Cognition and Learning: The Implications of a Situated Connectionist Perspective for Theory and Practice in Education.

John A. St. julien
Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/5829

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Cognition and learning: The implications of a situated connectionist perspective for theory and practice in education

St. Julien, John A., Ph.D.
The Louisiana State University and Agricultural and Mechanical Col., 1994

Copyright ©1994 by St. Julien, John A. All rights reserved.
COGNITION AND LEARNING: 
THE IMPLICATIONS OF A SITUATED CONNECTIONIST 
PERSPECTIVE FOR THEORY AND PRACTICE 
IN EDUCATION

A Dissertation
Submitted to the Graduate Faculty of the 
Louisiana State University and 
Agricultural and Mechanical College 
in partial fulfillment of the 
requirements for the degree of 
Doctor of Philosophy 
in 
The Department of Curriculum and Instruction

by
John A. St. Julien
B.S., Louisiana State University, 1973
M.S., Louisiana State University, 1988
August 1994
A man picks up a problem and calls it his, with perhaps slight appreciation that he is taking up a task which arises out of the conflict of insistent social processes, for the solution of which he has volunteered. He makes it his own, but he did not originate it. The academic attitude of creating problems for Doctors' theses is not favorable to the just realization of what problems are when they are genuine.

—Mead, The Philosophy of the Act, 1938, p. 99
ACKNOWLEDGMENTS

The social theme of this dissertation imposes both a special burden and a special joy in thanking those who have helped make the work possible. I believe—and I have tried to give others good reasons to think—that I cannot act alone in any real sense in the formulation of this work.

I act as a member of my community and not simply for it. The community of discourse at LSU has made this dissertation possible and the exciting and collegial atmosphere fostered there contributes profoundly to the work I do. The faculty, the staff, and the graduate students are without peer. I am and remain deeply grateful.

Without the rigor-inducing assistance of Layne St. Julien this book would be a substantially lesser product. Her readings and re-readings make for a much more accessible text.

In a more traditional vein, I would like to take this chance to honor my family, Betsy St. Julien, Layne St. Julien, Brannon May, and Erin May for the patience they have shown for the harried and the distracted.
TABLE OF CONTENTS

ACKNOWLEDGMENTS ................................................................. iv

ABSTRACT ................................................................................. viii

CHAPTER 1
Introduction: Learning Problems .................................................. 1
  Learning and Descriptive Theorization ........................................... 3
  Psychologism and Learning Theory ............................................. 6
  Education and the Need to Go Beyond the Theorization Given ....... 9
  Connectionist Learning, Situated Knowledge ............................... 13
  Strategic Choices ........................................................................ 15
  The Plan of this Dissertation ...................................................... 19

CHAPTER 2
Situating Connectionism: The Horizon Against Which It Appears ... 25
  The Current Context ............................................................... 26
  A Discursive History ............................................................... 31
  Reiss's Narration of the Rise of the Analytico-Referential Discourse ... 34
  Toulmin's Narration of the Development of Modernity ................. 39
  Mind, Thought, and Logic: The Development of an Identity ......... 47
  The Rise of Cognitive Science ................................................... 51
  Emergent Problems in Cognitive Science ................................... 54
  Emergent Answers to Problems in Cognition .............................. 60
  Conclusions ............................................................................. 66

CHAPTER 3
Connectionism: Computation and the Brain .................................. 68
  Connectionism Bounded .......................................................... 68
  Connectionist Analogies .......................................................... 70
  Connectionism Described .......................................................... 73
    An Example of a Connectionist Memory Model ......................... 74
    An Examination of a Connectionist Simulation .......................... 78
  The Plausibility of Connectionist Models ..................................... 87
  The Cognitive Plausibility of Connectionist Models ..................... 89
    Education and the Classical Cognitivist Account ......................... 89
    The Basis for Emergent Cognitive Properties in Network
      Architectures ...................................................................... 95
    Complexity and the Plausibility of Connectionist
      Architectures ....................................................................... 97
    Error and the Plausibility of Connectionist Architectures ............ 99
    History and the Plausibility of Connectionist Architectures ......... 101
  Learning: Ill-Formed Problems and Connectionism ...................... 106
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Connectionism and Theories of Practice</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>The Representation Debate</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td>Representation in Connectionist Networks</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>Representation Outside Connectionist Networks</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td>Abstraction, Embodiment and Displaced Knowledge</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>Vygotsky on the Word</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>Dreyfus and Rules</td>
<td>184</td>
</tr>
<tr>
<td></td>
<td>Theories of Practice</td>
<td>188</td>
</tr>
<tr>
<td></td>
<td>Situating Situated Cognition</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>195</td>
</tr>
<tr>
<td>5</td>
<td>Eukaryotes or Prokaryotes?</td>
<td>197</td>
</tr>
<tr>
<td></td>
<td>The Background of the Study</td>
<td>198</td>
</tr>
<tr>
<td></td>
<td>Connectionist Constraints, Practice-Based Design</td>
<td>201</td>
</tr>
<tr>
<td></td>
<td>The Study</td>
<td>208</td>
</tr>
<tr>
<td></td>
<td>The Experimental Materials</td>
<td>208</td>
</tr>
<tr>
<td></td>
<td>The Conditions of the Study</td>
<td>209</td>
</tr>
<tr>
<td></td>
<td>Statistical Measures</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>Discussion of Statistical Results</td>
<td>214</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>219</td>
</tr>
<tr>
<td>6</td>
<td>Designing a Programmed Environment to Induce Category Formation</td>
<td>222</td>
</tr>
<tr>
<td></td>
<td>The Challenge of Software Design</td>
<td>223</td>
</tr>
<tr>
<td></td>
<td>A Theoretical Background for Category Learning</td>
<td>225</td>
</tr>
<tr>
<td></td>
<td>Practices and Genres</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td>Activity and Temporality</td>
<td>231</td>
</tr>
<tr>
<td></td>
<td>The Category Learning Program</td>
<td>233</td>
</tr>
<tr>
<td></td>
<td>The Program’s Design</td>
<td>234</td>
</tr>
<tr>
<td></td>
<td>Connectionism and This Program’s Design</td>
<td>234</td>
</tr>
<tr>
<td></td>
<td>Practices and this Program’s Design</td>
<td>236</td>
</tr>
<tr>
<td></td>
<td>Interacting With the Category Formation Program</td>
<td>239</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>243</td>
</tr>
</tbody>
</table>
CHAPTER 7
Situating Connectionism: Summary & Implications ........................................... 244
  Reviewing the Path ......................................................................................... 246
  Absences ........................................................................................................... 254
    The Individual ............................................................................................. 255
    Institutions ................................................................................................. 258
    The Body ...................................................................................................... 259
  Implications ...................................................................................................... 260
    Recovery and Reconstitution ...................................................................... 261
    Intentionality ............................................................................................. 262
    Development ............................................................................................... 264
  A Pragmatic Contextualization ...................................................................... 266
  The Child ......................................................................................................... 270
  Conclusion ....................................................................................................... 276

WORKS CITED ................................................................................................ 279

APPENDIX ....................................................................................................... 290
Category Learning Program ............................................................................ 290
  Instructions and System Requirements ......................................................... 290

VITA ................................................................................................................. 291
ABSTRACT

In this dissertation I attempt to construct a framework for an alternate theory of instruction, starting from the position that education has no theory of learning; instead, what passes for theories of learning are actually descriptions of the conditions under which knowledge is acquired. Descriptive theorizing does not serve education well because it is not likely, being a description of what is known told in terms of assumed categories, to be generative or adaptive.

I question the naturalness of current assumptions about thought and learning by tracing the consolidation of the present discursive formation around a presumed unity of logic, language, and causality based on the forms of geometry. A crucial move in that consolidation was Descartes' formulation of thought as essentially logical. As our culture deals with the contradictions inherent in this formulation, new disciplines of knowledge arise. Most interestingly for education, the new disciplines of cognitive science and artificial intelligence have pushed the idea that humans are "computers," that we reason by calculation, to the breaking point. It has become increasingly obvious that humans cannot think according to the forms of logic and connectionist theories which propose alternate images of mind are gaining ascendance.

The implications for education are large. I extensively explore the implications of connectionist modeling for a distinctively educational model of learning. These connectionist theories substitute a shifting and uncertain web of associations for the solid storage metaphor common to most educational theorizing and methods. The stability which can no longer be located in the
sovereign self must instead be found in the world and in the socially-based practices that constitute both the world and the individual. Situated cognition, pragmatists, and poststructural sociologists are explored to understand the new constellation surrounding learning. A short exploratory study, based on the principles that emerged from the study, of an alternate way to teach categories is experimentally explored and found successful. This work was extended to a computer-based implementation which allowed theoretical ideas concerning time and activity to be explored. Includes a Macintosh disk.
CHAPTER 1

Introduction: Learning Problems

The cognitive revolution simply absorbed the concept of learning into the broader concept of "the acquisition of knowledge."

—Jerome Bruner,

The first task of a study such as this is to establish its problem as a problem for the reader and to leave that reader in a position to understand why the author is concerned. The explanation given is always translatable as a story, a story about a fascination, a problematic situation, an attempt to understand the situation and to draw a useful meaning out of the problem and our reaction to it. I offer the following as a metaphor for the story that I would like to tell.

I want to tell a story, a story about learning and a story about love; a story which is finally about meaning.

This is always a particular story. It is a simple story, perhaps the simplest of all stories. So simple and so true that we have never questioned the full implications of it....

I sit in a room and watch a young child in a crib beneath an open window. She is beautiful, or I think she is. I sit immobilized, fascinated. A cool gust of wind moves the curtains, annoying the child. Her mother, harried, returns and bends over the crib. The baby perceives motion, turns her head and recognizes her mother. She gurgles, smiles, reaches out and is softly gathered in....

There are any number of miracles in this story but I want to focus on a particular one: the child recognizes her mother's face.
It is miraculous, but in more than our common sense of being wonderful. It is miraculous in the older sense of being inexplicable, in the older sense of being quite literally impossible to explain logically.

But we do explain it logically, or at least we impute the forms of formal logic: if X, then Y. If the baby is cold and if the baby knows that Mother has been warm, then she holds out her arms to be picked up. I choose this story, in part, because we have an intuitive sense that in this case our explanation is inadequate. Something within tells us that there is more to this story.

Logic is how, almost universally, our culture frames our understanding of thought. But we are uneasy about making the claim that the baby actually operates in this formal, propositional way. We are not uneasy in making the claim that a six-year-old child learning his letters operates in this way. We do make the claim that our bosses operate this way. We even make the claim that we operate this way.

I think that we are wrong. I think that thought cannot be propositional. I think that it cannot take the form of formal logic. And I believe that this is not an inconsequential mistake. At stake here is the very making of meaning. Our society—and especially our schools and our academic practices—are predicated on the idea that the process of thought takes the form of logic and that what is not "logical" can only be a source of "error." Real meaning, canonical meaning, is logical; it is known by what it is not—not contextual, not particular, and certainly not uncertain. This leads us to ignore "lesser" forms of learning that are, obviously and intrinsically, contextual, particular, and ambiguous. By ignoring perceptually-based learning such as the child exhibits in favor of the acquisition of decontextualized facts, schools implicitly deny the sort of connected, meaningful learning that comes easily to every child in favor of a regime of knowledge that is disconnected and difficult. If it is a mistake to
believe that the kind of learning that the child does when she comes to recognize her mother’s face is inconsequential, and that such learning has a very direct relationship to the sorts of meaningful learning that leads to competence on the part of students, then rectifying that mistake may have far-reaching consequences for the practice of education.

Learning and Descriptive Theorization

The child’s coming to recognize the mother is arguably the first instance of learning for each child. That we have not been able to credibly explain how this is possible stands as a profound indictment of our understanding of learning.

We are generally content to simply describe the conditions under which such knowledge is acquired. Or, more accurately, we describe the conditions under which knowledge is acquired and then the conditions under which acquisition fails and attribute the failure to the difference between the two descriptions. The weakness of this approach is most vividly revealed in situations where learning typically succeeds. The success cannot be explained because there are no differences to which we may attribute it. Because we can find almost no conditions under which the child does not learn to recognize faces—or form categories or learn language—we conclude that processes which underlie these powers have little to teach us. But this conclusion depends on the crucial assumption that learning can be explicated by referring to the conditions under which knowledge (understood as learning’s residue) is acquired. The sleight-of-hand, behaviorist substitution of the observable conditions of knowledge acquisition for the process of learning is what allows us to dismiss,

---

as "unproblematic" and therefore uninteresting in the context of schooling, facial recognition, category formation, and language acquisition. We are led to abandon the study of learning, and especially learning that almost always succeeds, in order to study the conditions under which knowledge acquisition fails.

In this crucial regard the much-touted cognitive revolution fails to break with its behaviorist antecedents. It fails to move from knowledge and its acquisition to learning. It simply admits, in a structuralist response, that structures can exist within the head that correspond to the structures of the world. This simple admission does not differentiate cognitivists from behaviorists. Many behaviorists would not dispute the existence of mental structures; rather, cognitivism differs from behaviorism in its attitude toward the profitability of pursuing the question of mental structures. As such, simple cognitivism does not break with traditional approaches, and educators are left with the same old wine repackaged in new bottles.

By not attempting to understand context-dependent, successful learning and its implications, education effectively abandons the possibility of constructing an understanding of learning that is not fundamentally descriptive. Descriptive theorizing, while often necessary when the material processes which constitute phenomena are opaque, is fundamentally conservative—it describes already established practices and cannot generate new ones. Descriptive theories start with the categories that tradition offers. It takes these categories as "natural"—as self-evident. If these categories are taken as natural then any difficult, conflicting data resulting from further inquiry into the matter at hand is likely

---

2 This point is often underemphasized and a casual reader of cognitive literature could come away with the impression that the difference between the two schools is based on the existence of mental structures. See Howard Gardner, *The Mind's New Science* (Basic Books: New York, 1985).
to result in finer and finer subdivisions of the categories and more and more conditional rules being generated to account for the anomalous results of inquiry. Without the limits which a material process imposes on the possible relations that constitute objects the radical underdetermination of descriptive theory encourages the theoretician to postulate an ever more complex and numerous entities to elaborate the traditional scheme. It was just this process that resulted in the elaboration of the Ptolemaic system of astronomy with its circles within circles. The functional, modularist school of cognitive psychology, starting with the objects offered it by the traditional folk psychologies appears well advanced down this road. “Memory,” for instance is broken into long and short term memories with various relational rules governing their interaction and fundamentally different processes of thinking are argued for language and sense-based forms of thought.

While others have rebelled against descriptive theorizing, few have done so with such fine invective as John Dewey:

...such logic only abstracts some aspect of the existing course of events in order to reduplicate it as a petrified eternal principle by which to explain the very changes of which it is a formalization.³

Dewey was objecting to substituting a reifying, “objective” description, which he saw as grounded in an unhealthy reliance on the transcendent forms of logic for the preferable grounds of perception and inference in an attempt to ground our understanding in the interactions “of changing things.”⁴

Because descriptive theories can have access only to the “products” of learning, they are necessarily reifying theories of knowledge rather than


⁴ Ibid., p 7.
theories of learning. They can only describe what is known and the conditions under which this knowledge came to be known, abstract a simplified form, and impose this pattern on future instances. Critically, descriptive theories have no principled method to discover new ways to organize the data that they take as basic. By and large they must start with the categories that the history of their society offers them—and this systematically limits their explanation to the traditional categories and divisions.

The reliance on descriptive theorization has resulted in the conflation of theories of knowledge acquisition with theories of learning. Theories of knowledge acquisition are concerned with the conditions under which knowledge is acquired. Theories of learning, by contrast, should be concerned with the processes by which that change that we call learning occurs. In the final analysis, a better understanding of both learning and knowledge will be necessary to develop a fuller theory of instruction—but we most acutely lack a credible theory of learning.

It is a central contention of this study that coming to an understanding of learning which is not simply descriptive, but which instead is an account of material processes, is basic to reconceptualizing learning in a way that can lead to the improvement of educational practice.

**Psychologism and Learning Theory**

Descriptive theorizing makes a priori, culturally normative assumptions as to the appropriate level of description. Traditionally, learning theories have focused on the individual level. As such, learning fit—however uncomfortably—into the discipline of psychology. Academic psychology thus

---

became the theoretical well from which the practice in the applied field of education was drawn. In adopting this psychologistic approach, education has largely ignored the crucial social elements of learning and the fully social basis of knowledge.

Critique of psychologistic theory and practice has been a bracing part of the recent discourse of educational theory.6 Theoretical inspiration is now being drawn from such diverse fields as literary theory, sociology, anthropology, cognitive science, discourse theory, narrative theory, aesthetics, practice theory, post-structuralism, post-modernism, and linguistics.7 Either explicitly or implicitly (by emphasizing particular social practices as narrative theory does), such critiques move the focus of educational theory away from the sovereign individual. Nonetheless, the thrust of such critiques has been to oppose particular psychological theories and practices justified by these particular theories rather than to oppose psychologism itself.8 Generally speaking, interpretations drawn from these fields do not challenge the centrality of the individual in questions which concern the actual tasks of teaching and learning.

This appropriation of ambiguous or even patently anti-individualistic theories to the service of individualistic practice is neither new nor, in an historical context, surprising.9 But it is under-appreciated in the current

---


8 Exceptions distinguish themselves by their rarity. For two good examples, see: Jean Lave and Etienne Wenger, Situated Learning; Legitimate Peripheral Participation (Cambridge, MA: Cambridge University Press, 1991) and Valerie Walkerdine, The Mastery of Reason: Cognitive Development and the Production of Rationality (London: Routledge, 1988).

9 This misuse has been a particular problem with the pragmatists. Readers may wish to persuade themselves that the pragmatists were not individualists in the sense used here. See,
circumstance. Potentially progressive theories are bravely used to shore up the status quo while exuding the pretense of radical change. Taking one extreme as an example, E. D. Hirsh has combined elements suggested by cognitive science and discourse theory to suggest that discursive competence can be achieved via the didactic teaching and simple memorization of a list of culturally significant elemental facts. These unstructured facts are taken as examples of schemata which subserve discursive competence. Most proponents of either theoretical school would reject his construction as representing their theoretical convictions. Similarly, insights from sociology, anthropology, and organization theory are used to inform and buttress “cooperative learning strategies,” but such strategies are finally understood as more effective ways to achieve the unaltered endpoint: individual learning.

Quite simply, nothing has interrupted the preunderstanding of schooling in which learning is distinctively individualistic and hence psychologistic. As long as we regard learning as essentially a task accomplished by individuals to the end of increasing their store of useful knowledge, it is hard to see how the fundamentally psychologistic orientation of education can be altered.

The need for a broader definition of the problem of education has long been acknowledged. The educational work of Dewey, and the more thoroughly elaborated pragmatic descriptions of the constitution of the self and the importance of the act put forth by George Herbert Mead, give a firm historical

---


grounding to more social ways of understanding the context of learning.\textsuperscript{12} Current theories based on the work of such as Pierre Bordieu, Jean Lave and Valerie Walkerdine offer ways to extend theories of experience and action by a fine-grained analysis of the practice, or habitus, of which they are composed.

But, like older theories, modern work does not propose a learning process. Like theories that they are set in opposition to, they describe the conditions under which learning occurs. They differ by describing the distinctive social context in which learning occurs and by problematizing the relation of the individual to that social context. Still, such theories are finally theories of knowledge acquisition, not learning, and insofar as educational adopters remain free to assume that learning remains at its root the memorization of facts, the full value of this advance will not be available to students.

**Education and the Need to Go Beyond the Theorization Given**

Because descriptive theorizing is not generative—being a description of what is already perceived projected onto unknown situations—it often proves sterile in practice. Education, caught in a basically descriptive account of learning, has suffered from this effect. One of the most troubling aspects of education's long crisis is the oft-noted institutional tendency to persist in the practice of discredited methods and strategies of teaching. It is a contention of this work that—at least in part—current practices continue to exist, even though repeatedly and convincingly critiqued, because no plausible alternative is understood to exist for the current understanding of learning and knowledge that supports them.

\textsuperscript{12} Two fine examples are: Dewey and Bently, *Knowing and the Known* and Mead, *The Philosophy of the Act*. 
An example may serve: memorization is "bad." This is a central credo of almost any program of teacher education and is routinely repeated by practicing teachers. Such a critique is not empty of content, research-based or lived. Research demonstrates and teachers' experience confirms that material understood as lists for memorization is poorly remembered and (even when retrievable) seldom brought to bear on appropriate problems. Nonetheless, from teaching the alphabet to making lists of characteristics which distinguish eukaryotic and prokaryotic cells, memorization is the first and often the last task set for the student. The implicit contradiction between omnipresent practice and substantiated judgments concerning the value of such practice is a painful condition of educational life.

But given the classic assumptions about knowledge, the commitment of facts to memory is necessary. Learning simply is the commitment of facts to memory. At root, almost no one quarrels with that. The most progressive will insist that the facts that are learned must be useful in a real-world context—and will attempt to teach methods of problem solving to students. Ironically, what is most often offered is a list of strategies to be memorized in addition to the list of facts. What follows from the conjunction of the demonstrated ineffectiveness of memorization and the understanding that memorization is what learning consists of is the conclusion that it is how we teach students to memorize that is at fault.

Hence arises the plethora of "strategies" that constitute the staple of teacher instruction. The solution is to "sweeten" memorization, to make it fun, and to make sure peripheral factors like "low self-esteem" do not interfere with the process. Indicative of the poverty of this approach is the cyclical nature of educational reform in the area. There are only so many ways to "sweeten" memorization and as each strategy tries and fails to secure a basic improvement
in student competence, veteran teachers notice, with understandable cynicism, the recycling of previously failed strategies.

When we look at educational problems in this light, we begin to suspect that either experience—both research-based and lived—or our conception of learning is wrong. This dissertation is largely dedicated to the task of affirming experience and to demonstrating that a viable alternative to the traditional conception of learning is emerging and may be able to help us exit the (un)merry-go-round of educational reform.

In short, our psychologistic, descriptive orientation has focused our attention narrowly on the conscious individual to the exclusion of perceptual and unconscious processes which lie "before" the individual and to the exclusion of the socially organized regularities in which everyday cognition takes place. The sterile recycling of educational solutions press us toward the conclusion that learning needs to be reconceptualized in order to serve our students adequately.

