contents of memory are taken to be a single continuous trace. That is, we take the input stimulus s_t to be a continuous function of the time t, as is the induced response r_t. Similarly, the memory trace (S_x, C_x, R_x) is taken to be a continuous function of a continuous index variable $0 \le x \le K$ (analogous to i, $1 \le i \le K$, in the discrete logic). We now must say what we mean by continuity.

Suppose that S and T are bounded metric spaces with similarity metrics σ_S and σ_T. Further suppose that f is a function from S to T. Then we say that f is continuous at a point $p \in S$ if and only if for all ζ with $0 \le \zeta < 1$ there is an η with $0 \le \eta < 1$ such that $\sigma_T[f(p), f(q)] > \zeta$ whenever $\sigma_S(p, q) > \eta$ and $q \in f[S]$. That is, we can make $f(p)$ and $f(q)$ arbitrary similar by making p and q sufficiently similar.

For the sake of the continuous logic I postulate that s_t and r_t are continuous at all $t \ge 0$, and $S_{x,}$ C_x and R_x are continuous at all $x \ge 0$. Hence, I am assuming that the state of the neural systems cannot change instantaneously. This is a reasonable assumption for any physical system.

As before, define $\alpha_x = \sigma(s_t, S_x)$, $\beta_x = \sigma(r_t, C_x)$ and $\gamma_x = \Gamma(S_x, C_x)$. We define the total similarity of a response:

$$\sigma_{[a,\,b]}(r) \;=\; \int_a^b \sigma(r,\, R_x)\; dx$$

Similarly, the average is defined $\hat{\sigma}_{[a,\,b]}(r) \;=\; \sigma_{[a,\,b]}(r) \,/\, (b - a)$. Finally, the weighted similarity of a response to the memory trace is:

$$\varsigma(r) \;=\; \int_0^K \gamma_x \sigma(r,\, R_x)\; dx$$

These are just the continuous extensions of the previous definitions.[53]

53. Note that $\varsigma(r)$ is just the inner product $\gamma \cdot \sigma_r$ of γ and the $\sigma_r = \sigma(r, R_x)$.

The definition of the state transition function is the same: $s: c = \boldsymbol{\varepsilon} r[\varsigma(r) = \max\{\varsigma(r) \mid r \in B\}]$. Note however that this defines the state *at the next instant of time.* That is, since $c_{t+dt} = r_t$, the new context is $c_{t+dt} = r_t = s_t : c_t$. This is the differential equation that defines the behavior of the system over time. Recall that we require r_t (and hence c_t) to be continuous.

2. Activation of a Single Trace Interval

We can now derive results analogous to those for the discrete logic. For example, since the interval $[0, K]$ can be broken down into various size intervals $[0, x_1]$, $[x_1, x_2], \ldots, [x_n, K]$ the discrete analysis can be applied to the subintervals. We consider a specific case.

Suppose there are a, b, c and d such that $0 < a < b < c < d < K$. Our intent is that the region $[b, c]$ of the trace is activated, the regions $[0, a]$ and $[d, K]$ are not activated, and the regions $[a, b]$ and $[c, d]$ are partially activated. Hence there are $\varepsilon \approx 0$ and $\zeta \approx 1$ such that $\zeta \le \gamma_x$ for $b \le x \le c$, $\gamma_x < \varepsilon$ for $0 \le x \le a$ or $d \le x \le K$, and $\varepsilon \le \gamma_x \le \zeta$ for $a \le x \le b$ and $c \le x \le d$. We assume the concept is sufficiently sharp; that is, there is a $\delta \approx 0$ such that $b - a < \delta(c - b)$ and $d - c < \delta(c - b)$. We also assume that the activated responses are mutually similar; that is, $\mu[a, b] = \eta \approx 1$. Proceeding as in the discrete case,

$$\begin{aligned} \varsigma(r) &\le \varepsilon\sigma_{[0,a]}(r) + \zeta\sigma_{[a,b]}(r) + \sigma_{[b,c]}(r) + \zeta\sigma_{[c,d]}(r) + \varepsilon\sigma_{[d,K]}(r) \\ &\le \varepsilon a + \zeta(b-a) + \sigma_{[b,c]}(r) + \zeta(d-c) + \varepsilon(K-d) \\ &< \varepsilon a + \zeta\delta(c-b) + \sigma_{[b,c]}(r) + \zeta\delta(c-b) + \varepsilon(K-d) \end{aligned}$$