But if psychological perspectives do not and cannot create the intellectual apparatus to interrupt the present organization of schooling practices, then in what direction can the teacher turn? This question arises in education with particular force for, unlike many academic fields—most saliently, anthropology and cognitive science—education is intimately tied to the world of practical application. Simple disproof of an organizing theory is not enough to change practice; both a basic disproof and an alternative, positive conception are needed to effect fundamental change. Previous attempts to displace the individualistic or logocentric positions which have characterized educational practice have foundered by not producing a convincing case for abandoning the theoretical assumptions which underlie such traditional practice. This has made it possible to adopt the rhetoric and even the surface form of the practices
advocated by educators such as Dewey without making a fundamental commitment to a different conceptual universe. Dewey's work, and other forms of active learning, are (mis)understood as mere strategies to sugar coat in yet another way the traditional assumption that learning consists of listing and memorizing facts.

What is needed is an understanding of the material process of learning that competes directly with the logocentric assumptions of the Western tradition. Only a perspective that directly challenges the idea that thought is logic can provide a firm basis for altering educational practice. It will be argued below that education is prey to a specie of the “intentionality problem.” It is not apparent to our students what most of the curriculum is “about.” The decontextualization of school knowledge is a recurrent theme in curriculum theory. But it is seldom recognized how deeply the decontextualized knowledge that we object to in schooling is due to our basic understanding of thought and learning themselves. It is no accident that schools teach as if knowledge were constructed of atomistic, decontextualized fragments. The dominant tradition of our culture holds that just this is true. Schools simply enact what society proposes. What is needed is a way to tie the processes which are the basis for the child's recognition of her mother to the competencies that are sought by schooling. A child's recognition of her mother is manifestly about her mother. And her reaching out to be held is an inarguably competent response.

A student should leave school similarly competent to act in the world; it is becoming progressively clearer that this is not the same as being able to manipulate formal symbolic systems. Instead, current work empirically
demonstrates that competence in the world is embedded in the network of practices that social groups share and which organize their interaction.\textsuperscript{13}

\textbf{Connectionist Learning, Situated Knowledge}

This dissertation will attempt to make sensible the potential importance of connectionist approaches to learning when understood in conjunction with situated perspectives on knowledge. As developed here, this program attempts to evade the bifurcation of self and the world which has produced the problem of intentionality. The connection between the symbol and the object it represents is taken as entirely arbitrary and hence essentially meaningless in most popular theories of meaning. In opposition, the approach advocated here finds meaningful connections between symbols and objects in the world in the social, lived practices experienced by the actor.

Initially, connectionist learning and situated knowledge may seem like unlikely partners. They emerge from very different venues and appear to pursue different paths toward the question of human meaning-making. Connectionism emerges out of computational theory, systems theory, and neurology. Situated perspectives on knowledge emphasize sociological and phenomenological factors and are particularly indebted to the socially oriented theories exemplified by Lev Vygotsky.

In part they meet in their absences. Situated cognition and connectionism both avoid a focus on the traditional hero of the story of learning: the individual. They also provide complementary stories which fill out each other's shortcomings. Connectionism tells an intriguing and suggestive story of just how context-dependent learning, associative memory, and other educationally central mysteries can arise in the material world. But its explanation of these factors leaves new questions of how stability of representation is maintained. Situated cognition, with its emphasis on embodied practices, fills in this gap with socially organized regularities found in the world which substitute for stable symbols found in the head. Similarly, situated cognition, while providing a true advance in explaining the context dependence of knowledge and tracing out the ways in which this is intersubjectively organized, leaves us with no sense of how individuals can act across situations and is crippled in explaining how the powerful abstractions which are the chief object of school learning can arise. Connectionism can fill those gaps without forcing situated perspectives back onto the questionable path of individualism.

More positively, connectionism and situated perspectives meet in their common analytical stance. Both focus on network architecture and the material processes that constitute, respectively, learning and knowledge. Network architectures suggest that cause is not so much determinative as it is constrained. That is, Newtonian, billiard-ball conceptions of causality are traded for a dense network of relational connections which, in concert, limit the paths that their system can take over time. This analytical model, often associated with Darwin's introduction of evolution into scientific discourse, remains an underutilized intellectual tool which has proven powerful in the

---

hands of researchers in both of these fields. Both areas of research also
determinedly keep their eye on the ball of material process. They avoid purely
general, descriptive theorizing in favor of the analysis of the small-grained,
material, interlocking networks which constitute the phenomena they study.

It is worth noting that because of the style of theorizing they share, they are
not in a position to offer prescriptive advice about any and all situations in
which their analytical tools could be used. In an educational context this means
that while they both point toward constraints on educational practice, they do
not suggest “the best” strategy. Rather, they reveal what is unlikely to result in
the learning of useful knowledge. Neither is in a position to suggest a single,
best method for education. The choices, finally, are our own.

Strategic Choices

This project is committed to working toward a basis for a theory of
instruction in schools; it examines connectionist and situated perspectives with
an eye toward that goal and not a general, abstract theory of cognition. On the
basis of a synthesis of the relevant research, it suggests that a more useful
educational approach to learning would join the “about-ness” of connectionist
perception to the competent knowledge embedded in practices. This is not a
prescriptive project which seeks to define the best teaching method, but one
which hopes to constrain theory and practice by eliminating particular,
untenable assumptions about learning. Beyond eliminating approaches, it also
hopes to offer intellectual instrumentalities15 through which to perceive and
think about learning which will prove useful in the interpretation of research

15 The phrase “intellectual instrumentalities “ is drawn from Kliebard, who in turn attributes
it to John Dewey. Herber M. Kliebard, “What is a Knowledge Base, and Who Would Use it If
We Had One?” Review of Educational Research 63 (Fall, 1993): 295-303.
and the design of instruction. In fact, it provides support for the position that the creative act of teaching lies ineradicably in the teacher's hands by emphasizing unique particularity of any act of cognition in a way that casts doubt on the universalistic assumptions that support teacher-proof curricula.

I have argued that our culture in general and the dominant strand of educational discourse in particular ignore the possibilities of focusing on learning in favor of a culturally preferred focus on the conditions under which knowledge is acquired. Metaphorically, we ignore the child who learns to recognize her mother's face in favor of the adolescent who (sometimes) learns geometry. The descriptive, psychologistic theory that has resulted from this predilection has not served students well.

By substituting a theory of knowledge acquisition for a theory of learning, we are largely unable to reconceive our practice in a way which will benefit our students. Such a substitution occludes the fact that we have no theory of learning and this blind spot is a surer bar to our creative thought on these matters than an incorrect theory would be, for it diverts our attention from the heart of the educational enterprise—learning it does so by substituting what is taught and how reliably what is taught is transmitted for how learning takes place and how effective, meaningful learning can be encouraged.

The project, then, is not so much to critique the old framework—this has been convincingly done by others. Nor, finally, is it only to present a new theoretical framework that is more useful. This, too, has been convincingly assayed, not only in the current framework but as far back as John Dewey. Neither incisive critiques nor enticing alternatives have had much effect on the practice of education. I will ask, why not? If socially situated approaches to cognition have been implicit in the most sophisticated approaches to the problems of learning (including socialization and perception as forms of
learning), why haven’t practices been adopted which reflect these insights? Crucially for educators, why has education not been noticeably affected by these insights into what we take to be our basic task: encouraging learning?

I suspect that a major reason is that taking such a viewpoint on learning implies deep differences from the culturally canonical position on individuality and thought. It is difficult to imagine a more difficult and counterintuitive position than one which asks us to see that our very experience of ourselves and our thought is historically specific, culturally interpreted, and potentially problematic. Thorough-going situated perspectives go further and make the claim that not only is our self-conception problematic, it is wrong—or at least profoundly misleading. Previous impulses in the direction of a socially grounded, perceptually aware cognition have failed to overcome this barrier which is strategically situated in our very perception of ourselves. Insights emerging from the various isolated disciplines were simply swamped by the larger framework which supported the canonical position. Perception became an isolated technical field, anthropology was marginalized, and perspectives in psychology and education were ignored or their meaning was transformed in their implementation. The history of Dewey’s attempt to insert a social conception of learning into the discourses of both philosophy and education is uncomfortably instructive in this regard. G. H. Mead’s social psychology became a pale imitation of its basic intuition as it was transmuted into the symbolic interactionist school of sociology. James’ associative psychology and the gestalt psychology of pattern recognition (in both its perceptual and humanist forms) fared no better against the logicist alternatives.

16 As with much of Dewey’s work, it seems more fruitful to go to Dewey when discussing misinterpretations and misappropriations than to his commentators. See: Dewey, “Appendix: Letter to a Friend,” in Knowing and the Known, 313-329; Dewey, “Preface,” in The Influence of Darwin on Philosophy, iii-vi.
We need to ask ourselves, what is to prevent the same things from happening to a perspective based on connectionist pattern recognition and the socially aware alternative of situated cognition? How can we avoid having valuable insights swamped by the “natural,” already established framework of explanation?

Perhaps the greatest advantage of alternate perspectives in the present day is an awareness of this problematic history. An awareness that past attempts have not been as successful as could be wished leads to an attempt to understand the historical conditions under which such attempts have labored and the tactics that they have employed. More specifically, observing the history of past attempts allows us to see that the force of the canonical assumption that thought is logical cannot be countered by simply pointing out that this is a descriptive theoretical frame—an implication of Dewey’s work cited above—that has not worked and positing a presumptively better theoretical frame. The history of pragmatism gives testimony to this. The canonical assumption of thought as logic apparently needs to be challenged at the level of its root, not its fruit.

In addition, alternate perspectives have labored under the weight and against the systematicity of the way that our culture regards the self and thought, the world and the real. The pervasive and mutually reinforcing relationships that compose this system of thought work against change in any single area. The most strikingly different insights (and Dewey is always a

17 The need to embed particular theoretical advances in a larger context which supports them is beginning to enter educational discourse. Both Gee and Clancey in recent work argue that their particular work needs to be embedded in a larger story stretching from biology to sociality to be effective. See: James Paul Gee, *The Social Mind: Language, Ideology and Social Practice* (New York: Bergin and Garvey, 1992) and William J. Clancey and Jeremy Roschelle, “Situated Cognition: How Representations are Created and Given Meaning,” Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, 3-7 April 1991.
good example) are interpreted in ways that make sense in terms of the background understanding and are then implemented in “rational” ways. Any valuable, distinctively different, contribution that they might have made is washed out during this process.

Third, and this is a particularly educational demand, some real, tangible difference that can be seen as an improvement over current instructional strategies needs to flow from the new perspective.

This array of problems implies a triadic strategy. First, the basic understanding of learning itself needs to be the fundamental issue. Second, granting the deep embeddedness and systematic nature of the current conception, a wide spectrum of evidence which provides the foundations for an alternate conception of learning must be arrayed across historical, biological, social, and even computational fields. Third, the framework proposed should imply differences in how the field of educational practice—research, materials, and instructional design—is conducted.

The Plan of this Dissertation

This dissertation tries to fill in some of the holes in an alternate, more useful, story about the nature of learning. It is a daunting project. Broadly, I will argue that at a deeply cultural level the West has taken a turn away from learning which has proven unproductive from the point of view of education. This turn has deprived educators of the tools with which to think productively about crucial aspects of their enterprise. The task that is taken up here is to build new lenses with which to view education and new tools with which to build educational practice.
This first chapter attempts to bring the reader into the problem. It works to show why learning theory, or rather its absence, is a problem and why the absence of a learning theory based in material change has not been seen as a problem.

Chapter 2 pursues the history of the current discursive structure and the relationship between that structure, its contradictions, and the development of cognitive science. It attempts to demonstrate that the received understanding of our selves and our relationship to the world is not natural, but is a response to an earlier crisis of representation. In trying to bridge a gap which the dynamics of the earlier discourse of patterning had opened up between the word and the thing represented, the current analytico-referential discourse posits three parallel systems. Language, logic, and the world are held to have the same form—to operate according to the same rules over similar objects. Because they are identical in this way they can be arranged to operate in parallel. The world is taken as logical, logic is taken as analogous to causal relations in the world, and language is taken as capable of reflecting accurately both of these relations. As this discourse develops, its dynamics also produce evidence that this ordering is not so much a description of things as they are but an unrealistic assumption about how meaning is made.

Viewed against the background of an unraveling discourse, debates within cognitive science between logic and statistics, biology and formalism, and knowledge acquisition and learning take on heightened significance. Understanding the historical context allows the reader to see the emergence of connectionist alternatives to logical formalism as a particularly crucial site—the location of our own self-image—for reformulating our sense of how meaning is made. This context also highlights the reasons that perspectives such as connectionism are, at least initially, counter-intuitive: they are discursively
“unnatural” and threaten the current structures which organize our meaning making.

Chapter 3 dives into an exposition of connectionism as a theory of learning. This chapter attempts to reconstruct some of our common mental conceptions on a connectionist basis. Relying in places on computer simulations redeveloped by the author, a model of context-dependent associative memory and category formation is built which demonstrates a robustness which artificial intelligence constructs, founded as they are on the logocentric assumptions of the analytico-referential discourse, cannot match. This approach leads to understanding learning as a change which takes place in a material, relational network. A web of such networked relations makes possible the crucial phenomena of distributed representation, and increasingly complex layering and recursive interactions between layers are shown to enable more and more sophisticated forms of learning. The chapter examines the psychological, computational, and neurological plausibility of connectionist architectures, concluding that in each realm connectionist perspectives are more plausible than its competitors’. This, however, does not mean that accepting connectionism as a viable learning theory solves the educational problems that followed on the conflation of learning and knowledge acquisition. Disentangling these two concepts leaves us with a theory of learning but without a compatible theory of knowledge. Knowledge, the knowledge valued by the community, is what teachers are charged to teach, and a learning theory which hopes to aid educators in the construction of better a pedagogy must ally itself with a congruent theory of knowledge if a useful theory of instruction is to be developed.

Pursuing the question of a theory of knowledge adequate to the task of instruction, chapter 4 develops a focus on practices as recurring patterns of
interaction which form the basis for a social theory of knowledge. This chapter enters the subject by reviewing the ongoing “representation debate” in the academy and pointing to the distinctive position implied by connectionist architectures—a position which offers a third path to competing formalist and imagist alternatives. A connectionist viewpoint inverts the usual perspectives concerning representation and finds the stability, not the instability, of representation a problem in need of explanation. Because connectionist networks are inherently unstable, and because this instability is associated with the useful characteristics that move researchers to adopt the stance, some other source of representational stability must be sought. This chapter pursues the possibility of locating the necessary and observable stability in the intersubjectively regularized interactions which comprise our relations with the socially-organized world we inhabit.

While focusing on Vygotskian and situated approaches to cognition, chapter 4 remains mindful of the obligation that theorization has to schools and students. It will not be enough to claim that the sorts of learning most easily understandable from a situated connectionist perspective are not the learning that we honor in schools. Rather, it will be necessary to explain the sorts of displaced but nonetheless embodied learnings that are valued by the community. These are represented by three-digit multiplication. In working to explain abstract, displaced knowledge, the emerging alternative relies on the previously developed approach to category formation and suggests that abstract cognitive tools are built in the same way that more concrete categories are built. Thus the practices associated with addition (leftward displacement, vertical addition) are redeployed in multiplication to construct new, more abstract, and arguably more powerful, cognitive tools.
Chapters 2, 3 and 4 work to fill the first two goals in my "triadic" strategy for developing an alternative, more educationally useful approach to learning and cognition. They focus on building a learning theory in a broad historical context which is plausible given the constraints of neurology, situated cognition, and computational theory. Having laid out a skeletal framework for an alternate approach to learning and knowledge, I must demonstrate how this difference can make a difference. Chapters 5 and 6 begin to work in this direction.

Chapter 5 takes up the issues of research design and analysis in the context of the design of instructional materials. Using constructs suggested by the developing situated/learning perspective, a significantly different approach to teaching the distinction between eukaryotic and prokaryotic cells in the discipline of microbiology is constructed. Extracted from a larger experimental framework, the connectionist/practice approach suggests differences from traditional text genres which are statistically analyzed. By and large, the analysis indicates that the new material works as well and often better than the more familiar traditional material in enabling students to make a correct distinction between eukaryotic and prokaryotic cells.

While the results from chapter 5 are encouraging, the experimental study examined there was constrained by the experimental context and the nature of paper-based materials to slight two crucial and particularly distinctive aspects of the emerging perspective on learning: the importance of situated activity and particularly the hermeneutic, temporally embedded aspect of that activity. Chapter 6 takes on the task of making the implications of these aspects clearer in the context of developing a computer-based learning module which takes advantage of the underutilized plasticity of the medium to make a more dynamic presentation of the eukaryotic/prokaryotic distinction possible. This
program serves as a model for teaching the bifurcating, absolute distinctions which are typical of taxonomies and classical categorization in general. Sample screens are discussed, as is the progressive construction of categories through the manipulative activity of the user. The program itself is included as an appendix and the reader is encouraged to work the sequence in order to better understand the dynamically oriented position taken in the analysis.

Chapter 7 summarizes the work, characterizes the position taken on learning, and suggests that this position fits into a more general intellectual style of analysis emerging across a number of intellectual fields. It points to gaps remaining to be filled in the current position on learning taken by this dissertation and lays out associated areas for further exploration using this style of analysis. Development and the construction of the self in activity are two areas in which the author hopes to integrate the positions taken herein with progressive stances taken in those literatures.

The perspective offered in this work is largely a matter of reinterpreting our understanding of the world and ourselves in terms of realistic constraints on our theory and practice—our praxis. The attempt is not so much to provide a basis for radically changing what we do in schools (though in places that seems appropriate) as it is to give us new tools through which to view the ongoing successes and failures of our engagement with students, and thereby revitalize our work. If it is successful, it will allow us to view the child who learns to recognize her mother's face successfully and act competently as a part of that recognition as continuous with the students in our classroom—and as deserving of our wonder and appreciation as any newborn. It attempts, in contrast to the usual impulse of reformist movements, to pour new wine into old bottles.
CHAPTER 2
Situating Connectionism:
The Horizon Against Which It Appears

When a man reasons, he does nothing else but conceive a sum total from addition of parcels, for REASON . . . is nothing but reckoning . . . .

—Hobbes,
The Leviathan, 1958, p. 45

This chapter concentrates on providing a background against which connectionist theorizing can appear as a response to historically specific ways of knowing. It will argue that the received understanding of mental functions and their composition is not, as that understanding represents itself, “natural,” but is a historically contingent approach to self-understanding. This work will not represent connectionism as other than historically contingent.

Connectionism is understandable only as a historically specific way of viewing ourselves formed in direct response to problems created by the contradictions of the traditional understanding, and is arrived at by using tools developed within this tradition.

In pursuing the goal of placing connectionism against a cultural and historical background which makes it sensible, this section will trace its history on two widely divergent time scales. First, a broad historical overview will be offered. This will review the rise of the current way of understanding ourselves and the world, with a focus on tracking the broad shape of a coalescing understanding of mental activity that is the received or traditional interpretation. Second, against this background, the specific interdisciplinary
field of cognitive science from which connectionism emerges will be described. Battles within this field recapitulate positions—and outcomes—familiar from the larger historical context.

The Current Context

The pursuit of a goal such as situating connectionism does not appear simply as the disinterested intellectual “play” of a postmodern analytic. Rather, the need for such an exploration reveals itself in the dramatic clashes taking place within the disciplines concerned with understanding the realm of the human. One could, and many have, noted the turmoil in the scientific disciplines. The verities of our culture have lost their quality as truths. Newtonian physics gives way to quantum uncertainty. The center of scientific theorization shifts to the unstable ground of biology and evolutionary theory.18 Chemistry lays claim to fundamental principles of the organization of matter in the rise of nonlinear dynamics. There is a thriving academic industry in pointing out the collapse of traditional patterns of explanation in specific disciplines and the blurring of disciplinary boundaries themselves.19

Certainly these shifts are important, but they are shifts in our communities’ explanation of the world; they treat the nature of human understanding as a background factor. This particular discussion will focus more narrowly on the way this conflict manifests itself in the realm of cognitive science and academic psychology. The educator finds this realm of special importance because it is the well from which the practice of education is drawn—or at least the way in

18 M. Grene, The Know er and the Known (Los Angeles: University of California Press, 1974).

which teachers justify their practice. Our culture invests these disciplines with the authority to understand ourselves and, in some ways more pointedly, to understand the way we understand our world. At risk in the disorder within academic psychology and the newly ascendant cognitive sciences is our very understanding of ourselves. The terms of the disagreement could scarcely be more starkly dramatic. On the one hand, formalists such as Fodor\textsuperscript{20}, Papert\textsuperscript{21}, and Minsky\textsuperscript{22} accuse the insurgent connectionists of various forms of romanticism, behaviorism, gestaltism, and of being merely fashionable. Most tellingly, the repeated assertion is that the connectionists partake of the "holistic heresy."\textsuperscript{23} The thrust of their attack is that in one way or another—or in many ways—the connectionist resurgence is unscientifically motivated and methodologically suspect. For their part, connectionists and their allies accuse formalists of being deliberately blind to the interconnectedness of the material world\textsuperscript{24} and of being even more blind to the implications of humanity's own way of being situated in the world.\textsuperscript{25} They believe that the insistence that the formalists make on inviolable levels of analysis amounts to little more than


\textsuperscript{21} Seymour Papert, "One AI or Many?" \textit{Daedalus} (Winter, 1988): 1-14.


\textsuperscript{23} Ibid.


scientific dogmatism—a dogmatism that flies in the face of the history of science.26

Invective aside, the grounds for disagreement are substantial. The formalists take the traditional position that human thought is essentially logical. They hold the “physical symbol hypothesis;” Newell and Simon, the formulators of this phrase, define it succinctly and canonically as:

The Physical Symbol Hypothesis. A physical symbol system has the necessary and sufficient means for general intelligent action.
By necessary we mean that any system which exhibits general intelligence will prove upon analysis to be a physical symbol system. By “sufficient” we mean that any physical symbol system of sufficient size can be organized further to exhibit general intelligence.27

The terms in which this definition is cast are revealing because they refer to the traditional, logical way in which cause is established. A thing or event is said to be caused by another thing or event if it can be established that it is necessary (empirically, always co-occurring) and sufficient (no other thing or event is necessary). Not so subtly, the claim is being made that one need look no further for the “cause” of intelligence than a complex physical symbol system, and that intelligence is an expression of such a system. This, coupled by theorists with Turing’s famous demonstration of the universality of his abstract computational device, has been taken to mean that there is no need to consider particular instances—that there is no need to understand human intelligence in its particular, embodied form.

In its purest form this constitutes the essential position of the formalists. In one set of concessions, the formalist camp has come to the conclusion that a system must not only have the form of a physical symbol system referred to


27 Dreyfus and Dreyfus, "Making a Mind versus Modeling the Brain" 10.
above but must also possess at least some degree of context-specific information to manifest intelligent behavior.\textsuperscript{28} Much of the development of the field has surrounded the development of data structures which purport to model memory such as frames, schemata, and scripts.\textsuperscript{29} In a nutshell then, the formalists believe what I have labeled above the "theory theory." They believe that productive thought is a matter of logical operations over factual objects. The model, ultimately, is one of geometric proof.\textsuperscript{30} According to their account intelligence is an abstract quality divorced from the material substrate in which it happens to be found.

The task they set themselves is to model the logic which causes intelligent behavior in their scheme. They want to know how knowledge is organized to be useful.

Connectionists, on the other hand, stand opposed to such a formulation. They believe logical cause has little to do with the process which subserves intelligent behavior. Instead, they rely on the particular material matrix in which we find the phenomena which they believe does subserve intelligent behavior: the brain. For them, thought is not logical at its root; it is associative. Metaphorically, not geometry but poetry is the basis for human reason in their

\textsuperscript{28} The most extreme extant example is "CYC," a project headed by Douglas Lenat, which is literally attempting to encode all the semantic relationships necessary to reproduce "understanding." See the review of this work in: Jim Barnett, Kevin Kight, Inderjeet Mani, and Elaine Rich, "Knowledge and Natural Language Processing," \textit{Communications of the ACM} 33 (August, 1990): 30-49.


\textsuperscript{30} For a discussion of this genealogy from a point of view sympathetic to the Descartian roots of the project of traditional artificial intelligence, see: John Haugeland, \textit{Artificial Intelligence: The Very Idea} (Cambridge, MA: MIT Press, 1985), 28-36.
formulation. Intelligent behavior is intimately associated with the relationship between the actual neuronal architecture of the brain. While logic may be conceded to be sufficient in the sense that anything may be described by simple logical processes, connectionists aver that classical, formal logic is demonstrably not necessary; the human brain is an existence proof that other architectures can support intelligence. If Turing’s assumption of unlimited space and time is discarded, it is not sufficient—the human brain has neither enough time nor enough address space to work as ponderously as the logical story would suggest.

The task connectionists set for themselves is to model the material organization of the brain. They want to know how humans can learn from their experience in the world.

It is tempting to say that this opposition was inevitable; in the intellectual history of the West, the shifting configurations of opposition between faith and empiricism as the ground for understanding, between the material and the ideal as cause, between logic and poetry as making meaning, were bound to shake out this way at least once. But this is not a simple matter of a distorted opposition between Snow’s two cultures. It takes place entirely within the privileged realm of scientific discourse, and the participants agree on fundamental matters essential to that discursive community. They agree that


32 This point is discussed in more detail in chapter 3 of the present work. See also: John A. St. Julien, “New Understandings of Cognition,” Paper presented at the Tenth Conference on Curriculum Theory and Classroom Practice, Dayton, OH, 26-29 October 1988.

their task is one of material explanation, that human thought is in no way ineffable. They agree that the scientific method—what Dewey understood as disciplined common sense—should be the way in which such an answer is sought. Neither an external, unique cause in a concept of god nor an internal, unique reference to the human spirit is allowed. They even agree on the basis of their disagreement, with each accusing the other of being some specie of idealist; they simply disagree about what justifies the charge. This disagreement does not represent a polar division within the larger scientific community. The participants dissent together from some canonical positions that shape science. Most notably, they agree on the value of modeling as both a theoretical and empirical enterprise. That is, they are both willing, contra the scientific mainstream, to take a felicitously designed model as theory in that it can be a representation of the real which implies surprising, testable events. A model may also be taken as empirical data about a system in that the behavior of a well-tested one is taken as a legitimate reason to take action in the world.

This, then, is a sharply focused disagreement among participants loyal to a single strand of scientific discourse. It can be seen as a battle for the cultural high ground of science over the issue of what constitutes productive human thought.

What I want to ask, is how did it come to be this way? What makes their opposition possible, sensible, and contentious?

A Discursive History

This is a historical question. It asks how people, cognitive scientists among them, came to understand themselves and their fellows in such a way that this particular disagreement is both possible and compelling. To answer such a
question, we need to examine histories that focus on the way we understand our own cognition. This is a markedly different focus from that which informs more familiar histories which chart the transition to our times. We are more familiar with histories which focus on the products of science (Galileo’s theories as the start of modernity) or on political structures (the American and French Revolutions as the decisive break with medieval social organization) or on economic factors (the rise of the mercantile state or the appearance of the urban proletariat as the decisive moment shaping modern history). Our focus here needs to be on the transition to our current self-understanding, a self-understanding that connectionism problematizes.

A number of scholars of intellectual history have come to the conclusion that the view a people hold of their own intellectual capabilities is the hallmark of an intellectual age. In their differing ways, such authors as Reiss, Toulmin, and MacIntyre focus on different aspects of how the westerners came to hold their present self-understanding. Reiss focuses on the move from a previous discourse of patterning to our current analytico-referential discourse. \(^{34}\) MacIntyre traces four European cultural traditions in which the meaning of rationality varies significantly. \(^{35}\) Toulmin studies what he sees as the shift from reason to rationality in the sixteenth century. \(^{36}\)

These commentators are particularly congenial for this study because they adopt a common material approach to history. By this I mean they avert only to structures of explanation which are firmly rooted in human history itself,


chiefly to practices and more specifically to discursive practices. They do not refer to some internal, lower level "human nature" or to some external, higher level deity or teleology in explaining their findings. This attitude marries well with the explanatory processes at work in cognitive science, which also refuses patterns of explanation that cordon off difficult questions by assigning them to a trivial "black box" or by defining certain issues as so much beyond our grasp as to be ineffable. This allegiance to material of cause in the realm of the mental constitutes cognitive science's most consistent break with prior patterns of explanation—and its greatest legacy.37

These writers can be read as producing complementary works, with Reiss examining major discursive shifts, and with Toulmin and MacIntyre respectively emphasizing discontinuity and continuity of change within our current discursive regime. Taking such a stance allows the reader to begin to make judgments about the importance of the discussion within cognitive science by offering a framework against which to understand the particular claims of each side.

The narrative that I will lay out here is one of a developing series of conceptions of human self-understanding, each of which takes shape against the backdrop of emergent difficulties in the previous mode of understanding. The general pattern is that each successive solution is accepted because it allows the participants to escape the dilemmas in which they find themselves enmeshed. But each solution creates its own difficulties and it is against the

---

37 Although the influence of Chomsky is clear in cognitive science his "black box" approach little influenced cognitive scientists. What they received from Chomsky was his reassertion of the unity of grammar (syntax) and logic. The generativity of grammar was a result of its logical properties. Cognitive scientists broke with Chomsky precisely in taking up the challenge (which Chomsky declined) of delineating which logical processes were within the black box.
background of those difficulties that subsequent solutions arise. MacIntyre has phrased it well:

The best theory so far is that which transcends the limitations of the previous best theory by providing the best explanation of that previous theory's failures and incoherences (as judged by the standards of that previous theory) and shows how to escape them.\(^{38}\)

I have emphasized the parenthetical phrase “as judged by the standards of that previous theory” to point to the historical necessity to explain an event in the light of its own circumstances rather than by reference to what it has produced. There is nothing inevitable about the conflict within cognitive science that this chapter attempts to make sensible. If connectionism can be used to restructure the current problematic more usefully, this does not make connectionism a final solution, only a contingent resolution. Indeed, insofar as it acts to reconfigure the border between knowledge and learning, connectionist perspectives reveal the need for a more adequate understanding of knowledge in a dramatic way. Situated cognition as a way of understanding the social constitution of knowledge will be addressed in chapter 4 below.

**Reiss’s Narration of the Rise of the Analytico-Referential Discourse**

I have argued above that cognitive researchers are faced with a cultural preunderstanding of their task which shapes their research and offers them problems that are important in this canonical scheme. Classic research in cognitive science has treated this given framework as “natural.” Reiss’s work argues convincingly that this framework is not natural, and charts the transition to our current view of ourselves.

Reiss explains that new discursive structures arise out of contradictions inherent in the old pattern of understanding. As a discourse works out its own implications in the practices of those that participate in it, contradictions that were previously only implicit are made explicit, and the failure of the discourse to adequately handle all the issues that are created within its discursive frame becomes apparent. Slowly, but according to Reiss irresistibly, those in the community who encounter the contradictions become aware of the inadequacy of the formulation with which the then-current discourse frames the problem. Awareness of the very discursive structures through which sense is made allows the objectification of the current patterns of understanding and interpretation.39 Only then can an alternate pattern of explanation arise which provides a better “explanation of that previous theory’s failures and incoherences . . . and shows how to escape them.”40

In the particular case of the discursive formation previous to our own, Reiss, following Levi-Strauss, points to a form of reference which does not privilege the separation of the interior and the exterior of the enunciator; discourse is a part of the world and not separate from it. Such a discourse cannot distinguish between sign and object; all objects are signs and all signs, objects. The world is ordered by a resemblance in which the name of the object and the object itself are not distinguished. Knowledge consists of “a discursive exchange within the world” and is essentially interpretive.41 This discourse does not take into account the arbitrary nature of the sign. This pattern of thinking and reference has been discussed as “magical thinking” and “bricolage” by Levi-Strauss and

40 MacIntyre, “Moral Arguments,” 222-223.
41 Reiss, The Discourse of Modernism, 31.
others. For those steeped in such a pattern of thought, lying is an epistemological problem, not simply a moral one. Similarly, manipulating language should be as effective as manipulating objects—hence magical “incantations.” From our current vantage point, we can anticipate how these assumptions brought with them real problems in practice.

In such a system, much that we find commonplace or interpret very differently must be occluded, hidden. Lying, for instance, was a violation of the proper order of things. It interrupts the unity of the world and suggests that a lie may become the truth, overturning that order. This ties closely to the question of “magic.” Magic, within the discourse of resemblance, was not, as we tend to think of it, a matter of using language to affect the world though some mystical tie between the incantive use of the word and the world. Rather, under the discourse of resemblance, the sign and the object were not habitually regarded as separate and the resources of the culture did not make the difference available. But lying is possible and magic is not.

As it becomes increasingly obvious that the word and the object referred to are not inseparable, a space is opened up between the sign and the object that allows the sign and the language system as a whole to be detached from the world. There is a crisis of reference. Suddenly one is not sure what words refer to or upon what principles that reference is ordered.

The contradictions of the discourse of resemblance generated a crisis of representation in which a new discursive structure is possible. But in order for any new structure to gain legitimacy and displace the old, it must offer a resolution to the crisis. It must respond with a referential system which explains the relationship of words to the world and offer a set of principles

which orders that relationship. Our current “analytico-referential” is such a response. Reiss labors to demonstrate that in the intellectual life of the 1500s the issue of reference was a source of great anxiety, and he locates the rise of our current forms of both science and literature in the attempt to save the system of reference.

Reiss claims (and I will not repeat his complex analysis here) that the sixteenth century resolution to this crisis is to grant a separate status to the sign apart from the world. This allows a conception of the relationship between these two new entities in which they are yoked together as parts of the same unity. Lying is possible and is no longer an epistemological problem but only a moral one. One may speak incorrectly about the world without threatening the order of it, making the relationship between description and the world opaque and giving rise to the possibility of what today we call science. The crucial remaining task was to make the relationship between language and the world reliable once again. This is the crux of the crisis of reference. The sixteenth century repaired the wound which was opened up by separating language and the world by making the relationship strictly parallel by equating the syntax of language with the causality in the world. The tool that bridges this gap is logic, understood as equally descriptive of grammar and of the workings of the world.

The new analytico-referential discourse, then, rests on an identity between the syntax of language and the workings of the world based on the structure of logic. Crucially, this is not held to be an analogy. It is identity. The formal relations that hold between elements of language (grammar), logic (in the Euclidean, deductive sense), and the world (science) are held to be the same. Language and the world are no longer conceived of as being the same thing, but are now understood as properly corresponding in a set, logical way. In
attempting to repair the deficits of the previous discourse of resemblance, the new analytico-referential discourse emphasizes that the way language references the world is that language and the world share a pattern exemplified by formal, Euclidean logic. "It allows the world of phenomena and of concepts to be serialized into a grammar." Once serialized, it can be logically analyzed, and any resulting deductions can be taken to be true of the world. The speaker stands outside this process and enunciates the sentence. This argues that logic, the world, and language are but different expressions of the same underlying order. Thus, one knows a thing when one can describe it. One expects that the world is adequately describable in words. One expects that operations performed logically on objects identified in language will prove true of the world. And one knows that causal entailment in the world is simply logical. One knows that there is but a single order and a single truth. If this seems natural, it is because, in Reiss's account, this stance underpins the present dominant discourse, the analytico-referential. It must feel comfortable to us. But like any discourse, it contains its own contradictions.

What the discourse of analysis and reference ignores is that while the sign is arbitrary, the map is not the territory mapped. The relations that govern language are not the relations that govern the world. Language is not identical to logic nor is logic a full formalization of the relations that govern the world. That we can recognize these statements as true, and at the same time feel that the framework they contradict is natural, is evidence that we occupy a historical moment in which reference is again in crisis.

43 Reiss, The Discourse of Modernism, 32.

44 This is a reference to Alfred Korzybski's famous dictum: "The map is not the territory and the name is not the thing named." Alfred Korzybski, Science and Sanity (New York: Science Press, 1941), 58.
At some level, concrete practices in the world must take these contradictions into account and mitigate their effects. And if the analytico-referential discourse is to dominate, these sustaining practices must be occluded. Reiss’s discussion suggests that just as the analytico-referential discourse grew out of the elements that were occluded in sustaining the discourse of resemblance, so also are the elements that are occluded in the present discourse likely to hold the key to understanding the fracture points of the present discourse. Reiss’s analysis suggests that one would expect to see a challenge to the unity of sign, logic and the world. Such a challenge would be expected to emerge from an elaboration of the present system. There is no “outside” from which it can originate. Just as the elaboration of the resemblance mode of knowing led to the revelation of the occultations upon which the mode depended and so opened the way to the rise of the analytico-referential discourse, so would we expect the present discourse to be elaborated in such a way as to reveal its own occulted practices.

Toulmin’s Narration of the Development of Modernity

But Reiss’s analysis does not, itself, bring us to a point from which we can understand the background against which connectionism appears. Connectionism appears as a significant issue in the elaboration of the analytico-referential discourse. Reiss’s discussion gives us the building blocks and their broad relations but does not tell us what will be the final shape of a discourse described by these limits. Such a shape remains to be delineated by the specific history of actors attempting to work out the contradictions as they encounter them.

My analysis here will follow the same general pattern as my analysis of the emergence of the discourse as a whole. That is, I will describe the initial state of
the system before its elaboration, describe points of conflict, and discuss the ways the discourse was elaborated in response to the resultant problems. These elaborations, themselves intended to heal the gaps introduced by practice, will eventually lay bare the contradictions that were only implicit in the problems investigated. The tool used to make such a gap explicit may become the basis for the elaboration of a new discursive regime. In this way the rise of grammar in the sixteenth century was an attempt to heal the gap between the sign and the world but eventually functioned to make plain the gap and finally served as the basis for the shift to a different discursive framework.

I will use Toulmin’s work to chart a more specific history of the elaboration of the discursive structures that Reiss discusses. Some caution is in order here, as the intellectual frameworks that underpin Reiss’s Foucaultian project and Toulmin’s philosophical history of science differ, as do the pragmatic ends which shape their texts. But these same differences should give us some confidence about the area in which their analysis is parallel. Their analyses share an interest in the rise of the modern intellectual order and locate the beginnings of that rise in the fifteenth century. Both accounts find literary and scientific accounts of this era pointing toward a fundamental change in the way the world and thereby the self is understood. Their differences may be viewed, in part, as resulting from their differing interests—interests which in the context of the present work are complementary. Reiss is interested almost solely in the transition between discursive orders. Toulmin, on the other hand, is deeply concerned with what he sees as the mistaken elaboration of that order. While we may not join Toulmin in characterizing the elaboration of modernity as an error, preferring instead to treat it as a fact of our history, his analysis advances our project of understanding the trajectory of the current order. Both writers describe a discursive structure increasingly dominated by mathesis, the attempt
to model all aspects of the world on the basis of formal logic, and a preference for generalizability and objectivity.

In summary, Reiss leaves us with a discursive order which is based on an identity between the structure of language, Euclidean logic, and causality in the material world. Because these identities will not hold up, and indeed have not held up, contradictions are said to exist in the current discourse. The tensions that arise through these inaccurate mappings need to be accommodated in some way if the discourse is to continue to operate. The most obvious solution is to make one element of the discursive structure dominant (even while occluding this dominance because to "see" it would be to admit the failure of the discourse to solve the problem of reference). This sort of either/or choice is built into the discourse by the adoption of a logical form based on Aristotle's exclusion of the middle. If language, logic, and the world seem not to be identical, then we must choose which one to privilege in ambiguous cases. The patterns of language or of logic or of material cause must rule the triad. We are shaped by the discourse within which we are embedded to choose between the good offices of literature, mathesis, and science to explain our experiences.

Toulmin's history of the first two centuries of modernity points toward a conflict between literature on one side and mathesis and science on the other. It is a conflict which the humanists such as Montaigne lost and which Leibniz and Newton won by successfully uniting logic and the material world. The hard certainties of the ratio and rationality came to be preferred over the humanistic judgment that was the basis of Reason.

Toulmin's account, following Dewey, focuses on René Descartes' ideas as exemplifying a mistake which has led Western understanding down an unproductive road. Toulmin, however, criticizes Dewey for failing to ask,
"Why did this transition take place just when it did?" Toulmin’s explanation centers on the thirty years of war and hardship that followed the assassination of Henry of Navarre. This difficult period is said to have led to a pervasive thirst for certainty. But history has been a record of just such periods of unrest and misery: in European history the black death of the mid-fourteenth century which coincided with the beginning of the Hundred Years War comes to mind. The fourteenth century was not, however, a period in which the entire intellectual community of Europe embraced a search for certainty. Toulmin himself fails to ask a further crucial question: "Just how was this transition possible; what are its conditions of possibility?"

Without an intellectual substrate which could support certainty—the absolute certainty of logical proof, not, for instance, the certainty of faith—as a solution to disorder, Descartes could not have emerged as a watershed intellectual figure. Reiss’s account makes it clear that the fourteenth century did not yet have the intellectual apparatus of the analytico-referential discourse which linked causality, logic, and grammar; that was developed in the sixteenth century. Without such apparatus, “certainty” in the modern sense was not available. Only the analytico-referential discourse makes sensible the equation of logic and causality which allows us to be “certain” of the deduction that follows from physical laws. But the development of this logical apparatus in the sixteenth century makes it available to the seventeenth—and Descartes becomes possible.

---

45 Toulmin, _Cosmopolis_, 44.

46 A certainty which is undermined by our further study of the material world. We now regard Newtonian laws, by and large, as artifacts of basically stochastic relationships.
Toulmin charts the intricate conflict between the humanist attitudes of early modernity where particular cases, timeliness, and complexity were accorded as legitimate a status as the global, timeless, and linearly ordered solutions associated with geometry and syllogistic logic.\textsuperscript{47} Descartes was, and thought of himself as, a geometer. “He claimed he was in the habit of turning all problems into problems of geometry.”\textsuperscript{48} The tool of geometry seems to fill the role in elaborating the discourse that Reiss claims grammar filled in its development. Grammar had pointed to seriality, syntax, and regular transformative rules. This attempt to rescue reference led to an equation of Euclidean logic, material cause, and grammar on the basis of their similarity along this formal axis. As this uneasy equivalence almost immediately started to unravel, the pattern had to be maintained by occluding the problems with it. What occurred was that mathesis was accorded the dominant status but that this dominant status had to be obscured. The other two elements which formed the basis of the discourse had to be either subsumed or discredited—always without acknowledging that this was what was happening.

What emerged was a system in which what we now understand as the literary was pushed to the periphery, and its role in the support of mathesis and science was, as Derrida and others have recently pointed out, unacknowledged and even denied. On the other hand, mathesis and science were conflated. Logic and the experimental (scientific) method were understood as the same process. The differences between logic and science have recently been highlighted by historians and philosophers of science such as Kuhn, Holton,

\textsuperscript{47} Toulmin, \textit{Cosmopolis}, 30-44.

and Ernst Mayr. The results of further investigations into the relations that actually govern the material world—beginning with Einstein and continuing with Prigogine's work in chemistry—demonstrate that these relations are not, strictly speaking, logical.

Toulmin's contrast between the reasonableness of the early modern era and the rationality of its development in the seventeenth century prepares us to understand the situation in which modern cognitive science, and the intellectual background which shaped it, arose. Formal logic, with its universality, timelessness, and axiomatic basis, replaced practical judgment's emphasis on particularity, timeliness, and concrete diversity.

In such a context, every area of human endeavor was judged by its congruence with this rationalist model. Following this account, then, it is not simply "scientism" to adhere to rationalist standards. Indeed, this pattern underlies the dominant form of late modernism and as such informs all areas of life. Folk psychology, for instance, participates fully in the results of Descartes' solution to the problem of knowledge. We have no trouble

---


51 See MacIntyre for a healthy example of the conflictual and incomplete way that discursive elements within a broad tradition may differ. No tradition, including the tradition of Western rationalism, is unitary—Western rationalism is notable in that it needs to claim that its tradition is unitary to substantiate its universalist claims. Alasdair MacIntyre, *After Virtue* (Notre Dame, IN: University of Notre Dame Press, 1981).

52 See the illuminating work by Stich which traces this relation and draws out the implications for current work in cognitive psychology: Stephen P. Stich, *From Folk Psychology to Cognitive Science: The Case Against Belief* (Cambridge, MA: MIT Press, 1983).
understanding rationality as something which is objective: that which is formal and universal. Error is associated with the body in opposition to the mind. Thus, we explain emotionality as one of the chief causes of faulty reasoning. We commonly insist that another source of error is the failure to get our facts right, that is, to adequately secure the axiomatic basis which underpins our reasoning. Science and the science of mind with which we will be concerned, do not arise in opposition to folk or common understanding. Science is, as Dewey clearly understood, only a rigorous extension of our common sense.