Conversely,

$$\begin{aligned} \varsigma(R_x) &\ge 0 \cdot \sigma_{[0,a]}(R_x) + \varepsilon\sigma_{[a,b]}(R_x) + \zeta\sigma_{[b,c]}(R_x) + \varepsilon\sigma_{[c,d]}(R_x) + 0 \cdot \sigma_{[d,K]}(R_x) \\ &= \varepsilon\sigma_{[a,b]}(R_x) + \zeta\sigma_{[b,c]}(R_x) + \varepsilon\sigma_{[c,d]}(R_x) \\ &\ge \zeta\ (c-b)\ \mu[b,c] \ = \ \zeta\ (c-b)\ \eta \end{aligned}$$

Noting that $\varsigma(r) \ge \varsigma(R_x)$ allows the inequalities to be combined as before. Simplifying and solving for $\sigma_{[b,c]}(r)$ yields:

$$\sigma_{[b,c]}(r) \ > \ \zeta\ \eta(c-b) - \varepsilon(K + a - d) - 2\zeta\delta(c-b)$$

Hence, the average similarity is:

$$\hat{\sigma}_{[b,\,c]}(r) \;>\; \zeta\,\eta + \varepsilon(K + a - b)/(c - b) - 2\zeta\delta$$

Hence,

$$\lim_{\substack{\delta\to 0\\ \varepsilon\to 0\\ \zeta\to 1}} \hat{\sigma}_{[b,\,c]}(r) \;>\; \eta$$

Hence, as the concept becomes sharp, the interference becomes small and the similarity of the stimulus/context to their stored counterparts increases, the response approaches its stored counterparts.

3. Learning

So far we have described the memory trace as a function (S_x, C_x, R_x) defined for $0 \le x \le K$. How does the memory trace get extended? The simplest way is if we let all experiences be recorded with equal weight. That is, we set $K = t$, and let the memory trace be precisely the previous history of the system, (s_x, c_x, r_x), $0 \le x \le t$. The required modifications to the formulas are simple. The result is that the memory trace wanders through mental space under the influence of its own past history.

More realistically, we might assume that experiences are not equally likely to be recalled. Thus we can postulate a continuous function $\pi_x = \pi(R_x)$ that reflects the *inherent relevance* (such as pleasure or pain) of the mental experience. This is an indirect basis for other measures of relevance. The required modification is simple: $\varsigma(r) \;=\; \int_0^t \pi_x \gamma_x \sigma(r, r_x)\, dx$.

IV. CONCLUSIONS

I have had several goals in this paper. The first was to claim that AI is moving, and must of necessity move, into a new phase that comes to grips with nonverbal reasoning. My second goal has been to claim that traditional AI technology, based as it is on idealized verbal reasoning, is inadequate to this task, and therefore that the new AI requires a new logic, a logic that idealizes nonverbal reasoning. My final goal has been to show, by example, what such a logic might be like. This logic is at present in a very rudimentary form. My only consolation in this is that Boole's logic was a similarly rudimentary form of modern symbolic logic. I would certainly be very gratified if my logic were as near to the mark as his.

V. REFERENCES

Armstrong, S. L., Gleitman, L. R., and Gleitman, H. (1983). "What Some Concepts Might Not Be," *Cognition 13,* pp. 263-308.

Arnheim, Rudolf (1971). *Visual Thinking,* Berkeley and Los Angeles: University of California Press, 1971.

Buchler, J., ed. (1955). *Philosophical Writings of Peirce,* New York: Dover, 1955.

Dreyfus, Hubert L. (1979). *What Computers Can't Do: The Limits of Artificial Intelligence,* revised edition, New York: Harper & Row, 1979.

Dreyfus, H., ed. (1982). *Husserl, Intentionality and Cognitive Science,* Cambridge: The MIT Press, 1982.

Dreyfus, H., and Dreyfus, S. (1986). *Mind over Machine,* New York: Macmillan, The Free Press, 1986.

Gardner, Howard (1985). *The Mind's New Science: A History of the Cognitive Revolution,* New York: Basic Books, 1985.

Goguen, J. A. (1969). "The Logic of Inexact Concepts," *Synthese 19,* 1969, pp. 325-373.

Goudge, Thomas A. (1969). *The Thought of C. S. Peirce,* New York: Dover, 1969.

Haugeland, John, ed. (1981). *Mind Design: Philosophy, Psychology, Artificial Intelligence,* Cambridge: The MIT Press, 1981.

Haugeland, John (1985). *Artificial Intelligence: The Very Idea,* Cambridge: The MIT Press, 1985.

Heidegger, Martin (1962). *Being and Time,* seventh edition, transl. J. Macquarrie and E. Robinson, New York: Harper and Row, 1962.

Heidegger, Martin (1982). *The Basic Problems of Phenomenology,* transl. Albert Hofstadter, Bloomington: Indiana University Press, 1982.

Husserl, Edmund (1962). *Ideas: A General Introduction to Pure Phenomenology,* transl. W. R. Boyce Gibson, London: Collier, 1962.