At this juncture we begin to close in on the object of our quest: to understand the background against which connectionism appears and thereby to understand the depth of the implications that it holds for our culture's conception of thought and learning. Mathesis, the attempt to formalize all relationships on a model of formal logic, as has been argued above, has been a basic move in formation of the current discursive structure. As this discourse has evolved, Descartes' division of phenomena into mental and material realms has been a crucial and discourse-shaping move. It is another expression of the discursive underpinning that identifies the same pattern of relation at the bottom of all phenomena and postulates parallel but directly analogous logics which underlie the various realms. Descartes' move to separate mind from body and to identify mind with pure logic, and his failure to see a problem of coordination between body and mind, are understandable within the discursive history sketched above. It is simply a recapitulation, in a different but complementary form, of the idea that language, the world, and logic are parallel and noninteracting systems which nonetheless stay in synch because of their identical underlying form and processes. Descartes, of course, believed the world to be as logical, fundamentally, as the mind; it is this congruence which made science, and particularly his geometrically-based mathesis in the form of
the calculus, a sensible enterprise. Descartes' move, and his identification of the logical with the mind, was perhaps the crucial move in the development of the late modern discursive structure. Without it, and his companion translation of geometric logic into the calculus, the rise of modern science as a central activity of modernity is not understandable. The mind, being logical, can know the world which is organized logically, and can express itself in language, logically. Descartes' introjection of logic as the order of mind and his uncoupling of geometry from the form of the syllogism (and thereby the entanglements of language) made formal logic an "objective" tool. Without this formalized, unified basis for "proving" a thesis about the world with the certainty of geometry, "certainty" as we know it in late modernity—the certainty that Dewey and Toulmin protest—is not available.

Thus, assumptions about the nature of the mind have been crucial to the development of the modern discourse. It is not by accident that Descartes postulated an immaterial mind with only the rhetorical use of the tiny pineal gland as the passageway between mind and world. The original formulation of modernity declared language, formal logic, and the world to be governed by the same set of principles. These were bridges over the chasm of meaninglessness opened up by the collapse of the discourse of patterning. But they were only bridges, airy and insubstantial supports over a great gulf. Descartes helped secure modernity by obscuring that terrifying chasm. He made logic the basis for the operation of human mind, making possible a more secure obscuration of the pit. No longer was humanity attempting to coordinate three external systems—however alike—that were foreign to human understanding. Instead, "man" was essentially logical, a being whose very essence was to reach out and know the logically ordered world through the offices of his most godlike facility: logic. Finally, the logical nature of language
would assure that he would to be able to reliably represent that knowledge. The terrifying gap that humanity had to bridge coordinating three independent systems was closed to the minuscule one represented by the pineal gland. There, hidden deep within the brain, the transaction between a logical if insubstantial mind and the equally logical substance of the world could be assumed to be natural. It is this relationship which makes "rational" the long-time assumption that in some sense we can know the world directly. If the mind is assumed to be as logical as the world, language, or logic itself, there is no real problem of coordination—only problems of confusion when the mind is distracted by material events. In a very real sense the mind can be said to already "know" the logical relations that constitute the world. Descartian introspection, then, suggests only that we let the mind do its task unencumbered by an actual engagement with a confusing world. It is in this way, Descartes hopes, that meaning will be most securely grasped.

Mind, Thought, and Logic: The Development of an Identity

The identity introduced between thought and logic in the Descartian scheme makes more sensible the sometimes puzzling association of computing and neurology in Western self-understanding. In the present account, formal logic, like Reiss's grammar and geometry in Descartes' analyses, is a tool that takes shape in the attempt to deal with problems produced by the then-current discursive contradictions. But like these earlier tools, the elaboration of the area of contradiction does not, finally, serve to close the gap, but only to reopen the wound. The elaboration of this tool has served to demonstrate how limited a tool it actually is, and how little suited it is to the role of unifying the current discourse in which it has been cast.
The exaltation of formal logic has been a dominant thread in the intellectual history of the West with Descartes’ body/mind dualism and Leibniz’s monads representing extreme and influential outgrowths of the belief in a “soul,” a human essence, which was strictly logical. A further, influential development of this line of reasoning was Boole’s famous book *The Laws of Thought*, which set forth the now famous “Boolean logic.” This development is crucial in the development of the ideas this chapter traces. Following Boole’s lead, Russell and Whitehead formulated their *Principia Mathematica*, which was intended to provide a rigorous, complete, bi-valued basis in logic for mathematics and all knowledge. Their project failed to meet the challenge of Gödel’s incompleteness theorem and was abandoned by its authors before completion. Gödel’s theorem constituted a fundamental challenge to the Cartesian project shared by virtually all the mathematicians and logicians of the day. It proved that any systemization strong enough to account for arithmetic must be incomplete.

The choice which faced logicians was galling: they had to choose between strength and completeness. The Cartesian project of logical totalization fails. Gödel’s theorem strikes directly at the discursive basis for totalization via mathesis—formal logic—for if logic is a *limited* system, it will always remain inadequate to model the seemingly limitless world or to be the sole defining feature of the mind. It was particularly galling to discover that formal logic was not only not adequate to the task set it by the discursive assumptions, but it was not even adequate to model everyday arithmetic. The one-to-one correspondence between formal logic and the world that both modern forms of scientific mathesis and the discourse itself assumed is simply not available.

53 The title of this book stands as a succinct exposition of the conviction of the rationalist school that logic and thought were one and the same. George Boole, *The Laws of Thought* (Chicago: Open Court, 1854/1940).
Bringing us to the focus of the present day, machine computation and computers, Turing reacted to Gödel’s dashing of Russell’s hopes by embarking on his own project to cordon off "undecidable" propositions. He hoped that by isolating undecidable propositions, he could rescue the Cartesian. He failed. His failure, though, produced the famous Turing machine thought experiment. This paper is usually remembered for its proof that such a machine was universal and could solve any formally statable problem. The “universal Turing machine” idea—that such a machine could solve any well-formed problem—has functioned to legitimate computer science in general and computational cognitive science in particular. But the idea of such a generally powerful computational device actually appears in Turing’s arguments as a means to show that even such an ideal machine would fall prey to the limits Gödel had intimated. The universal Turing machine, the acknowledged conceptual forerunner of computers, actually seals off the possibility that all knowledge could be quantified. The modern serial digital computer is directly modeled on Turing’s conception. The irony, an irony which points to our culture’s lack of awareness of its own intellectual history, is that while the Turing machine is a direct outgrowth of the failure of the Cartesian project to totalize all understanding under the banner of logic and mathematics, computers based on its principles have been used largely in pursuit of that very goal. The history of recent research in artificial intelligence is a history of the ignorance of this basic philosophical contradiction.

A final element in the social history of modern computing was provided by von Neumann and the idea of program, which simply allowed the machine’s


55 See Gardner, *The Mind’s New Science*, 18. This idea is often and correctly attributed to Babbage and his sponsor, Lovelace. The point made here however is not a question of primacy,
instructions to be stored within the machine. With this step and Putnam’s pointing out that such programs could be considered mind and would be implementable on any Turing machine (body), a full-blown model of mind based solely in the dictates of formal logic emerged. This, often seen as a pivotal insight, simply replays Descartes’ division of mind from the material plane of the body and the accompanying substitution of logic for material relations that Descartes placed at the center of our discursive structures. The essential, crucial difference is that this form of philosophic rationalism could be put to the empirical test. Steeped in the discourse in a way that its originators could never be, the modern cognitive scientist, unconsciously certain that logic was the way the world worked, was unable to see the need to maintain the immateriality of logic that Descartes insisted on. This discursive “conflation” and the resultant confidence in the metaphor of the mind as a program and the computer as a brain was to lead directly to the possibility of artificial intelligence. This, as we have begun to see, was a new tool that was fated not to explicate and solidify our dominant discursive practices but to erode them.

but of social history. In the context of modern computing it is Von Neuman’s insight and the way that he contextualized it that determined the subsequent course of its use.


58 It is striking to see the two historically contentious philosophies of rationalism and empiricism confront each other again. It does not bode well for rationalism that the current situation requires that the battle be fought on empiricism’s home ground. The historical irony is extended if we notice that again, the specific grounds for contention lie in whether knowledge is gained by reason or experience. Hobbes vs. Descartes redux! (It is indicative of the strength of the rationalistic tradition that while these worthies disagreed on what they felt were fundamental grounds, they both assumed that formal logic was the foundation of thought, a point currently under debate.)
The predominant model of mind in cognitive psychology is currently based in a computational model labeled “the serial digital computer.” Such explanations have proceeded from the powerful early explorations of such as Turing and von Neuman, whose work helped open the possibility of creating logic in the material world by imagining a machine that could compute—a “computer.” In a world that finds such an idea commonplace, it cannot be emphasized enough what a radical insight this was. Prior to the development of logical machines, the classic method of distinguishing men from the rest of creation was reason, and in this view mathematics was often seen as the pinnacle of human achievement. The “von Neuman architecture” which used on/off logic gates, logical programs, and local, site-addressable memory, was a combination of then-current cutting-edge logic and the best ideas of how the brain functioned. At the time, there was assumed to be no contradiction. The research trajectory which demonstrated the invalidity of this assumption is one way to understand the story of the rise of cognitive science.

The Rise of Cognitive Science

The history of the rise of cognitive science and the connectionist reaction parallels the story that Toulmin tells about modernity as a whole. The field arose in a moment of agreement; a wide range of methods and a general agreement about the task were initially evident. But as the field matured, one side of the discourse, the one that focused on the particular nature of learning and that drew on biological and statistical metaphors, was displaced by a logicist model that insisted on the purities of axiomatic, propositional logic and
assailed the holistic heresy that it claimed\textsuperscript{59}—accurately—its sibling represented.

Clearly, the rise of computers raised questions as to the nature of mind. This was accentuated by the fact that the machine said to be the first digital computer was modeled on the classic McCulloch-Pitts neuron.\textsuperscript{60} There is humor in this since a machine modeled, in part, upon the then-prevailing models of human brain function came to be the predominant model of mind. As our understanding of the brain has progressed further, this model of mind has come under attack as being falsely modeled on a machine!

A critical juncture in the history of computing arose at this point when the analog computer, a computer which directly modeled the problem rather than solving equations which described it, lost out in the competition to become the tool of choice in computing. A key factor was the development of reliable transistors to provide a stable basis for the bi-valued computation that Russell had developed and for which Turing had suggested a physical implementation. It is worth noting that as soon as it became practical, the more culturally acceptable, logical formulation of thought which embodied the mind/body dualism was chosen. Analog computers were perfectly possible to construct with the new technology but were not seen as desirable if one could have “the real thing.” Analog computers and the associated processes of modeling the problem rather than “solving” it via formalisms were to undergo a long eclipse.

Operating within the world view that formal logic was at the basis of thought, computer programs running on serial digital machines became tools

\textsuperscript{59} Minsky and Papert, Perceptrons, 19-20.

for implementing the theories of researchers in cognitive psychology and related fields. The tool proved very powerful and was quite successful at exhibiting intelligent (in the sense of logical problem solving) behavior. Almost all the effort in implementing these theories went into writing the software, or deciding what representations and processes could be used to produce the desired results. This was reasonable; hardware considerations were practically nil (and this is only slowly beginning to change). Quite naturally, the program became a metaphor for the mind, the computer a metaphor for the brain. The metaphors were developed as models, models which were seen as successful because they were able to handle what was then felt to be the pinnacle of human achievement: logic.

These powerful and successful models carry with them an implicit set of assumptions based on the nature of the serial digital computer and its characteristic programs. The serial model presupposes a machine based on a central processing unit which sequentially addresses a passive, discrete, site-locatable memory store by means of a program. This program consists of a series of commands which manipulate abstract symbols according to rules; these rules are, in turn, only symbols themselves. Thus the mind, and by extension people, come to be seen as symbol-processing machines. Further, the sequential nature of the commands necessitates the use of a hierarchical logic structure.

In this way, an internally consistent image of mind and brain and their relationship arose which was in alignment with many of the deepest presuppositions of our culture regarding this relationship—an image we now take as reality.

Emergent Problems in Cognitive Science

As with any paradigm successful enough to attract widespread acceptance and large amounts of research based on its insights, problems have arisen. Those problems with a particular educational importance will be discussed below. Among these are those arising directly from the neurology of the human brain: time problems and space constraints. Neither does the dominant, logicist paradigm deal adequately with a criticism which many psychological theories have proved vulnerable: that by failing to specify a mechanism by which thought could occur in the brain, any such theory has a hidden assumption of a "little man" to categorize and judge—the homunculus.

All the problems mentioned here relate in a greater or lesser degree to the particularizing move of situating thought in the human brain. There is a strong functionalist argument which states that the neurological problem is a phantasm and that the particular implementation does not matter. Fodor is probably the chief proponent of the view. In his article on the mind-body problem (1981) he points out that any software can theoretically be implemented on any Turing machine. This is certainly the overwhelming conclusion of computational theory and is the point for which Turing is usually remembered. Any well-formed problem can be solved on any Turing machine.  

---

62 A well-structured problem is one that can be stated in terms which can be examined via logic. Generally this means that we know what would define a good solution and we know the constraints upon finding that solution. As such, "well-structured" problems exclude paradox and vague, incomplete, and poorly parsed problems. In practice, well-structured problems appear suspiciously predigested, with factors such as attention, perception, and motivation entered as pre-given. Considered in this light it is suggestive that most school problems are just such predigested, "well-structured" problems—and most other problems are not. As Dewey and others have recognized, imposing an appropriate structure on the diffuse problem area is the first and, as artificial intelligence workers have discovered, often the most difficult part of handling difficulties.
machine. To go further and conclude, as Fodor does, that quarrels about how mind is implemented are therefore irrelevant, does not necessarily follow.

The conclusion seems to fail on two points. First, it assumes that the phenomena of interest, "the mind," is completely capturable by the logic of a Turing machine. This assumption, as we have seen, is entirely in line with our current discursive formation. Second, while the argument does show that any Turing machine can exhibit the phenomena of mind (given the above assumption), it does not follow that any mind can be reasonably implemented on any such device.63 As Toulmin’s analysis would suggest, particularity matters. Two key characteristics of Turing’s universal machine are that it has an infinite memory capacity and an infinite time in which to process. Turing’s proof of universality depends on these conditions. The human brain has neither. Cognitive psychologists—and educators—are interested in the phenomena of the human mind. The human mind is based in the human brain and it is this particular implementation which imposes limits on the models which may be reasonably postulated. The claim that will be made here is that the digital serial model is not a reasonable one in the particular context of the brain. Fodor’s claim that such particularization is trivial is backed by the weight of a traditional understanding which favors the search for universal, certain knowledge over the particularity of the problem. One response is to claim, with Toulmin and others, that it is not reasonable to dismiss particular cases, or to make basically timeless and placeless analyses of situations that depend crucially on both time and space. Another response, and one which such a tradition legitimates within its own understanding of knowing, is to show that

63 For an excellent review of some of the reasons to disbelieve the logical story of mind by a senior artificial intelligence researcher, see: Douglas R. Hofstadter, "Waking Up From the Boolean Dream," chap. in Metamagical Themas: Questing for the Essence of Mind and Pattern (Bantam: New York, 1985), 631-665.
this, as particular instance, does not conform. Time and space constraints, discussed below, will seek to substantiate these points.

Given that brain does matter, the question must be: in what way does our knowledge of the brain constrain our theorizing?

First, and perhaps foremost, the brain operates in parallel. Neurological evidence supporting this view is discussed in detail in Anderson and Hinton. Such evidence shows that there is no central processing unit to send out calls one after the other along a hierarchically organized tree of inquiry; rather, the evidence shows a complex network of interconnection without a central processing unit, a parallel architecture which makes many calls at once. Such parallelism has become a well-accepted basis of neurology.

Strictly speaking, this alone defines only the architecture as parallel. The programs that run on it could conceivably be serially ordered. Architecture is not destiny, and however inelegant the idea seems, the mind is not logically required to take advantage of the possibilities of parallel processing.

There are considerations, however, that seem to eliminate the serial model from being considered a realistic representation of how the human mind operates. These are time and space constraints.

The question of time rises from the observed speed of the processing units (neurons) in the brain. Neurons work on a millisecond basis, while typical computer gates work on a nanosecond basis. This means that computers' gates operate a million times faster than neurons. In its fundamental operations the

---

64 I am unwilling to go on and conclude, as those within this zeitgeist would presumably be willing to do, that this is the instance which disproves the general rule Fodor attempts to establish concerning the triviality of particularity—but such an approach is fascinating and I do not see how the logical formalist can avoid it; there is certainly no less trivial example than the way thought itself is constructed.

human brain is a million times slower than a computer. Therefore it is not surprising that computers take moments to return a computation that would take a person years to do, if the human could do it accurately at all. Yet there are many jobs which humans can do much faster; those involving pattern recognition, for instance, can be done in milliseconds in the human brain but take hours—where the recognition is possible—to be done by a computer.66 Something is clearly wrong with an explanatory model that takes orders of magnitude longer to complete an operation using machinery that is orders of magnitude faster. The human brain works far too slowly to be organized as the metaphor of the program suggests, for if a computer takes hours to do recognition tasks, the human should take centuries to accomplish the same task, given the disparity between the rates at which they work. The combined disparity between the known speeds of the basic "machine" and the observed differences between the two in pattern recognition tasks is approximately four orders of magnitude! No fiddling with a more efficient program will bridge a gap of such magnitude. The inherent possibilities of parallel processing must be utilized.

The problem of space similarly arises from consideration of the particular human brain. Lashly, as far back as 1950, in his oft referenced paper "In Search of the Engram," toted up a truly impressive amount of research—including thirty years of his life's work—that led him to conclude that memory traces located at any particular site were impossible in light of what was known of human brain physiology. His disproof proceeded from two bases. First, extensive destructive testing had failed to locate any single spot for any single

memory or skill. Even if truly massive portions of the cortex were destroyed, the skill was retained. Often this process could proceed to the edge of reducing the animal to a completely non-functional vegetable state before the "trace" was lost. Lashly's second line of reasoning was more radical. He stated that in his judgment, there weren't enough neurons or even synapses to encode all of the recall humans have access to at the available number of discrete sites. He observed that the large ratio of sensory to cortical neurons, coupled with the observed fact that virtually all of the brain could be observed as active during sensory input, foreclosed the possibility that there were any neurons left over to be dedicated to serving as memory sites. Lashly's almost despairing observations were to remain unchallenged, and largely unreconciled with prevailing theory, for almost twenty-five years. Thus physiological considerations make serial, site address models of mind untenable.

A persistent and even more long-lived dispute involves the concept of the homunculus. Psychological theories have traditionally been attacked, especially by those outside of the field, as implicitly calling for a "little man" to choose and to categorize. The critics felt that this was an intellectually suspect position since it only "hid" the problem of how decisions were made. The dominant school of artificial intelligence has claimed that its model eliminates the homunculus and that this is one of the strongest points of the perspective.67 In such a view, the program makes the decisions and there is nothing mysterious

67 See, for instance: Fodor, "The Mind-Body Problem," 124-133. More recent work, such as: John H. Holland, Keith J. Holyoak, Richard E. Nisbett, and Paul R. Thagard, Induction: Processes of Inference, Learning and Discovery (Boston: MIT Press, 1986), make it clear that even traditional artificial intelligence researchers feel the need to address the problem of learning. They do not, however, necessarily agree on how learning should be pursued. Thagard and Holyoak have been working within the connectionist paradigm; Keith J. Holyoak and Paul Thagard, "Analogical Mapping by Constraint Satisfaction," Cognitive Science 13 (July-Sept., 1989), 295-355. Holland, on the other hand, has pursued a more rule-based approach with genetic algorithms. See the account in: Russell Ruthen, "Adapting to Complexity," Scientific American 268 (January, 1993), 130-140.
about it; the operations of this program, which this research program sees as mind, are quite mechanistic and very transparent. It is not, however, clear that what earlier psychologists were accused of having hidden in space—in the brain—the computationalists who rely on the serial model have not hidden in time—via the program. Who programs the program? In the serial computer the answer is clear: at some point in time the programmer, who stands outside the computer, imposes order via the program. In a human no such programmer is apparent.

The appealing argument has been made that evolution provides the programmer. Certainly no computational model would be possible without the assumption of a basic “drive to organize.” Similarly, the case for certain behaviors’ being “programmed” or at least as existing before experience—fight or flight reactions and falling reactions, for instance—is unassailable. The case for even more complex behavioral patterns such as schizophrenia and personality structures appears to be building. Still, learning is difficult to see as programmed in this sense; a genetically-based understanding of Euclidean geometry seems unlikely. The great mass of what has interested people about their own cognition is clearly the class of things which is unknown but can be learned. What accounts for the flexible, autonomous, self-guided learning that we all experience and observe? What seems to be missing is a concrete mechanism for self organization. If we are to remain materialists and reject Descartian dualism, we will agree that such a mechanism must exist; almost as clearly, it must be based in human biology and its evolutionary history and its focal site must be the brain. The trouble here is that the serial model provides no such mechanism and in not addressing the problem leaves itself open to the accusation that the homunculus has not been vanquished, but has merely receded in time.
This is not to say that such a defect is characteristic of all computational models; one of the advantages of the interactive distributed memory discussed below is that it provides a self-organizing principle which clearly emerges from its assumptions about the human brain structure. Hopfield, among others, indicates that stable patterning in a relaxation response to input is an emergent characteristic of complex networks. Such distributed patternings spread across the whole field and serve as the analog of memory sites in the serial model.

Thus, given its failure to correlate with the known neurological characteristics of the brain, its failure to answer questions concerning the possibility of a program so organized actually running in the brain given the known time and space constraints, and its failure to adequately dismiss the homunculus, the serial model cannot be considered a valid model of brain functioning.

**Emergent Answers to Problems in Cognition**

Granted the serial model will not work, what will? What we are looking for here is a single unified solution to the problems of time, space, and the homunculus discussed above. A suitable solution would be fast, very fast, would have tremendously more memory than the site address model implies, and would exhibit a clear self-organizing principle which allows for learning without recourse to an externally originating program.

---

Research and thought on the problems with the serial model have tended to focus on one problem at a time. This paper will follow that pattern and will follow the development of these problems in an order which seems logical given the synthesis which seems to be emerging. It is worth noting that the apparent "logic" so imposed is not historically accurate. That is, some pieces—specifically distributed representations—developed long before the necessary theory of brain structure (parallel processing) to implement it arose.

The first problem is the neurological one: the brain does not function according to the patterns of a serial computer; rather it is a parallel processor. Bringing an idea of brain function in line with this fact immediately takes care of the problem of time. Parallel processors are fast. Perhaps more cogently, serial processors are of a necessity slow. Each command must be sent out to a specific site and the encoded response must be returned. Then and only then may further demands be made. Everyone must get in line and wait a turn. The larger the data base searched and the more commands sent, the worse the problem becomes. Parallel processors can, theoretically, be as fast as their gates—all demands are made simultaneously and all the requisite data is returned in one pass.

The obvious objection is that such a return must still be organized—how can a parallel processor be governed? Clearly, one solution is to place a serial processor over it to give it "order." Such a mixed system is conceptually possible, and some experimental data such as that reported in Anderson.

---

69 Paul Smolensky, "On the Proper Treatment of Connectionism," Behavioral and Brain Sciences 11 (March, 1988), 1-74. In this influential article Smolensky argues for a mixed model of human cognition but acknowledges the limits to the "depth" of processing which it is feasible to assume. Mixed models built on connectionist principles leave very little room for the sorts of programs which traditional artificial intelligence presumes.

concerning extended response time for increasingly complex computation problems can be interpreted as pointing to this. However, given the time constraints, it is clear that to explain adequately the problem of the rapidity of human recognition (children's recognition of familiar faces occurs in milliseconds), the "program" which retrieves the recognized object may contain no more than 100 "calls"—a daunting limit for any extensive serial governance.\textsuperscript{71} Very little in the way of symbolic logic can be done in such an extremely short program. Thus the role of serial governance in any mixed model must be very small in organizing human thought. If the old solution won't work, what will? An answer is suggested by the solution to the second problem with the traditional AI approach: memory.

The third pressing problem of the serial model is that of memory space. This problem is associated with site address and the parallel processing model does not address this problem directly. In fact, modern parallel processing "super computers," such as the Cray supercomputer, use site address (and a serial executive), and models based on such a computer would retain the memory problem. Lashly, in addition to being the first to succinctly state the problem, was also able to formulate what seems still to be the only answer: distributed representation. Such a model of memory depends upon abandoning the concept of site and substituting patterns of relationships as the "location" of representations. Distributed representation solves the space problem by increasing exponentially the amount of potential memory. Consider a simple situation in which there are four memory sites, four locations which are either "on" or "off." If these are treated separately, only four bits of information are encodable. If, however, these four are understood as constituting a unity, we

\textsuperscript{71} Hofstadter, "Waking Up From the Boolean Dream," 631-665.
are able to consider the pattern of relations between the parts that constitute the whole. There are sixteen such distinct patterns and sixteen bits of information may be stored. Increase this to a twenty-location unity and there 1,048,576 separate patterns of relations versus only twenty for its independent site address cousin. When one realizes that each neuron in the cortex is said to average 2,000 synaptic connections to other cells and that the Purkinje cells—implicated in information processing—have as many as 100,000, the impact of this exponential progression in the expansion of memory space is quite staggering. It is little wonder then that the idea appealed to Lashly, for it neatly solves his problem of memory space; he need not be limited to the simple one-to-one correspondence that site address entails.

Some mention should be made of the engineering objection to distributed representation. This criticism says that such randomly connected elements could never function reliably, since they would oscillate wildly. Even disregarding the real world disproof that the brain appears to do a fine job of utilizing such architecture, networking theory in every field militates against this objection. From ecology to classic systems theory, networked relationships are considered much more stable than their more nearly linear cousins. Hopfield's mathematical treatise is quoted as giving the definitive lie to the idea that such systems are not stable.

Taken separately, parallel processing and distributed representation seem to solve the time and memory problems of serial processing that are based in

---


neurological fact. It is a central thesis of this work that what appears to be occurring is the tentative emergence of a new psychological paradigm of brain functioning based on a synthesis of these ideas.

Such a synthesis would appear to be well-conceived, since each concept provides solutions for problems that its "brother" generates. Parallel processing, for instance, gives distributed representations somewhere to live. Though Lashly’s "In Search of the Engram" article\textsuperscript{75} arrived at distributed representations as a logical necessity, there was no way, under the then-current paradigm based on site address, to access such a representation. Parallel processing, inherently involving the simultaneous address of many sites, gave such an idea an intellectually viable substrate.

Additionally, distributed representations gives parallel processing, by raising the possibility of associative recall on the basis of similarity and of content addressable memory, possible primitives that would allow its coordination without extensive recourse to a serial executive. This allows the designer to make minimal use of a serial executive, the extensive use of which would work to negate parallel processing’s speed advantages over serial processing.

It is these new primitives of distributed representations that account for some of the most attractive characteristics of the new models. A particular memory site has no meaningful relationship to what it represents in the serial model; the meaning of a one or a zero in any particular location is completely arbitrary and any meaning that it can have is imparted by the program. A distributed representation is, on the other hand, inherently meaningful in the sense that the internal structure of the pattern formed in response to the

perceived object is not arbitrary, but bears a consistent relationship to the object that is perceived over time. Further, this internal structure of connectionist distributed representation, in which many differing representations may share large parts of their constitutive pattern, leads representations to interact.\textsuperscript{76} Similar symbols have similar patterns and therefore similar interactions. Any modification of the strength of the hardware connections comprising one symbol will tend to alter similar representations in the same way. It is this interactive aspect that yields such powerful implications as associatively organized memory, content-addressable memory, and the resultant self-organizing possibilities.

These new primitives radically reduce the demand for a program and the serially based executive in which it is based. The rules that such a program is based on are largely replaced by the structure given by the patterns of interactions between distributed representations, while its memory retrieval and relating mechanisms are replaced by content-addressable associative memory. Thus the last problem this paper originally proposed—the implication that a homunculus removed in time lurked in the concept of the program—has been diminished in scope by the substitution of the regularities imposed by emergent structure in interaction with its environment for the rules imposed by the program.

In this way the constellation of parallelism, distributed representation, and the new primitives which are implied by the first two factors’ interaction, suggest a unified and interdependent whole which avoids the pitfalls that make the traditional artificial intelligence model untenable as a way of understanding

\textsuperscript{76} Anderson and Hinton, 1981.
the functioning of the particularly human brain with which psychologists and educators are concerned.

Distributed representation, an emergent property of the network architectures we have discussed here, form a fundamental challenge to the discursive structures of modernity. It posits an alternative basis for cognition to the logic-centered story our tradition gives us and which has proven untenable. We buy the explanatory power of network architectures at a cost, though. We must give up the foundational discursive assumption that mind works according to the patterns of logic. The Cartesian center of a logical mind does not hold.

Conclusions

This chapter has traced the development of the discursive background against which connectionism has appeared as a dissenting field in the area of cognitive science. A chief thesis has been that connectionism, by dissenting from the canonical assumptions of the AI branch of cognitive science, is staking out a position which threatens the stability of the current discursive formation. Artificial intelligence took up the materialist challenge to close Descartes' gap located at the pineal gland by making logic material and locating it in the organization of matter. The concept of the serial computer as a brain and the program as the mind is dualism without Descartes' gap. This attempt to instantiate the mind as logic has failed and has the effect of undoing Descartes' identification of mind with logic. Bereft of the crucial assumption that the mind is logical, the further assurance of congruence between the underlying structures of language and the world are undone. Not coincidentally, both language and the world are no longer assumed to follow the pattern of formal
logic. Modern science, developed in response to the late modern assumption that the logical mind could know the logical world and express that understanding clearly within the logical structures of language, has eaten its own discursive base. The crisis of representation, noted in parts of the academy, is, on this account, real.

Education's response to such a situation is properly a practical one: that is, given such a situation, how do we discharge our obligation to educate students? Without a viable theory of thought, we are left uncertain as to what learning and knowledge are. How do our students learn? What is it that they know? The recent disputes within education, the extensive search for alternate methods, and finally the search for alternate theoretical bases are a testament to the urgency of this problem for educational practice. In the current account connectionism, itself a crucial factor in undermining the current understanding of how learning takes place, will be presented as an alternative to canonical understandings of "how they learn," and situated cognition will be taken as pointing toward a more adequate conception of "what it is that they know."
CHAPTER 3
Connectionism:
Computation and the Brain

[N]othing seems more possible to me than that people some day will come to the definite opinion that there is no copy in the . . . nervous system which corresponds to a particular thought, or a particular idea or memory.

—Ludwig Wittgenstein,
*Last Writings on the Philosophy of Psychology*, Vol. 1, 1982, para. 504 (66e)

Connectionism has been discussed in terms of learning theory and as a point in the trajectory of Western understanding of thought. In this chapter I will discuss the field itself. I will describe connectionism, work through an example of a connectionist network, and detail cognitive, neurological, and computational evidence regarding the plausibility of connectionism.

Connectionism Bounded

Connectionism in its recent form grows from treating two constraints seriously. First, connectionist theory focuses on the modeling of learning in a cognitively realistic fashion—it aspires to model human cognitive abilities fully, both its successes and its errors. Second, it is constrained to models which are plausible given the actual material substrate of thought: the human brain.

These commitments constrain the models that connectionists will consider reasonable to ones that learn realistically and to ones that could possibly be based in the human brain. These constraints are congenial to educators who are, by definition, concerned that people learn. Educators are not directly
concerned with what might be true of knowledge structures in the abstract or in the possible efficiencies of systems of calculation in computer science.

The split between those committed to modeling the process of learning and those committed to modeling logic and knowledge structures is one of the defining issues in the history of cognitive science. The commitment to actually modeling learning is a commitment to building a system which mimics what we know from observation are characteristics of human learning. It leaves open the question of the mechanism which produces these effects. This is in distinct contrast to the opposing camp's assumption that logic is the mechanism that produces valuable thought. The logicist asserts that anything which is not logical is outside thought as such. Thus emotion and context, for instance, are primarily understood as sources of error.

Crucially, we know that people context dependent reasoners and that a large portion of this context dependence lies in the use of contextually variant categories. Researchers focused on learning treat this as datum to be accounted for in a productive way. That is, they wonder what role context dependence plays in productive thought. Those committed to a logicist position can—within the tradition of objectivity—treat such particularities only as


78 This has been a large part of the discourse of cognitive science, a field broader than the relatively narrow artificial intelligence community on which this dissertation focuses. A textbook overview of much of the work is available in: John R. Anderson, Cognitive Psychology and Its Implications, 2nd edition (New York: W. H. Freeman, 1985), 60-72.

sources of error and would attempt to continually refine their models to eliminate this as a source of error.

People are neither serial computers nor Turing machines. As has been noted in chapter 2, calling attention to this fact points to the particularity, to the finitude, and to the time-bound nature, of the human cognitive machinery. The model of serial computation, the logic machine, cannot be instanciated usefully in the "hardware" of the human brain. Some fundamentally different model must be in operation. That connectionists have found a credible model in light of what we know about the brain allows us to see how incredible the traditional explanation has been—the traditional model is plausible only in the absence of any realistic alternative.80

Connectionist Analogies

Connectionist explanations, precisely because they are not organized on the traditional logical model of declared axioms and deductive conclusions are difficult to grasp using our usual patterns of understanding. It shares this problem with ecological and evolutionary models of explanation. In all these cases small differences in the particular history of any situation can result in very different final outcomes.

One naturally-occurring example of a situation in which small initial differences can make a large difference in the final outcome is erosion on a slope. Water falls randomly on the upper reaches of the slope and rolls downhill picking up minute pieces of soil and creating a small—very small—tendency for succeeding droplets to follow the same path but they mostly

---

80 And, as the history of romanticism amply demonstrates, not always credible even in the absence of a credible alternative.
cancel each other out and the surface wears down smoothly; there is no systematic change. Eventually, through sheer chance, several drops may succeed one another in the same path, carving out a deep enough groove to capture an increasingly large area, and hence capture more raindrops. Soon a small groove has formed that future patterns of random droplets cannot change. In short order it is a ravine that has become a permanent part of the landscape.

Similarly, a connectionist network, before it has learned anything, starts out with no real differences—it is like a smooth plane. As a network receives “drops” of input from the outside small changes in connection strengths propagate across the network which make it slightly more likely that future “drops” will follow the same path. As long as those drops are random, the patterns formed on the plane of the network will also be random—and largely self-canceling. But any patterns in environment, in the input, will predispose the net to form grooves. These grooves are the basis for the connectionist analog to memory—and thought. It is important to realize, however, that this pattern is not itself “a memory.” It only “works” when it is activated, when, as the analogy would have it, activation “flows” through the system. Before it is stimulated by environmental input the pattern is only potential. As a connectionist network learns more and more patterns, the network is increasingly crisscrossed with worn-in “ravines.” Initially each array of different inputs results in a different output array. However, as more and more ravines are worn into the network, we reach a practical limit on the number of independent patterns that may be maintained. In short order new patterns of sensation begin to fall into already established ravines. If the input is sufficiently different it will groove a new ravine. But at some point similar patterns of input will begin to fall into almost identical patterns of activation.
And eventually these small differences in the initial patterns of activation will not be enough to activate different output patterns—the new, slightly different pattern will be captured by an already established "ravine." Lead astray by our traditional epistemology we will be tempted to say that the network has made a mistake. But almost the opposite has happened: the network has *learned* to categorize. When differing input patterns are assimilated to the same output pattern a new organizing principle has surreptitiously emerged. The network has evidenced that most crucial of all learning activities: it has learned to classify difference as the same.

This ability to respond to different inputs as if they were the same leads to such crucial abilities as pattern completion, error tolerance, categorization by family resemblance, and automatic inference. We may consider pattern completion the most basic, and most valuable, of all these qualities.

Classical theories of knowledge, and the theories of categorization on which they are based, are weak and brittle in our messy, work-a-day world because they require perfect knowledge to complete an act of categorization. We must know whether a particular instance evidences the "necessary and sufficient" features that make up the category. The real world seldom obliges us by presenting neat, unconfused and complete information. To handle such a world the agent acting in it must create categories on the fly. The classical story gives us no idea of how categories are created; in order to recognize a member of a category, one must already know the necessary and sufficient features. But until the necessary and sufficient features are themselves recognized there is no way to see the items that are to be categorized as the same. There is no logical way into the classical story of categorization other than the one advocated by innatists from Plato to Chomsky: we, somehow, already know and can remember the ideal forms.
The pattern completion abilities of networks, on the other hand, give us a way to get started in this process without innate knowledge—overloading the network space results in categorizing small differences as the same in a straightforward, material way. Neither mystical nor innatist ideas are necessary. The messy, incomplete, slightly wrong information that we get from our world is easily enough assimilated to the same output pattern. Learning occurs and it is this pattern of relations that underlies our cognitive abilities.

**Connectionism Described**

Connectionism, then, is a model of the relations that underlie cognition. But what is it that we want to explain? The story that we want to be able to tell, with wonder but without turning our attention away from its implications for our practice, is the primal story of the child’s recognition of her mother’s face.

I sit in a room and watch a young child in a crib beneath an open window. She is beautiful, or I think she is. I sit immobilized, fascinated. A cool gust of wind moves the curtains, annoying the child. Her mother, harried, returns and bends over the crib. The baby perceives motion, turns her head and recognizes her mother. She gurgles, smiles, reaches out and is softly gathered in....

The child’s recognition of her mother is a wondrous thing. We have only the exciting beginning scraps of how it might be possible. Recognition is precisely the point at which artificial intelligence, grounded in the modernist story of mind, has failed. Connectionist models offer a radically different story of how this can be accomplished.

The problem that the tradition has found insoluble is just that of recognition. In chapter 2, I discussed a disproof of the possibility in logical terms. In practice the problem appeared less forcefully and more intractably.

Consider the real problem that faces the young child. She must recognize her mother’s face. It is only when we try to specify more fully what this task entails,
a task that AI workers have attempted, that we realize the enormity of the problem. The child never sees the same face twice. She must identify as the same her mother in varying intensities of light, under radically different shadowing conditions, from different distances, from different angles, with and without earrings and glasses. The range of experiences and the degree of overlap with similar experiences that must be disentangled to experience her mother's face as the same is truly staggering. How does she know that her tanned aunt, also brunette and well practiced at bending over babies, is not her mother in a slightly darker light? How is the mask of harsh shadows that occurs in the late afternoon's setting sun disembedded from the diffusely lit face of the morning's dusk. Why isn't that creature with dangling, glistening earrings seen as an unknown and threatening creature? AI programs fail, and fail miserably, at similar tasks. Logic cannot account for it. How does the child succeed?

No artificial neural net has nearly the complexity required to react as the youngest child does to the repeated appearance of her mother's face. But neural net architectures do generate intriguing analogs to the child's competence. The child learns to experience difference as the same. She solves the classical problem of categorization effortlessly, faultlessly, and fluidly.

An Example of a Connectionist Memory Model

An example of a very simple neural net's behavior may give the reader a sense of the striking power of the architecture associated with connectionism. The example developed below focuses on categorization as an emergent property of distributed representation. It is a redevelopment and simplification of the Jets and Sharks exemplar which McClelland & Rumelhart use to
demonstrate the Interactive Activation and Competition (IAC) model of memory.\footnote{This program was published by Rumelhart and McClelland in their handbook: James L. McClelland and David E. Rumelhart, \textit{Explorations in Parallel Distributed Processing} (Cambridge, MA: MIT Press, 1989). This book comes with a disk which contains the IAC program. The program runs on various personal computers and allows the user to go into the defining matrix and alter the parameters, the number of nodes, and how those nodes are connected. The implementation discussed here differs from the original in that the “jets” gang has been eliminated from the matrix, the number of nodes reduced, and the decay parameters eliminated. Rebuilding the example in this way makes the points I make about associative memory easier to explain. For further explanations of this program and examples of varying use, see: Rumelhart and McClelland, \textit{Explorations}, 11-48, and James L. McClelland, David E. Rumelhart, and Geoffrey E. Hinton, “The Appeal of Parallel Distributed Processing,” in \textit{Parallel Distributed Processing, Volume 1: Foundations}, ed. David E. Rumelhart, James L. McClelland, and the PDP Research Group (MIT Press: Cambridge, MA, 1986), 3-44.}

Abstractly, an interactive activation and competition network consists of a collection of processing units organized into different pools. There are excitatory connections among units in different pools and inhibitory connections between units in the same pool. Units in the network are given initial values that change as the program runs through multiple iterations. These values are understood as the activation level of each unit. The change in each iteration is governed by a function which computes the input to each unit by taking into account both the current activation of the unit and any input from other units in the system or from the outside. The function then uses that input figure to change the activation level of the unit by factoring in values that define the maximum and minimum activation levels and a decay factor. All of these values can be changed to explore the effects of changing various parameters on the history and development of the system.

The effect of arranging the relationships in this way can be read from the name of the model: Interactive Activation and Competition. It is interactive because it sets up a relationship between the processing units which makes them into a linked network in which a change in the activation value of one unit effect the value and the linkage strength between all the others. It is competitive
because the relationship between processing units within a pool is negative. A negative connection will mean that the unit within a pool, which in early iterations of the net's history has the highest activation, will suppress the activation value of all others in the same pool. Configuring a network in this way results in one element of each pool coming to dominance. The positive relationships between units in different pools means that the overall level of activation will remain high.

As a model of memory, the properties of an IAC are very suggestive. Like the child, it can categorize—that is, produce "the same" output for different cases. It can recognize an example as a member of a category even under conditions that are adverse, and it automatically fills in missing details which in the experience of the network have co-occurred with other members of this category. All of these features of network relations are difficult to explain formally without elaborate logical structures which, like the Ptolemaic system in astronomy, become only more complicated with each new case. This network model exhibits these characteristics as emergent properties of the system.82

Attaching our relatively technical explanation to a story, a concrete example, may help make these claims more understandable. For the sake of discussion, let us imagine that we are participating in the tryouts for a play based on the situation found in West Side Story. There are two gangs, the Sharks and the Jets; each group has its separate members and they have their separate

82 An emergent property of a system is a property which is born of the interaction of the parts which cannot be extrapolated from the characteristics of the parts taken as individuals. A relatively familiar example would be carrying capacity in ecology. The number of rabbits that a particular ecology will support is not a property of rabbits, or of foxes, but an emergent property of the relations that comprise the system. It can be most usefully conceived as a limit parameter that emerges from the relationships of the system as a whole. For a useful and rigorous discussion of emergence, see: William Bechtel and Robert C. Richardson, "'Emergent' Phenomena in Interconnected Networks," chap. in Discovering Complexity (Princeton, NJ: Princeton University Press, 1993), 202-229.
characteristics. The situation that you, as the casting director (acting the part of the confused inquirer) have set up is to test the person auditioning for the character’s role by putting them in a situation where they have to realistically act and identify a gang member, “Ken,” in response to your questions during a cold improvisation. The actors have come prepared by reading the script and have, supposedly, attended to the instruction to “get to know” the characters. The part that is being auditioned for is that of a young neighborhood man who is in neither gang but knows the people in the Sharks relatively well. You tell the players that the inquirer is trying to get the name of one of the Sharks. How do they handle the scene? The following dialog emerges:

The inquirer asks: “You know that guy, uh, the dark haired guy, always got something to hock, hangs out down at Joe’s.”

The character responds: “Come on, everybody hangs at Joe’s.”

To which the inquirer says: “You know, 30s, different girl every weekend...you know.”

The character reacts: “......Dunno, Rick, maybe.”

Inquirer: “No, I know Rick, he lives on Myrtle.”

Character: “Maybe Ken, but he’s younger.”

Inquirer: “Yeah, Ken, that’s it—he’s younger?”

Character: “Yeah.”

Inquirer: “Yeah.”

As a casting director you are probably pleased; for a cold ad lib, this sounds pretty good. But why are you pleased? What sounds real? Our concern with a realistic modeling of cognitive abilities causes us to focus on the authentic feel of this portrayal of the process of recall. It sounds like normal, everyday associative recall—not like a wooden query about a checklist of traits.
To examine the example a bit, The Jets, like the Sharks, are all Puerto Ricans. They are mostly in their thirties and they all hang at Joe’s. There is little in this first attempt that helps the character find a difference on which to base a judgment—a fact which the character calls attention to in his first line. The inquirer expands, exasperated, to more specific information. The trouble is, he has gotten Ken’s age wrong; Ken is in his twenties. This is an understandable mistake, since most of the Sharks are in their thirties but it confuses character up for a while.\footnote{Actually, the “mistake” that assumes that all Sharks are in their thirties is just the sort of “mistake” that neural nets make. It is intriguing to consider the possibility of modeling two nets feeding their outputs into each other until they produce the same “recognition” but not on the basis of the same pattern. This would be one way to model intersubjectivity.} Finally he comes up with Ken, who fits the bill except for his age—he’s a burglar and he’s divorced, with a major reputation as a ladies’ man. The inquirer affirms the choice by recognizing the name and filling in more details. Finally they engage in a small ritual of agreement.

How would a connectionist approach model this process? What follows is an connectionist model which helps to explain associative memory.

\textit{An Examination of a Connectionist Simulation}

A few caveats are in order as we begin an examination of this model. As models necessarily are, this is an abstraction. In this instance it will be worthwhile to note that the present example is an abstraction on two levels. First, “horizontally” it is an abstraction from its broader context. As McClelland originally developed this example, it was about two gangs and had sixty-eight rather than thirty-seven nodes. This added complexity made it difficult to follow but did enable “bias” effects from this broader context. Second, “vertically” this example implicitly assumes that there are something like “\_Phil” units distinct from name units “Phil,” which in their turn are distinct
from "Shark" units. This is unlikely to be the case. To the contrary, the chief rhetorical purpose of this program is to demonstrate how intertwined and inseparable such concepts are in actual memory and recall.