Johnson-Laird, P. N., and Wason, P. C. (1970). "A Theoretical Analysis of Insight into a Reasoning Task," *Cognitive Psychology 1,* pp. 134-148.

Kahneman, D., Slovic, P., and Tversky, A., eds. (1982). *Judgement under Uncertainty: Heuristics and Biases,* New York: Cambridge University Press, 1982.

Kahneman, D., and Tversky, A. (1982) "The Psychology of Preferences," *Scientific American 246,* 1982, pp. 160-174.

Kahneman, D., and Tversky, A. (1984). "Choices, Values and Frames," *American Psychologist 39,* 1984, pp. 341-350.

Kanerva, Pentti (1986). *"Parallel Structures in Human and Computer Memory,"* RIACS TR 86.2, Research Institute for Advanced Computer Science, NASA Ames Research Center, 1986.

Kickert, Walter J. M. (1978). *Fuzzy Theories on Decision-Making,* Leiden: Martinus Nijhoff, 1978.

Köhler, Wolfgang (1947). *Gestalt Psychology,* New York: New American Library, 1947.

Kohonen, Teuvo (1977). *Associative Memory: A System-Theoretical Approach,* Berlin: Springer-Verlag, 1977.

Kolb, Brian, and Whishaw, Ian Q. (1985). *Fundamentals of Human Neuropsychology,* second edition, New York: W. H. Freeman and Company, 1985.

Kosslyn, Stephen Michael (1980). *Image and Mind,* Cambridge: Harvard University Press, 1980.

Kuhn, Thomas (1970). *The Structure of Scientific Revolutions,* second edition, Chicago: University of Chicago Press, 1970.

Miller, Arthur I. (1986). *Imagery in Scientific Thought,* Cambridge: MIT Press, 1986.

Oettinger, Anthony (1972). "The Semantic Wall," in *Human Communication: A Unified View,* E. David and P. Denes, eds., New York: McGraw-Hill, 1972, p. 5.

Paivio, Allan (1979). *Imagery and Verbal Processes,* Hillsdale: Lawrence Erlbaum Assoc., 1979.

Polya, G. (1968). *Patterns of Plausible Inference,* second edition, Princeton: Princeton University Press, 1968.

Ryle, Gilbert (1949). *The Concept of Mind,* Chicago: University of Chicago Press, 1949.

Rosch, Eleanor (1977). "Human Categorization," in N. Warren, ed., *Advances in Cross-cultural Psychology,* vol. I, London: Academic Press, 1977.

Rosch, Eleanor (1978). "Principles of Categorization," in E. Rosch and B. B. Lloyd, eds., *Cognition and Categorization,* Hillsdale: Lawrence Erlbaum Assoc., 1978.

Shepard, Roger N. (1975). "Form, Formation, and Transformation of Internal Representations," in R. L. Solso ed., *Information Processing in Cognition: The Loyala Symposium,* Hillsdale: Lawrence Erlbaum Assoc., 1975.

Shepard, R. N., and Metzler, J. (1971). "Mental Rotation of Three-dimensional Objects," *Science 171,* pp. 701-703.

Tiercelin, C. (1984). "Peirce on machines, self control and intentionality," in S. B. Torrance, ed., *The Mind and the Machine: Philosophical Aspects of Artificial Intelligence,* Chichester: Ellis Horwood Ltd. and New York: John Wiley, 1984, pp. 99-113.

Tversky,A., and Kahneman, D. (1983). "Extensional vs. Intuitive Reasoning: The Conjunction Fallacy in Probability Judgement," *Psychological Review 90,* 1983, pp. 293-315.

Wason, P. C. (1966). “Reasoning,” in B. Foss, ed., *New Horizons in Psychology,* vol. 1, Harmondsworth: Penguin, 1966.

Wason, P. C., and Johnson-Laird, P. N. (1972). *The Psychology of Reasoning: Structure in Context,* Cambridge: Harvard University Press, 1972.

Wertheimer, Max (1959). *Productive Thinking,* Enlarged Edition, New York: Harper & Brothers, 1959.

Wittgenstein, L. (1958). *Philosophical Investigations,* transl. G. E. M. Anscombe, third edition, New York: Macmillan Company, 1958.

Zadeh, L. A. (1965). “Fuzzy Sets,” *Information and Control 8,* pp. 338-353.

Zadeh, L. A. (1975). “Fuzzy Logic and Approximate Reasoning,” *Synthese 30,* pp. 407-428.

Zadeh, L. A. (1983). “Commonsense Knowledge Representation Based on Fuzzy Logic,” *IEEE Computer 16,* 10 (October 1983), pp. 61-65.