This model can be understood as a series of concrete examples with their relationships specified which exhibits a surprising ability to both "generalize" and "specify."

The set of relations from which this particular model starts is:

<table>
<thead>
<tr>
<th>Instance</th>
<th>Name</th>
<th>Gang</th>
<th>Age</th>
<th>Hair</th>
<th>Status</th>
<th>Job</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phil</td>
<td>Phil</td>
<td>Sharks</td>
<td>30's</td>
<td>Brunet</td>
<td>Married</td>
<td>Pusher</td>
</tr>
<tr>
<td>Ike</td>
<td>Ike</td>
<td>Sharks</td>
<td>30's</td>
<td>Blond</td>
<td>Single</td>
<td>Bookie</td>
</tr>
<tr>
<td>Nick</td>
<td>Nick</td>
<td>Sharks</td>
<td>30's</td>
<td>Black</td>
<td>Single</td>
<td>Pusher</td>
</tr>
<tr>
<td>Don</td>
<td>Don</td>
<td>Sharks</td>
<td>20's</td>
<td>Brunet</td>
<td>Married</td>
<td>Burglar</td>
</tr>
<tr>
<td>Ned</td>
<td>Ned</td>
<td>Sharks</td>
<td>30's</td>
<td>Brunet</td>
<td>Married</td>
<td>Bookie</td>
</tr>
<tr>
<td>Karl</td>
<td>Karl</td>
<td>Sharks</td>
<td>40's</td>
<td>Black</td>
<td>Married</td>
<td>Bookie</td>
</tr>
<tr>
<td>Ken</td>
<td>Ken</td>
<td>Sharks</td>
<td>20's</td>
<td>Black</td>
<td>Single</td>
<td>Burglar</td>
</tr>
<tr>
<td>Earl</td>
<td>Earl</td>
<td>Sharks</td>
<td>40's</td>
<td>Black</td>
<td>Married</td>
<td>Burglar</td>
</tr>
<tr>
<td>Rick</td>
<td>Rick</td>
<td>Sharks</td>
<td>30's</td>
<td>Black</td>
<td>Divorced</td>
<td>Burglar</td>
</tr>
<tr>
<td>Ol</td>
<td>Ol</td>
<td>Sharks</td>
<td>30's</td>
<td>Brunet</td>
<td>Married</td>
<td>Pusher</td>
</tr>
<tr>
<td>Neal</td>
<td>Neal</td>
<td>Sharks</td>
<td>30's</td>
<td>Black</td>
<td>Single</td>
<td>Bookie</td>
</tr>
<tr>
<td>Dave</td>
<td>Dave</td>
<td>Sharks</td>
<td>30's</td>
<td>Black</td>
<td>Divorced</td>
<td>Pusher</td>
</tr>
</tbody>
</table>

*Figure 3-1: List of Sharks*

Figure 3-2 represents the basic sets of pools and their resting state as the simulation begins. The external input is listed to the left of each unit. The activation level is listed to the right of each unit. It represents the sum of all inputs, external and internal, for the current iteration. The iteration level, which tracks the history of the developing system, is to the far right labeled “cycle.” The initial activity level of each begins at the “resting level,” which is set at
Activation is passed from the feature units (including the name unit) to instance units. From there it propagates back out to the feature units.

As we model the effects of the story on the characters' associations, we will set a few ground rules which will simplify this model. All input activations will be either a positive or a negative 100 and there will be no decay in the input values on successive cycles. We will be watching the instance units for their pattern of activation as successive external inputs are entered. Each changed input will be denoted by italics. We will assume that no unit is available to "consciousness" until its activation level goes above 50. When a unit reaches this level, it will be depicted in boldface. At that point the character guesses that the individual represented by that unit is the one that the inquirer is seeking.

<table>
<thead>
<tr>
<th>External input</th>
<th>Name units</th>
<th>Instance Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Sharks -10</td>
<td>0 Phil -10</td>
<td>0 _Phil-10 cycle 0</td>
</tr>
<tr>
<td>0 Ike -10</td>
<td>0 Nick -10</td>
<td>0 _Nick-10</td>
</tr>
<tr>
<td>0 Don -10</td>
<td>0 Ned -10</td>
<td>0 _Ned-10</td>
</tr>
<tr>
<td>0 Karl -10</td>
<td>0 Ken -10</td>
<td>0 _Ken-10</td>
</tr>
<tr>
<td>0 Black -10</td>
<td>0 Earl -10</td>
<td>0 _Earl-10</td>
</tr>
<tr>
<td>0 Brunet-10</td>
<td>0 Rick -10</td>
<td>0 _Rick-10</td>
</tr>
<tr>
<td>0 Ol -10</td>
<td>0 Neal -10</td>
<td>0 _Neal-10</td>
</tr>
<tr>
<td>0 Neal -10</td>
<td>0 Dave -10</td>
<td>0 _Dave -10</td>
</tr>
<tr>
<td>0 Pusher -10</td>
<td>0 Burglar -10</td>
<td></td>
</tr>
<tr>
<td>0 Bookie -10</td>
<td>0 Divorce -10</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3-2: Sharks Interactive Activation and Competition (IAC) network in resting state

---

84 This is a truncated explanation of the processes modeled in the McClelland and Rumelhart program. For a discussion of resting levels, decay rates, gamma levels, and other interesting but here secondary phenomena, see: Rumelhart and McClelland, Explorations, 11-48.

85 Thus earlier "clues" are no less salient than earlier ones—a psychologically unrealistic assumption.
As our improvisation starts, the players already know that the topic of conversation is a Shark. Thus they are acting the part of someone who is already biased toward recalling that which is typical for Sharks. In figure 3-2 we see the result of setting the external input of the Sharks unit to 100, or full activation. As activation propagates through the network for 10 cycles, interesting effects begin to emerge. Before hearing anything about the particular Shark that the inquirer was interested in, the network is predisposed to find Nick, Dave, Neal, Rick or Dave the likely choices. The net would find it easier to “believe in” a thirty-year-old, black-haired, married man. These are the initial biases of the system, biases born of the architecture’s interaction with these specific instances.

<table>
<thead>
<tr>
<th>External input</th>
<th>Name units</th>
<th>Instance Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharks 80</td>
<td>Phil -7</td>
<td>_Phil 6</td>
</tr>
<tr>
<td></td>
<td>Ike -7</td>
<td>_Ike 6</td>
</tr>
<tr>
<td>0 in20s -5</td>
<td>Nick -7</td>
<td>_Nick 7</td>
</tr>
<tr>
<td>0 in30s 11</td>
<td>Don -7</td>
<td>_Don 5</td>
</tr>
<tr>
<td>0 in40s -5</td>
<td>Ned -7</td>
<td>_Ned 6</td>
</tr>
<tr>
<td></td>
<td>Karl -7</td>
<td>_Karl 6</td>
</tr>
<tr>
<td>0 Blond -8</td>
<td>Ken -7</td>
<td>_Ken 5</td>
</tr>
<tr>
<td>0 Black 8</td>
<td>Earl -7</td>
<td>_Earl 6</td>
</tr>
<tr>
<td>0 Brunet 0</td>
<td>Rick -7</td>
<td>_Rick 7</td>
</tr>
<tr>
<td></td>
<td>Ol -7</td>
<td>_Ol 6</td>
</tr>
<tr>
<td>0 Single 0</td>
<td>Neal -7</td>
<td>_Neal 7</td>
</tr>
<tr>
<td>0 Married 6</td>
<td>Dave -7</td>
<td>_Dave 7</td>
</tr>
<tr>
<td>0 Divorce -4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pusher 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burglar 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bookie 1</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 3-3: Sharks IAC network at 10 iterations*

The initial query asks for a man among the Sharks who is dark-haired and a burglar by trade. We simulate this by setting the value of the external input to the appropriate units to 100. This new configuration, added to the activation of the Sharks unit which remains active, propagates through the network. Among
the Sharks, only Ike, the lone blond, is definitively eliminated, and his activation value drops to a negative number (though not quite to the resting state of -10). The identification of the sought-after Shark as a burglar increases the activation level of all the burglars: Don, Ken, Earl, and Rick—but not, yet, to the threshold rate of 50. (See figure 3-3)

<table>
<thead>
<tr>
<th>External input</th>
<th>Name units</th>
<th>Instance Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 Sharks 84</td>
<td>0 Phil -3</td>
<td>0 _Phil 16</td>
</tr>
<tr>
<td></td>
<td>0 Ike -8</td>
<td>0 _Ike -9</td>
</tr>
<tr>
<td>0 in20s 2</td>
<td>0 Nick -4</td>
<td>0 _Nick 7</td>
</tr>
<tr>
<td>0 in30s 42</td>
<td>0 Don 4</td>
<td>0 _Don 37</td>
</tr>
<tr>
<td>0 in40s -6</td>
<td>0 Ned -3</td>
<td>0 _Ned 16</td>
</tr>
<tr>
<td></td>
<td>0 Karl -6</td>
<td>0 _Karl 0</td>
</tr>
<tr>
<td>0 Blond -15</td>
<td>0 Ken 2</td>
<td>0 _Ken 32</td>
</tr>
<tr>
<td>100 Black 80</td>
<td>0 Earl 5</td>
<td>0 _Earl 40</td>
</tr>
<tr>
<td>100 Brunet 77</td>
<td>0 Rick 7</td>
<td>0 _Rick 43</td>
</tr>
<tr>
<td></td>
<td>0 Ol -3</td>
<td>0 _Ol 16</td>
</tr>
<tr>
<td>0 Single 7</td>
<td>0 Neal -4</td>
<td>0 _Neal 7</td>
</tr>
<tr>
<td>0 Married 39</td>
<td>0 Dave -4</td>
<td>0 _Dave 3</td>
</tr>
<tr>
<td></td>
<td>0 Divorce -2</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3-3: Sharks IAC network at 20 iterations

In the next exchange the inquirer says that the Shark he is seeking is, like most Sharks, in his thirties and that he is an active ladies’ man. This guidance is modeled here as turning on the inputs to the in30s, Single, and Divorced units. Another 10 cycles of iteration yields the following pattern:
At this point Rick has gone over the threshold activation level of 50 and is therefore reported as a guess. Rick, of all the Sharks, is the only one who matches all the criteria. He is a thirty-year-old, black-haired, divorced burglar. This is a good example of content-addressable memory.

The trouble is that this guess is wrong. The inquirer has mistakenly said that the person that he is seeking is thirty. But the inquirer is certain that it is not Rick; after all, he knows Rick. That means that no one really matches the search criteria. In contrast to the current IAC model systems based on simple deduction break down at this point. They cannot handle “bad” input. The computer maxim, “Garbage in, garbage out” applies, but with the added caution that it takes only one “byte” of garbage to turn all the output into trash. Humans don’t just quit in this way, though they may be confused.

With the IAC network modeled here, we simply continue to do what we have been doing: we input the datum that Rick is not the name of the instance
sought. We set that unit to a negative 100 and again cycle the networks relations another 10 times.

<table>
<thead>
<tr>
<th>External input</th>
<th>Name units</th>
<th>Instance Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 Sharks</td>
<td>0 Phil</td>
<td>0 _Phil</td>
</tr>
<tr>
<td></td>
<td>0 Ike</td>
<td>0 _Ike</td>
</tr>
<tr>
<td>0 in20s</td>
<td>0 Nick</td>
<td>0 _Nick</td>
</tr>
<tr>
<td>100 in30s</td>
<td>0 Don</td>
<td>0 _Don</td>
</tr>
<tr>
<td>0 in40s</td>
<td>0 Ned</td>
<td>0 _Ned</td>
</tr>
<tr>
<td>0 Blond</td>
<td>0 Ken</td>
<td>0 _Ken</td>
</tr>
<tr>
<td>100 Black</td>
<td>0 Earl</td>
<td>0 _Earl</td>
</tr>
<tr>
<td>100 Brunet</td>
<td>-100 Rick</td>
<td>0 _Rick</td>
</tr>
<tr>
<td>100 Single</td>
<td>0 Neal</td>
<td>0 _Neal</td>
</tr>
<tr>
<td>0 Married-13</td>
<td>0 Dave</td>
<td>0 _Dave</td>
</tr>
<tr>
<td>100 Divorce</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 Pusher</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 Burglar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 Bookie</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 3-5: Sharks IAC network at 40 iterations*

This pushes the Ken instance barely above 50 and allows our character to offer Ken as a possible candidate. This choice is confirmed by the character and we finally have identified our Shark. But before we go on, note that even when receiving negative input to his name unit, Rick has actually picked up 2 points, to 65. What is going on here? This, like the ability to retrieve a particular instance by its content, is an artifact of the interaction of the architecture and the particular instance we have explored. Recall that every member of a “pool” is connected to every other member by inhibitory connections. The instance _Rick has gained enough dominance locally that it continues to suppress others in its pool independently. Had this same pattern of inputs into the net been input initially, if the inquirer had made a simple list of all that he knew, this effect would not have appeared and Ken would have been the dominant unit—and the clear answer. But in this sort of network, history does count. The net is in
very real danger of “blocking on” Rick. McClelland and Rumelhart call this phenomena “hysteresis:” “... prior states of the network tend to put them into states that can delay or even block the effects of new inputs.”

Luckily for the sake of this exercise, Ken emerges from the background enough to become an active possibility. The character offers this as an alternative with the caveat that Ken is too young. When the inquirer agrees that Ken is the one and that he is younger, we activate the in20s unit and deactivate the in30s unit and cycle the program once again.

<table>
<thead>
<tr>
<th>External input</th>
<th>Name units</th>
<th>Instance Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 Sharks</td>
<td>85</td>
<td>0 Phil -12</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>_Phil -11</td>
</tr>
<tr>
<td>100 in20s</td>
<td>78</td>
<td>0 Nick 5</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>_Nick 42</td>
</tr>
<tr>
<td>0 in30s</td>
<td>53</td>
<td>0 Don -10</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>_Don 21</td>
</tr>
<tr>
<td>0 in40s</td>
<td>-15</td>
<td>0 Ned -12</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>_Ned -11</td>
</tr>
<tr>
<td>0 Blond</td>
<td>-16</td>
<td>0 Ken 22</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>_Ken 64</td>
</tr>
<tr>
<td>100 Black</td>
<td>83</td>
<td>0 Earl -9</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>_Earl -5</td>
</tr>
<tr>
<td>100 Brunet</td>
<td>74</td>
<td>100 Rick -17</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>_Rick 60</td>
</tr>
<tr>
<td>100 Single</td>
<td>80</td>
<td>0 Neal 5</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>_Neal 42</td>
</tr>
<tr>
<td>0 Married</td>
<td>-15</td>
<td>0 Dave 3</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>_Dave 41</td>
</tr>
<tr>
<td>100 Divorce</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>0 Pusher</td>
<td>-3</td>
<td></td>
</tr>
<tr>
<td>100 Burglar</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>0 Bookie</td>
<td>-12</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 3-6: Sharks IAC network at 50 iterations*

With the confusion about the age cleared up, Ken becomes the dominant node in the instance pool and the clear choice.

What is exciting about this style of explanation is that it is a robust model of memory that produces the sort of associative recall that seems intuitively plausible and that it does so without recourse to elaborate, implausible rules.

---

While, like all computer programs, this one is built from rules, it is important to note just what is modeled in this medium. Unlike conventional logicist models it models a pattern of relationship—and a pattern of relationship that is particular to the actual content. The “structure” is dependent on the content. In a traditional account, the structure is independent of the particular content. The forms are eternal while the content is ephemeral. While connectionist models blur this distinction, they introduce a new distinction: architecture. This term, adopted from computer science, denotes a set of constraining parameters within which particular instances are constructed. Architectures limit but do not determine the structure/content interrelationship that is at the heart of connectionist models.

The point is not that this particular example could not be solved through logic (fuzzy logic is the most obvious logical strategy), but that to produce the same output, a logical model would have to be extremely complex. This is due to its inability to handle dynamics. Change occurs over time and models built on the model of deductive logics are timeless. To produce the same effects, a logicist model built of “if . . . then . . .” statements would have to chart out every possible path that the input could take and specify a set of relationships that would hold under each condition. Essentially, it would have to know all the branch paths in advance and specify a relationship in each case. In a strictly

---

87 And, of course, but less germane to the point made here, the content would not be available as content without the patterned relationship that I have here called structure.


89 Though this is the topic for a different dissertation, fuzzy logics also constitute a fundamental departure from classical reasoning in that they, like connectionist models, proceed from a point that assumes that objects are not unitary. In common use, this is yet another instance where statistics are used to dodge the effects of history on the systems that logicists want to study so that the relationships could be idealized as time-free.
logical system, all possible outcomes are calculable and propositional models in artificial intelligence build on this possibility—they calculate the intervening steps on the fly. But even for so simple a situation as the Sharks simulation given above, such a strategy is extremely unwieldy. Classical assumptions have resulted in an increasingly complex model of nested, conditional rules for any situation which it models. This resembles nothing so much as the Ptolemaic system of astronomy in which the assumption of perfect circles, made on the grounds of a particular culture’s belief that circles were the perfect forms behind the real, elaborated epicycles upon epicycles to explain each new observation. In the current instance we are faced with a situation where the classical assumption, based on our particular culture’s assumption that deductive logic is the pattern that underlies both the world and the mental, is producing an increasingly complex system of self-similar rules which shows no sign of coming to any natural end. We are ripe for a Copernican revolution in the way that we think about thought. Current models of cognition are less and less plausible as we are forced to elaborate them in the face of empirical findings in cognitive science and psychology.

The Plausibility of Connectionist Models

Connectionist modeling is a powerful tool that is coming into use in a surprisingly large number of areas. The differing interests of various

---

90 This example is one made popular by Kuhn in a different context but brought most forcefully into this discussion by Smolensky. See: Jeremy Campbell, The Improbable Machine (New York: Simon and Schuster, 1989), 191.

91 We are also ripe, not incidentally, for a new Copernican revolution in how we see the world. See, for instance, Ilya Prigogine and Isabelle Stengers, Order Out of Chaos: Man’s New Dialogue With Nature (New York: Bantam, 1984).

92 This aspect is accessibly covered by Waldrop (see particularly Farmer’s remarks) in: M. Mitchell Waldrop, Complexity: The Emerging Science at the Edge of Order and Chaos (New York:
investigators leads them to emphasize differing values in the tool. Some are interested in it largely because of its ability to convincingly model psychological and cognitive phenomena that had remained puzzling under the dominant models of investigation. Others are chiefly interested in the neural plausibility of such models and their ability to connect disconcerting discoveries about neural architecture to basic perceptual and motor abilities shared by all animals. Still others are fascinated with the computational properties of network models. Central processor, serial, local address computers—the sort of computation most supercomputers and your desktop PC share—have run up against distressing performance bottlenecks which networked architectures do not share. While one of the exciting things about this field is that the cross-fertilization between disciplinary areas is very evident, most of the motivation for pursuing network models falls into one of these areas.

Our interests in pursuing this field are both narrower and broader. They are narrower in the sense that this work focuses chiefly on learning and the process by which learning occurs. While this is an important focus of some researchers


93 This, for example, is the main interest of the PDP group. See: McClelland, Rumelhart, and Hinton, "The Appeal of Parallel Distributed Processing," 3-44.


in the field, it is not generally central to their rationale for pursuing work in the area. For an educator, a focus on learning and the development of a credible theory of the material change that constitutes learning are the central motivations. Education's interest is broader because education is directly engaged in a practical project in the social world, a project which is peripheral to the interests cited above; education seeks to enhance our students' practical competence.96 As we will see below (and will cover in more detail in chapter 4) this latter interest will push us to understand that connectionism and the insight it provides into learning is not fully adequate to fulfill our distinctly educational interests.

In this section we will examine the plausibility of connectionist models as cognitive, neural and computational models of cognition with special attention to what they have to say about learning.

The Cognitive Plausibility of Connectionist Models

One of the most appealing aspects of connectionist models is their psychological plausibility. Connectionists models generally produce a pattern of success and, perhaps more crucially, of failure that is analogous to the pattern of success and failure that we find in human cognition.

Education and the Classical Cognitivist Account

Educational interests, as outlined above, help us narrow our focus to the problems in psychological approaches that are most important for education.

---

96 This is to be sharply distinguished from so-called “competency-based” education which is not about practical competence at all, but is about a schoolish testing regime which is antithetical to the worldly competence which is education’s legitimate aim. For an excellent exposition of this distinction, see: William B. Stanley and James A. Whitson, “Citizenship as Practical Competence: A Response to the New Reform Movement in Social Education,” The International Journal of Social Education 7 (Fall, 1993): 57-66.
One range of problems in psychology that is suggestive for the practice of education is that of competence. This range of issues has to do with understanding how agents can act competently in the world. As we ignore the successful learning of the baby's coming to perceive her mother's face, a real engagement with the positive issue of competence is often ignored for a focus on various types of failures of competence. But generally people, like babies, do act competently given a reasonable history in the appropriate context. This competence is therefore held to be "unremarkable." But again, like the baby's ability to recognize her mother's face, it is a remarkable and wondrous thing.

Consider for a moment what the standard story about human competence has to say about the conditions for competent action. As we have seen above (in chapter 2), there is a very straightforward way of understanding our abilities which says, essentially, that we know the world through knowing the objects of the world and the relations which connect them. Objects are understood as either pregiven simples or categories in the classical sense. Classical categories are known by their necessary and sufficient features and are sharply distinguished from members of other categories. This sort of category is necessary in any system which relies on formal logical relations to connect the objects of understanding since, as Aristotle observed, objects which are allowed to overlap cannot be handled through the offices of logic.

Given this presumption one would expect that any study of human competence made from within this tradition would look first to cataloging the objects of the world and the relations between them. Indeed, this has been just the way that Western philosophy has preceded. As Liebniz observed:

[T]he most important observations and turns of skill in all sorts of trades and professions are as yet unwritten. Of course, we can also write up
this practice, since it is, at bottom, just another theory more complex and particular.97

But, "of course," Liebniz's task is too large: how can we list all the things that a person must be able to distinguish, know the use of, and their relations to objects both absent and present, just to make it through the everyday task of shopping at a modern grocery store? As the task of listing objects and their relations becomes more and more obviously a practical impossibility, those attempting this path are led to wonder how it is that people, who in the analytico-referential account must possess just such an annotated list, accomplish so difficult a task. They finally either conclude that people come to the world already filled with the appropriate knowledge98 or they postulate that there exist structures in the human perceptual apparatus and knowledge storage structures that "naturally" pick out the objects of the world.99

Such synthetic a priori are, ironically, justified as logical necessities. How else can the burden be borne? Logically, say their advocates, these are the only solutions. But the force behind the confident deferral to logic is derived directly from the presumptive unity and naturalness of the analytico-referential discourse. This discourse is founded on the presupposition of an identity between language, the world, and logic. Further, the assumption that this must be the way that the mind works is due to our common acceptance of the

97 Leibniz, Selections (New York: Scribner, 1951), 48. quoted in Dreyfus and Dreyfus "Making a Mind Versus Modeling the Brain."

98 This tack is used most commonly in restricted domains; for example, Chomsky's certainty about the black box of language is derived from the apparent complexity of grammar and the poverty of stimulus from which to learn such complexity. Plato, however, in his doctrine of the eternal forms (that we remember imperfectly), made it a general strategy.

99 Frege's Sinn, Husserl's noema, and Kant's schemata all share elements of this; see Dreyfus's historically oriented discussion in: Dreyfus and Dreyfus, "Making a Mind versus Modeling the Brain," 15-43.
Cartesian introjection of logic into the workings of the mind. Without our particular intellectual history, this approach would be neither obvious nor natural.

Given this history, however, it has been taken as both. Most recently we see the development of formal memory structures, which act as the sort of prescient filters that Frege, Husserl, and Kant postulated, in cognitive science. These have taken the form of Minsky’s frames, 100 Shank’s scripts101 and Rumelhart’s schemata. 102 Each of these has been put forward as a way to deal with context—a context which had stymied earlier more purely logical attempts to build models of intelligent agents in artificial intelligence and cognitive science. 103 These earlier attempts were based on the simple understanding of thought as taking the form of logic. While enjoying success in limited realms—such as solving logical theorems 104—they functioned poorly in the complexity of open


104 I am thinking here of the area of expert systems and, especially, Newell and Simon’s early program “Logic Theorist.” Interestingly, considering the intellectual history presented in chapter 2, this program was developed for and had its first use in proving theorems from Russell and Whitehead’s Principia Mathematica. See Gardner’s discussion of this period under the heading “The programs of the Dartmouth Tetrad” in Gardner, The Mind’s New Science, 145-155.
situations. Workers in this programme,\textsuperscript{105} echoing their historical antecedents, decided that the problem was a matter of inadequate or inaccessible background knowledge of the world. So background knowledge, understood as facts and propositions, were added to their models in ways that were structured to increase their availability and to capture context. Importantly, but not surprisingly given the history of the term, these schemata were conceived as hierarchically structured sets of propositions concerning things in the world—as structures of logic connecting fact. John Anderson, in his influential text,\textsuperscript{106} gives the example of a schema for a house which contains “slots” for categories such as “superset,” “function,” and “location.” A given instance may fill all of the slots in terms of its “default” designation and hence be a “central” instance, with less ideal instances being “peripheral.” What made this revival possible in the late seventies was, as Rumelhart et al. (1986)\textsuperscript{107} remarked, the development of the computer as a research tool. This made possible a specificity which allowed schema theorists to state their theories with a rigor whose lack had led to the dismissal of previous models.

These models, lent power by their material presence—they actually “ran”—became the basic building blocks of an understanding of knowledge by the new cognitive school of psychology. They did not, however, fulfill all their theorists’ dreams. In their formal instanciation, they were unable to answer some of the questions that researchers had been asking; they were “a pale representation of

\textsuperscript{105} The spelling “programme” is intended to reference Lakatos’s understanding of research communities. Lakatos emphasized competing research programmes, an analytical framework which seems particularly apt in the context of this discussion. See: Imre Lakatos, \textit{The Methodology of Scientific Research Programmes} (London: Cambridge University Press, 1978).


the underlying intuitions."\textsuperscript{108} Among others, they were fixed by their original structure, in that empty slots were filled but the structure itself could not be elaborated; and there was no way to place variable constraints on the filling of default functions, which would have made possible a situation where the filling of one slot would have changed the default on the filling of another.

In the vocabulary of education, this failure meant that these formalized schemas could "learn" only in the most mechanical of ways: it could fill slots with the nearest match to an expected value. A more profound alteration of structure in response to a history of changing input or a more dynamic contextual refrigeration was impossible. They were essentially memory and retrieval structures and they answered researchers' questions concerning anomalies in this area, such as the relative ease with which associated ideas are memorized and certain kinds of contextual blockage and enablement of recall. The difficulties with regard to learning were not breached and researcher dissatisfaction continued.

But the computer power that enabled the return of schema theoretic perspectives to psychology also made possible a return of the associative models which had suffered a similar eclipse in the absence of a convincing way to instanciate their complex networked ideas.\textsuperscript{109}

The Sharks simulation is an example of how plausible models of cognitive functions can be built using connectionist architectures. The Sharks simulation


\textsuperscript{109} In fact, the intuition which informs connectionist theories goes all the way back to William James who, like other pragmatists, struggled to make his theoretical position a material one. The image in his \textit{Briefer Course} of the connections that make up associations and the accompanying discussion of how they change during experience is eerily prescient of more recent work. See: William James, "Chapter 7: Association," chap. in \textit{Psychology: The Briefer Course} (New York: Harper and Brothers, 1892/1961), 120-146.
is designed to deal with the problems of static structures that had characterized frames, scripts, and schemata.

Generally, the problems with static logical structures such as those discussed above are subsumed under the rubric of "brittleness." Brittle programs work adequately in the restricted domain in which they are developed but do not scale well when exposed to a large and more complex universe. They deal poorly with several sets of conditions: 1) situations in which the complexity of the problem space makes it impossible to anticipate (and program for) every particular possible state of relations between interacting units; 2) situations in which the input contains errors; 3) situations in which history or order makes a difference; and 4) situations in which the problem itself is poorly defined. The Sharks simulation deals with the first three of these problems in ways that indicate the value of connectionist models of memory. Problem four, the problem of "poorly defined problems," is deeper and requires a broader solution, to be discussed below, than the Sharks example can support.

The Basis for Emergent Cognitive Properties in Network Architectures

The Sharks simulation exemplifies a primitive operation of connectionist architectures: pattern completion. Pattern completion is an emergent property of connectionist networks which is based on the tendency of a network to settle into similar patterns when given similar inputs. Put so baldly, such a comment seems almost trivial, but this is not a characteristic of competing models of cognition where only identical inputs, or inputs which can be transformed to be identical, reproduce the same outputs. This relative stability vis-a-vis incomplete or distorted input is crucial to understanding learning. The child

110 Several of the phrases used in this sentence are technical uses of common terms. While the common meaning indicates adequately for the present purposes the intended meaning, the reader should be aware that "problem space," "unique decision," and "poorly defined problem" have computationally precise meanings.
who learns to recognize her mother's face has never seen the same face twice. Logicist systems theoretically require a massive amount of prescient discarding of sensory experience to achieve identical input such that recognition can occur. Under such a logicist interpretation, the child would somehow know, before she knew the category being learned, just what she must eliminate as a non-essential feature of that category. Connectionist models do not have this problem. Pattern completion's simplicity turns out to be deceptive. It is actually a robust primitive under conditions where we have limited resources of either time or storage space.

Distributed representation is the architectural feature of connectionist networks that supports pattern completion. In a connectionist model there is no single location that serves as the site of any particular memory. Instead, memory is distributed across many nodes. As we have seen in the figures above (1-6) these networks are designed to implement incremental changes to connection strengths between nodes based on the history of external input. Because of this there is no single node or location within the network that uniquely encodes the characteristics of, say, Ken. Similarly, though it was never explicitly entered, the network does contain an "idea" of the ideal Shark. Figure 3-2 above shows the initial configuration where a bias toward Neal, Dave, Rick and Nick and against Ken and Don are established after only 10 iterations. This is in sharp contrast to localist models of memory developed in analogy to von Neuman computer architectures, where each atom of information is stored in a defined location.

Complexity and the Plausibility of Connectionist Architectures

Distributed representation enables a network to evade the first problem cited above, of situations too complex to anticipate all possible relationships, by its ability to respond "reasonably." In figure 3-3 above, the simulation is asked to suggest a dark-haired burglar who is otherwise a typical Shark. It returns Rick, who is a dark-haired burglar, but suggests him over other dark-haired burglars because his pattern of activation is closer to the "typical" dark-haired Shark. Rick's other characteristics are that he is in his thirties and is divorced. His closest competitor, Earl, is in his forties and is married. A typical Shark is in his thirties and is married. Earl is one of two Sharks aged forty and Rick is one of two Sharks who is divorced. So how does the program "know" which one to choose? In this case the simplest explanation is that the age thirties is more typical of Sharks than is the status married and that a reasonable answer to the query would return the most "Sharklike" qualifying member: Rick. It is instructive to trace this out in some detail. There are eight Sharks in their thirties and only two each in their twenties and forties. But there are only six married Sharks, four who are single, and two who are divorced. The negative relationship that characterizes within-pool connections means that as the iterations proceed, the eight thirties in the age pool will come to dominate their pool more quickly and thoroughly than the six married instances will dominate the marital pool. The age pool's activation strength will propagate to the "Rick" unit earlier and with greater strength than the marital status pool's activation will spread to the "Earl" unit. Once this early dominance is established, Rick will suppress Earl within their common pool more than Earl will suppresses Rick. The architecture is thus sensitive to and magnifies small initial
differences. Ultimately you get the most "reasonable" answer to a question which no designer of the simulation ever expected would be asked.

Nowhere in this program is there a rule which relates to this instance nor is there an ideal, prototypical model to which the actual cases are compared. This is worth noting because descriptions of psychological reality often insist on one or both of these conditions. The system demonstrated above can be described as having a rule which says something on the order of, "Take all the instances which satisfy this list of characteristics; if this list is larger than one, then take a look at all the other characteristics that all the members of the new group share which are not already listed; if there are any differences, count the number of times each different element appears in the larger population; if this number is the same, take the categories in which the difference appears and determine if it appears more often than any other subcategory; If it does, count the absolute number of times it appears as a member of its category . . . then award the nod to the instance with the largest number of times to appear." This is a complicated and unreadable sentence even with the many alternate branches not explicitly stated and elaborated. It is a bare-bones rendition of the logical, propositional approach to this problem. No such rule appears in the Sharks program, and no large number of other rules which could be used to describe other possible outcomes of this same configuration actually exists.

112 The phrase "small initial differences" references a key concept in what has come to be known as chaos theory and nonlinear dynamics. For a good popular treatment, see: James Gleick, *Chaos: Making a New Science* (New York: Viking, 1987). For a more philosophically oriented approach, see: Prigogine and Stengers, *Order Out of Chaos*.

113 The rule-based approach has been the dominant methodology in artificial intelligence and remains extremely influential. The descendants of Liebniz and Husserl remain active in their attempt to exhaustively describe the world in a way that would allow answers to questions to be calculated. See: Jim Barnett, Kevin Kight, Inderjeet Mani, and Elaine Rich, "Knowledge and Natural Language Processing," *Communications of the ACM* 33 (August, 1990): 30-49.
Very similarly, no prototype exists; that is, no “perfect dark-haired, burglar Shark” image exists which is somehow a smooth interpolation of all the Sharks with these characteristics plus an average of all known values in the larger category of Sharks for each unknown value.\textsuperscript{114} Though the program may be described as “acting” as if this is true, there is nothing of this sort to locate.

**Error and the Plausibility of Connectionist Architectures**

While both logical and prototypical approaches can be used to describe the functioning of this program, neither of these alternate approaches can adequately account for error. This is the gist of the second problem cited above: situations where the input is “noisy.” Noisy information is the bane of logicist models of intelligent action. By “noisy,” cognitive scientists usually mean information which is, in the most extreme case, simply wrong, is in excess, or is internally inconsistent. Logic has always had trouble with wrong information. You must be certain of your premises if your deductions are to be trusted. Excess information, information which is more than is necessary and sufficient to derive the correct conclusion, introduces biases into logical systems which can only produce deviation from the perfect answer. What is wanted is all the information and only such information. Internally inconsistent information is also deadly to such systems. To the logicist this means that somewhere there is something wrong—if the world is held to be fundamentally logical, as I have argued above, then true inconsistency is not really possible. But the world is such that often cognitive researchers may know that there is an inconsistency in the data set, but are unable to locate the particular “error” that causes it.

Humans have much less of a problem with such data. People can usually pick out easily any anomalous data in an area with which they are familiar—it

\textsuperscript{114} For a good discussion of the current status of prototype theory, see: Lakoff, *Women, Fire, and Dangerous Things*, 39-57.
doesn't fit the pattern with which they are familiar. In practice we simply discard such data; more often than not this turns out to be a good idea. Similarly, people often consider data which is in excess of what is logically required to come to the correct conclusion. While human behavior which shows these characteristics is often criticized as “illogical,” any cognitive theory which purports to be psychologically realistic must produce such effects as effortlessly as people seem to do.

In the original Sharks example, we observed the way in which the Shark network was used to model a discussion in which one character asked another to identify a person unknown to the inquirer but known to the first character. Some of the drama of the example was due to tension as to whether the program would be able to recover from incorrect information. In some ways the example given was a particularly powerful one because the target Shark, Ken, was not a typical Shark. Ken was one of only two Sharks who were in their twenties and also one of only two divorced Sharks. Because of this the network was biased against finding Ken a suitable Shark of any kind. The network displayed some of the flexibility and potential power of such architectures in being able to overcome error in such an unfavorable context. A more common, though less dramatically interesting situation (the casting director would find it “uninteresting”) would have been for the interlocutor’s admitted ignorance of the Sharks and his familiarity with other gangs to lead him to assume that the Sharks must all be younger than they were. He would be fitting them into his previous “biases.” This would lead to his saying that he wanted a twenty-year-old Shark when he was actually referring to thirty-year-old Rick. In this instance the same network would return three candidates at above the

threshold of 50 units of activation: Don, Ken, and Rick. In returning the instances Don and Rick, the simulation would have effectively been throwing away anomalous information which could lead to poor conclusions. It may not be strictly logical, but it does work. The nuanced nature of this discarding should be noted, however; Ken, who fits all of the given parameters, is the first choice but he, the odd Shark, is not the only choice.

History and the Plausibility of Connectionist Architectures

History effects are included as a type of error discussed above in the tradition of cognitive science. Connectionist models, however, give us a reason to focus on history as an element not directly associated with error. In a connectionist account history is not simply a source of deviation from the timeless, general, logical account. The passage of time can, instead, be seen as a source of much of the power of connectionist explanations.

A connectionist architecture blurs the distinction between content and structure. The effects of this architecture in making possible a reasonable instanciation of associative memory through the concept of distributed representation have been noted immediately above. Such models reproduce a much fuller range of human psychological phenomena. Pointedly, such models also produce, as byproducts, an account of human “error” that is integrated with the account of human ability.

Considered over time the coupling of content and structure modeled by a connectionist network opens a dynamism that simple structural models cannot model. Put simply, in a connectionist network both the order of events taken as data and duration matter. A network will settle into very different patterns depending upon the order in which it gains information and upon the length of time it processes that information.
Again, we can use the Sharks to examine a concrete example of what this might mean. Recall the original story that we told:

The inquirer asks: “You know that guy, uh, the dark haired guy, always got something to hock, hangs out down at Joe’s.”

The character responds: “Come on, everybody hangs at Joe’s.”

To which the inquirer says: “You know, 30s, different girl every weekend . . . you know.”

The character reacts: “… Dunno, Rick, maybe.”

Inquirer: “No, I know Rick, he lives on Myrtle.”

Character: “Maybe Ken, but he’s younger.”

Consider this slightly different story. Here, the order of the information presented in the first two inquiries is exchanged.

The inquirer asks: “You know that guy, uh, thirties, different girl every weekend, hangs out down at Joe’s.”

The character responds: “Come on, everybody hangs at Joe’s.”

To which the inquirer says: “You know, the dark haired guy, always got something to hock . . . you know.”

The character reacts: “… Dunno, Rick, maybe.”

Inquirer: “No, I know Rick, he lives on Myrtle.”

The Sharks network I have devised is sensitive to this difference in a way that turns out to be catastrophic for the inquirer’s finding the person he is seeking. In the original example the network had developed to the point that both Rick and Ken were activated at above-threshold levels and with the knowledge that the person sought was not Rick, the second choice, Ken, could be confirmed (figure 3-7).
Modeling the second story, however, makes for a radically different situation after we have gone through 40 cycles. Rick, as in the first simulation, was the first instance to make to the threshold mark of 50 immediately after 30 cycles. But at 40 cycles, after the inquirer claims that the person he is seeking is not Rick, the following, quite different, pattern emerges. Instead of Ken emerging as the second choice, Nick, Neal, and Dave emerge. Given the same information but in a different order, this simulation has failed to find the sought-after name: Ken. Further cycling of this simulation in this condition—thinking “longer” about it—without any new information in terms of different input strengths will only deepen the bias toward Rick as his stronger activation level within the instances pool leads to the suppression of all other instances. Under these new conditions Ken will never emerge as an alternative.
This model fails to return the “correct” answer in this instance. And this failure is not a fluke—it is a result of the basic architecture of the simulation, and to one degree or another it is a problem with all connectionist architectures where each node is active in more than one representation. McClelland and Rumelhart call this phenomena “hysteresis” and it is one way to explain the common phenomena of blocking, where an incorrect recall seems to block the correct recall. One might not be able to recall a friend’s daughter’s name as Sally after recalling, incorrectly, “Sarah.” Such historically, contextually

---

116 It is possible to design neural net architectures in which each pattern which corresponds to a representation of a pattern of input is fully independent of other patterns learned. This is chiefly a matter of the relative complexity of the learning tasks versus the complexity of the network. This is seldom a useful strategy because it forfeits the interactivity of representations—distributed representation’s chief advantage over its more traditional alternatives.

117 McClelland and Rumelhart, Explorations, 16-17. In McClelland and Rumelhart’s usage, “hysteresis” is a broader term than might be concluded from this example. Hysteresis means “delay,” and their use of the term includes the sort of delayed coming to the correct answer that my original example displayed after it received incorrect information as to the sought-after person’s age.
contingent recall is a feature of human cognition which is common and difficult to explain.

In this example we have shown how the order of the input can make a psychologically real difference in the result that the network settles (or does not settle) on. As mentioned above, in such networks duration also matters. Generally, given a steady diet of a single pattern of input, a network will settle into a pattern which is stable. Beyond a certain number of iterations it does not change. It has reached a stable state. This, too, can make a difference in the eventual pattern that the network settles into.

The network that we have been using as an example settles into a stable state by the time it has cycled 100 times with a particular set of input. If we allow the network to stabilize between inputs, the original story that we have been telling comes to a more rapid conclusion. After only the first set of information is fed through, the network settles on two cases with an activation level above our threshold of 50: Rick with an activation of 61 and Ken with an activation of 51. This would allow us to skip forward to the final part of our original story and refuse Rick as someone we already know and then accept Ken as the person we seek, effectively bypassing the age error which misled us in the first instance.

The preceding remarks on the cognitive plausibility of connectionist architectures have focused on areas which are particularly important in modeling the way human agents acting in the world might show their particular pattern of competence and failure. That connectionist architectures show an analogous pattern of success and failure in understanding such instances has been taken as evidence of the cognitive plausibility of such architectures for educational purposes.
Learning: Ill-Formed Problems and Connectionism

The Sharks example and the discussion above focus most directly on issues of contextual recall. While such examples gives a sense of the power and the flavor of connectionist models, of more fundamental interest to education is the way such models illuminate issues of learning. To date, learning theory has consisted almost exclusively of descriptions of the conditions under which learning occurs. In part this has been because there was no credible theory of how the differences that underpin learning take place. We could not answer a question which asked what material process supported learning. In part, though, the failure to pursue the actual mechanisms of learning was founded in the presumption that the material instantiation was relatively trivial. We already knew that thought took the form of logic. The tradition from Plato through Boole and Kant and on to such educationally important derivations as Bruner's and Piaget's have all assumed that real, valuable thought was logical.118

Cognitive science has been seen as an extension of this tradition, with the crucial difference that these researchers took the presumption that logic is the basis for thought and combined it with the materiality of logic that the digital serial computer represents. Faced with the problem of brittle programs—programs which could not transfer their expertise out of the very narrow range for which they were designed—it became apparent that logic alone was insufficient. Researchers concluded that there had to be a much larger amount

118 Writers as different as Piaget and Bruner agree that classification is done through features or attributes, leaving completely unexplained how such features are recognized. While both authors change many positions in later work, their early, influential writing serves as a good example of the way in which logical conceptions of thought have been dominant. See: Barbel Inhelder and Jean Piaget, The Early Growth of Logic in the Child, Classification and Seriation (New York: Harper and Row, 1964) and Jerome S. Bruner, Jacqueline J. Goodnow, and George A. Austin, A Study of Thinking (Huntington, NY: R. E. Krieger, 1977).
of contextually appropriate, domain-specific knowledge.\textsuperscript{119} There were at least three ways of attacking this problem: one could claim the knowledge was wired in by evolution, one could claim that the problem space could be made manageable by rules of thumb (heuristics) and thereby establish a tractable space within which logic could operate, or one could lean on learning.\textsuperscript{120} The prewired solution works only for restricted domains—it is not considered reasonable to suggest that we have evolved to go grocery shopping. The second solution is the route that most cognitive scientists and workers in artificial intelligence have taken. This has led to projects to describe the world,\textsuperscript{121} to the elaboration of domain-specific agents which use local heuristics to evade the problem of complexity,\textsuperscript{122} and, building on this latter idea, to concepts of people built of many such domain-specific agents.\textsuperscript{123} The ideas of domain specificity, though adequate for certain engineering objectives, does not meet the needs of those who want an understanding of how an agent can act competently in the world. Domain specificity and the complexity of domain-specific programs make it very difficult to credit the possibility that all these agents come prewired by evolution. Finally the hope that material processes which undergird competence can be known rests on finding a credible explanation for the phenomena of learning.


\textsuperscript{120} These three positions are exemplified by Chomsky, Simon, and Rumelhart et al. respectively.

\textsuperscript{121} Barnett, Kight, Mani, and Rich, “Knowledge and Natural Language Processing,” 30-49.


\textsuperscript{123} Minsky, \textit{The Society of Mind}. 
The trajectory that cognitive science has followed, from a belief in logic and the forms of logic, to a dawning realization of a more pervasive need for domain-specific knowledge to avoid the brittleness problem, to a focus on learning and encouraging learning, is very suggestive of the pattern in education—a path that educators trod long before cognitive science arose. Built on the same presumptions as the current research program in cognitive science, American education has passed from an emphasis on learning as properly training the forms of the mind (and the accompanying emphasis on math, the grammar of Latin, and natural law) to an emphasis on the many specific facts that one must know to operate successfully, to an emphasis on the process of learning. This describes, I suspect, the development of any competence-oriented project which originates within the framework of the analytico-referential discourse.

Educators, having reached the conclusion that they should concentrate on the process of learning, have been at a loss to say what is, exactly, learning. Without a theoretical grasp on the material process that supports learning, there is very little choice but to go with descriptive theories of the conditions under which learning occurs. At its least sophisticated this has been a simple take-up of behaviorist stimulus-response psychology; at its most sophisticated it has posited a complex individual competent within an environment filled with cognitive resources. Both are finally theories of knowledge acquisition and not theories which actually suggest how the change we call learning happens.

Learning and Distributed Representation

Connectionist networks offer a model which suggests that learning need not be regarded as entirely mysterious. It suggests ways of understanding the process of learning which go beyond description to suggest why some things are difficult to learn and others easy. It suggests ways to understand context
and history as useful and necessary parts of the learning process and not merely as sources of error. A nonmysterious process of learning is at the heart of any hopes for a truly effective pedagogy.

The Sharks example allows us to make a distinction between some phenomena that we often refer to as learning and the sorts of learning done in connectionist networks. The Sharks network can be interpreted as being capable of inference. If we look at inference as a specie of the "best fit" problem where an adequate response is made to a query on the basis of incomplete information, the Sharks network infers quite well. It can tell you who is the most Shark-like Shark or who is the closest to a dark-haired, burglar in his thirties who is unmarried—even if there is no such individual. When someone comes to a conclusion of this sort, we often say that the person has learned something about the group. Indeed, that person has done something, something we sometimes label induction, which has been very difficult to explain.

Similarly, we can claim that the Sharks simulation generalizes. It is able to conclude from a list of associated characteristics, "Sharks are in their thirties." As inaccurate as this is, strictly speaking, it is invaluable in our everyday lives. Too, this is not a simple matter of counting up numbers of Sharks and comparing that number to a total number of Sharks and concluding according to a plurality rule that Sharks are most likely to be in their thirties. It is entirely possible to get nuanced generalizations and suggestive failures to generalize. With three elements in the age pool (twenties, thirties, and forties), it would be possible to come up with a distribution of instances in which there was a plurality of thirty-year-olds but in which, otherwise, the forty-year-olds were more typical (more likely to have the "ideal" profession and martial status, for instance) which would so bias the overall network dynamics as to result in a
network which settled on forty as typical, since it is most typical of the most
typical instances. If the balance were closer to being even, it might be unable to
settle on any of the three ages reliably; the return would depend upon the
particular inputs activated and the historical pattern of that activation. This sort
of contextually sensitive recall, a recall which changes as the saliency of various
elements of the situation changes, often looks like a nuanced and carefully
reasoned judgment. Indeed, on the account given here it is carefully reasoned.
But it is not deduced.

Learning in Simple Multi-Layer Connectionist Networks

As suggestive as is the phenomena allowed by the distributed representation
in the Sharks net, there is a much richer tale to tell. While we may be able to see
how the “family resemblance” between different instances of the concrete
mother can lead to a network returning the same label “mother,” (close is good
enough for neural nets) we still have very little idea how the baby can learn the
category in the first place.

The Sharks net was a “hand-crafted” example of one model of recall, a
model rich enough to yield interesting phenomena and simple and explicit
enough to be observable. But the categories and the patterns of connection
between the instances were directly coded. Beneath the differing patterns of
activation these patterns remained stable. As the coder, I would need to specify
the relations that any new member of the Sharks would have to all the old
members. Until we are able to account for the creation of such patterns, we will
be short of the theoretical framework that we need in order to account for
learning.

In general, three factors are needed to move from the distributed network
that we have described so far to one which is capable of generating its own
patterns from patterns of input. First, and perhaps most important, there needs
to be some way of altering the connection strengths that does not depend
(directly) on a programmer. Second, there will need to be more than two layers
in the network, and, third, the activation function will need to be nonlinear.
Such a network would, in theory, be a complete Turing device, able to compute
any statable computation.124

The simplest case of a network that learns is a two-layer associative net. In a
two layer network the patterns of change are relatively easy to comprehend.
The design of such a network connects every node in the input layer to every
node in the output layer. A pattern of activation is presented to the input layer
and is propagated to the output layer. This is one way to regard the basic
design of the Sharks network if we consider all the feature units to be input
units and all the instance units to be the output.125 A two-layer associative net is
set the task of associating a given input pattern, which is taken as an “event,”
with an arbitrary output array, which can be taken as the “name” of the event.
This type of net does not take the strength and sign of the relation between the
input and the output layers as fixed. Instead, the strength of the connection
between units of the two layers is altered slightly in each iteration in the
direction that would produce the correct outcome. It is fairly easy to see in such
a situation that the network would eventually “learn” to produce the correct
“name” for the input. Even with such a simple network some interesting effects
are possible. A single network can learn to associate several different inputs
with their respective names. It can also be trained on distorted versions of each
of their input patterns and will correctly “generalize” by categorizing each

124 Patricia S. Churchland and Terrence J. Sejnowski, The Computational Brain (Cambridge,

125 This is not an entirely accurate analogy, chiefly because in simple two-layer networks
there is no pattern of connection between units in the input array.
slightly distorted version by its correct name. As in the Sharks example discussed above, the typical associative network exhibits interaction effects tied to its distributed representations that are very similar to those exhibited by human memory.

Two layer networks have known computational limits. The most significant basic advances in network architectures have centered around the development of "hidden" layers of units and nonlinear activation functions. The computational limits of two-layer nets have to do with what they can learn and so are of particular interest here. Two-layer networks cannot learn the solution of the classic XOR (exclusive or) connective in bi-valued Boolean logic. In problems of this sort, two conditionals of one type, say "+" and "-", yield one value when the conditionals are the same and another value when they are different. The sign that results from multiplying negative and positive numbers shows this pattern: two dissimilar signs yield a negative and two similar signs yield a positive. This is a simple case of a function which is not linearly disassociatable. That is, there is no single, defining element for which knowing the value is to know the answer to a question posed. Instead, the answer is relational. One must know and compare the values of both elements of the conditional to draw a conclusion. This is a large class of problems and not simply an exotic computational nicety. It goes to the heart of being able to draw relational conclusions.

---

126 The computational limits, particularly the inability to find a solution to the XOR problem, was the major theme of Minsky and Papert's Perceptrons. Marvin Minsky and Seymour Papert, Perceptrons: An Introduction to Computational Geometry; Expanded Edition (Cambridge, MA: MIT Press, 1988).
The solution to this problem lies in expanding the network architecture beyond two layers. Three-layer networks can solve for the XOR relation.\textsuperscript{127} The middle layer, usually called a hidden layer, can be designed to fire when both the conditionals are, say, negative, and thereby send enough activation to the output layer to cause it to yield the correct answer. This simplest case establishes, in principle, that relational conclusions can be drawn from network architectures. It stops short, however, of demonstrating what we are most interesting in seeing demonstrated: it does not show that such a net can "learn" in the incremental fashion discussed above in two-layer associative nets. We need to know that an algorithm exists that will converge on the pattern of weights which will yield the correct answer without explicit external coding.

This problem is difficult just because it is relationally conditional. In a two layer network, the difference between the desired output and the actual output can be computed for each output unit, and that error can be used to adjust the connection weight to each input unit proportionally to the input unit's effect on the net error. Metaphorically you just divvy up the error among those elements that caused it according to whether their activation strength contributed to the error. A too-weak connection gets incremented slightly and a too-strong connection is decremented. But in a three-layer (or more complex) architecture, it is not clear how to assign credit.\textsuperscript{128}


\textsuperscript{128} This is a specie of the more general credit-assignment problem in computational theory. See: William Bechtel and Adele Abrahamsen, \textit{Connectionism and the Mind} (Cambridge, MA: Blackwell, 1991), 86-87.
The solution turns out to allow a small change in the input to the hidden unit to make a large change in its output. This is done by making the output function of the hidden unit sigmoid (see Figure 3-8).\textsuperscript{129}

![Linear and Sigmoid functions](image)

A common function is graphed as a line drawn at an angle on a plane but a sigmoid function, is shaped roughly like an “S.” The function drawn as a straight line describes a situation in which a change in the amount represented by the Y axis always corresponds to a set amount of change in the amount represented by the X axis.

But in a sigmoid function large differences in the flat lower and upper ends of the range produce virtually no difference in output while small differences in the middle of the range produces large differences in output, Given the larger global context this behavior effectively shifts the sensitivity of the unit over its output range “seeking” the area in which its activation strength makes the greatest difference. This magnification of difference allows, after multiple

\textsuperscript{129} Churchland and Sejnowski, \textit{The Computational Brain}, 107-112.
feedback iterations, the 3-layer or more complex network to settle into a pattern of relations which minimizes output error.¹³⁰

This ability to learn the canonical set of logical functions is, in one sense, a task set for network theorists by the presumptions of the analytico-referential discourse discussed in chapter 2. We may express some doubt as to the importance of the task outside of an unquestioned acceptance of the primacy of logic. Nonetheless, to again refer to MacIntyre, a theoretical framework can be considered superior to its predecessor if and only if it can solve the problems generated by its predecessor.¹³¹ We may be in the waning days of a discourse, but we too must respond to the problems that discourse has created. More concretely, much of what students are expected to learn in education is couched in just such a set of logical functions. The ability of networks to model successful problem-solving of this sort is crucial to their plausibility in a specifically educational context. A recurrent critique of networks is that while they may be fine for certain low-level sorts of learning like perception and categorization, they are not capable of handling higher level functions—by which is meant logic relations and deductive conclusions over the symbols of language taken as discrete objects.¹³² It is true, and possibly an element of a connectionist critique of schooling, that school tasks are quite often

¹³⁰ Technically, the point is a good bit more complex than this, with some algorithms finding the absolute minimum at considerable computational cost (Boltzman class automata, for instance. See: Churchland and Sjenowski, The Computational Brain, chapter 3), but with a larger class of algorithms and their associated virtual machines finding local minima at a much lower computational cost. Whether these cheaper and more biologically realistic solutions learn quickly enough and well enough to suit human purposes will be an empirical matter in each particular instance.


characterizable in this way. The results set out above demonstrate that this need not be a reason to reject network architectures as an important element in a revised theory of learning. The exclusive or (XOR) relation in networks cited above can be inverted to solve for the case of not exclusive or (NXOR). Such a relationship is the formal equivalent of the more familiar "if and only if" relationship. Networks are capable, in other words, of learning to reach a correct answer even in those cases where a single datum among arbitrarily many determines the outcome.

Explanations of learning based on logical relations, which posit the ability to solve such relations as primitives, model relations of this nature more cleanly than networks. But the question for educators is whether the additional explanatory powers of distributed representations and the additional constraints on theorization are reason enough to abandon the traditional logical model of the transformation that underlies learning and adopt a theory based on networked relations. Chapters 5 and 6 below will seek to expand on the practical implications of this possibility.

Learning, the Teacher, and Environmental Regularities

As we have discussed the basis for learning in network architectures, we have progressed from talking of the learning-like qualities of distributed representations possessed by network architectures in general to learning in two-layer associative networks to learning in multi-layer networks. It has been shown that network architectures have patterns of success and failure in contextually sensitive situations that are suggestive of the patterns of success and failure of human reasoning. Multi-layer networks can also correctly solve formal logical problems.

To this point the most complex networks described have been multi-layer networks incorporating an algorithm which incrementally adjusts the
connection weights so that the gradual, internal reorganization of the network finally returns the desired output. These multilayer networks are usually called "feedforward" networks in reference to the direction of the propagation of activation. Activation feeds from the input layers forward to the hidden layers and then to the output layer. There are two broad critiques of such networks which are tied closely to the feeling that such networks are not "really" learning. One is that the learning that is shown is not real learning because the back propagation of error messages constitutes an omnipresent teacher. The other is that such networks are actually time-bound in significant ways and that they cannot adequately represent crucial psychological effects that depend upon, in effect postponing, the meaning of a current input until its appropriate context is revealed. Syntax in language, and particularly subject/verb agreement and embedded clauses, are often proposed as areas in which network architectures are inadequate.

From the standpoint of an educational theory of learning, the complaint that real learning is the sort of thing which happens without a teacher is largely without force. It seems unremarkable that learning would require a teacher—the assumption that real learning takes place spontaneously has been suggested in education, but not in a context in which spontaneous excluded teacher-arranged regularities which appear in "learning environment." In an


135 I am thinking of Montessori doctrine in particular, which holds that learning is a spontaneous accompaniment of play but which rigorously organizes the environment in which such play occurs to contain the regularities which are to be learned. Piagetian theory, of course, had very similar assumptions.
educational context the sharp distinction between "environment" and "the teacher" seems naive. It is true, nonetheless, that neural net architectures such as the ones that we have discussed depend upon repeated interaction with an environment that is stable (in the sense that it presents the sort of slightly varying regularities that networks excelling in extracting pattern from during its history of interaction.) It is further the case that networks are very slow at extracting such pattern unless the network has been designed with the particular task in mind or has been trained on a set of regularities that were specially singled out for their pedagogical value. Again, this does not, in the current viewpoint, appear as a reasonable indictment of using network architectures to model of learning. Accepting the use of specially designed neural nets only amounts to the admission that there is some organization of the brain above the very micro level of the neuron or cell assembly that most networks model. Connectionist researchers have no quarrel with such an assumption and, in fact, much work in the area is directly inspired by modeling particular human perceptual architectures. Depending on environmental regularities only amounts to saying that the learning environment must show systematic regularities and that a part of the teacher's task is to arrange for the appropriate regularities to be significant for the student. While these sort of objection is not very salient considering our current purposes, they do highlight the need for a broader framework than that which connectionist theorists usually adopt. Connectionists do show a certain naïveté in assuming that such regularities are readily apparent in the hurly-burly of the world. Education's more practical project pushes educators toward making explicit what cognitive scientists may reasonably treat as peripheral to their concerns of explicating

136 Mead's work stays particularly close to the perceptual substrate. For example, see: Misha A. Mahowald and Carver Mead, "The Silicon Retina," Scientific American 264 (May, 1991): 76-82.
cognition. The implications of this point will be expanded upon in chapter 4, where we will examine the specific utility of situated cognition in providing a basis for understanding the environmental regularities upon which network architectures rely.

Learning Temporal Patterns in Recurrent Networks

While the interests which motivate this work may sharply differentiate the educational and cognitivist concerns regarding the teacher, such is not the case in the second objection recently cited: that such time-sensitive effects as have been shown are insufficient to explain large areas of human capacities and the associated abilities to learn these capacities, which are necessary if we are to take network architectures as a reasonable starting point for an educational theory of learning. Indeed, if it were true that networks could not learn recursive and temporal patterning as has been charged, and if a serial architecture coupled with stable, site-addressable memory were necessary for this class of learning, as has been charged, it would be a strong indictment of the connectionist position. A large class of humanly learnable patterns would be covered under such rubric and, unlike the agnostic position that this work adopts toward the demands that learning model formal logic, such a deficit would affect the ability to explain many practical, in-the-world abilities that people conspicuously share. Beyond the syntactic and linguistic problems already mentioned, temporal coordination is required for such commonplace faculties as the capacity to appreciate music, the ability to dance and even such simple activities as grasping a pencil.

In large part, such critiques are in reaction to feedforward networks which learn by the backpropagation of error—the sorts of multi-layer networks
discussed above. While these are powerful learning architectures which effectively model perception and associative recall, such networks do have at least some of the limits that critics have asserted. What such networks lack are internal feedback. Internal feedback has been generally avoided, in part because the interpretation of network activity in the presence of such feedback, especially in the boundary case where the feedback crosses layers, is conceptually very difficult. In fact such boundary-crossing feedback can lead to conditions which are formally chaotic under the definitions used in the study of nonlinear dynamical systems—a point which will be discussed below. Internal feedback can take such limited forms as the inclusion of nodes whose activation strength is predicated in part on its activation level in the previous cycle. It can extend to include general within-layer interconnections and, most radically, to propagation of activation strength from higher to lower levels of the network. Networks which include such recursive feedback are known as recurrent nets—nets in which activation is passed in such a way that parts of it recur. This allows such networks to store previous states of the network and


138 Network architectures, in general, are flexible enough that separating architectures by type and attributing capacities to these types as I have done here is a heuristic, rather strictly accurate, characterization. For instance, speech recognition networks typically make use of “time delay” neural networks. In such nets the initial activation pattern is passed to only some of the input nodes, and their activation is preserved by being propagated sideways in the input layer as well as forward through a standard feedforward, multilayer network. Generally, this lateral propagation does not involve a transformation and the connections to the hidden units symmetrical to those receiving current input. Viewed from the standpoint of the hidden units, each input “moment” consists of several serial moments of “real time.” Is this a recurrent or a feedforward architecture? The answer largely depends on the use to which it is put.

139 “Backpropogation of activation” should be sharply distinguished from the backpropagation of weight adjustment in a standard feedforward network. It has serious computational consequences but is a biologically more plausible learning method. For the initial theorization of such “tangled hierarchies,” see Douglas R. Hofstadter: “Waking Up From the Boolean Dream,” chap. in *Metamagical Themas: Questing for the Essence of Mind and Pattern* (Bantam: New York, 1985), 631-665.
implicitly “compare” past states to current conditions. Doing so allows one to arbitrarily extend the time frame that can be considered far beyond the relatively short time that it takes a pattern of input to propagate forward to the output layers.\textsuperscript{140} Recurrence allows networks to compare temporal patterns and, interestingly, to resolve figure-ground ambiguities and segmentation ambiguities by allowing multiple, discrete “interpretations” to be compared.\textsuperscript{141}

\textbf{Complex Networks, Self-Organization, and Extended Learning}

Recurrent networks open up vast realms of possibility in connectionist inspired modeling and theorization. So far, we have avoided the use of meta-theoretical frameworks such as chaos theory, nonlinear dynamics, automaton theory, or complexity. In part this has been a response to the awareness that a theoretical structure such as complexity theory is too strong. It can be used to model almost any phenomena, including those to which it adds very little in the way of explanatory power. Arguably, feedforward networks trained by backpropagation are cases of complex nonlinear systems. But they can be understood fairly easily without reference to the more general arsenal of theoretical constructs offered by such frameworks. With recurrent networks, however, we enter a realm of complexity so great as to make recourse to these tools a useful measure.\textsuperscript{142}

\hspace{1cm}

\textsuperscript{140} Not all network theorists consider this a central problem. Particularly those working in perception make the point, “Let time be its own representation.” See the remarks of Carver Mead reported in Churchland and Sjenowski, \textit{The Computational Brain}, 120.

\textsuperscript{141} Churchland and Sjenowski, \textit{The Computational Brain}, 117.

\textsuperscript{142} For an extremely useful characterization of the circumstances under which separate research programs in neurology, biochemistry, and genetics have led to a rejection of localization of function and an endorsement of emergent phenomena in scientific explanation, see William Bechtel and Robert C. Richardson, \textit{Discovering Complexity} (Princeton, NJ: Princeton University Press, 1993).
At the heart of the additional complexity that recurrence introduces is the possibility that some elements of a network can be relatively isolated from change vis-a-vis other elements of a network. These relatively stable but still plastic regions of the network constitute semiautonomous entities within the network. Such semiautonomy is the potential basis for explaining a wide array of phenomena. Among these is the emergence of hierarchy and the possibility of large scale “resonance.”\textsuperscript{143} More grandly, imagination in the form of projective possibility, the concurrent ability to plan, and the existence of progressively more complex forms of abstraction which can feed backward to affect perception are enabled.

Such a prize is well worth seeking. But to do so more fully we will need to take a short side trip into the terminology of complex systems.\textsuperscript{144} Much of this will seem familiar to the attentive reader for the simple reason that the connectionist work on learning that we have been reviewing is a major source of such theory. In fact Doyne Farmer, in a seminal article, suggests that connectionist networks can be seen as the prototypical general case for the sciences of complexity and introduces a terminology based on its usages.\textsuperscript{145} The phenomena that is of the most interest in discussing the influences of the complex systems theory on a theory of learning is emergence. Emergence is


\textsuperscript{144} Tracking the emergence of dynamical classes of explanation is a project as yet undone. It appears to be a case of multiple emergences, with ecological and ultimately Darwinian conceptions claiming precedence.

said to occur in complex systems when the system exhibits characteristics which are not properties of the smaller units. Holism of this sort is not mysterious—it is explained on a relational basis. Generally the units or nodes are simple objects whose relationship is governed by simple rules. This has been the case in the networks discussed above. Dynamical systems are treated as open systems, meaning that such systems are open to outside perturbation or disturbance which can change the patterns of relation of the nodes. However, it is crucial that such systems be bounded. If there is no limit, however porous, the systemic relations simply dissolve. Such systems are said to be self-organizing in that the regular patterns of relations that connect nodes lead, in response to external perturbation, to the creation of new internal structures. This arises, in the abstract, when out of the flux of the initial disorder occasioned by the perturbation, particular patterns of relationships between the simple units arise which are more stable than alternate patterns of relationship. These patterns, by virtue of their relative stability, become the dominant ones within the system. There may be one such pattern indefinitely repeated or there may be several competing patterns.

In either case, under a stable pattern of external perturbation a stable pattern of relations will emerge. These structures are understood as self-organized states of the complex system. Typically there are multiple stable states the system can take under different levels of perturbation. In cases which are especially interesting in view of our present purposes, such states lead to internal conditions which are not only relatively stable in regard to the level of perturbation but which also interact with the environment in a way that stabilizes the level of external perturbation. Such systems can be understood as
adaptive. Insofar as these states enable adaptive changes, the system can be said to have mapped its environment. The stable states that are possible under such systems are not time-reversible; that is, history matters. The evolution of the system closes off possibilities as interactions with the environment continue.

Connectionist systems map closely onto such systems. Simple units are connected by simple rules. Input to a net can be considered perturbation. Feedback, whether in the guise of connection strengths or activation strengths, alters the internal structure of the network. On a purely local, broadly parallel basis, each unit or its relationship to other units is altered in response to relationships that are internal to the network. Over time, or multiple iterations of input, the network settles into a state in which the difference between desired output and produced output is minimized. It has mapped its environment as well as the original structure and its history allow. Artificial networks of this kind can be viewed as networks which have been designed to seek stability.

Recurrence is the key to seeing network models as adaptive complex systems of more than the simplest kinds. Without recurrence—internal

---

146 While "adaptive" is a term particularly associated with Holland's rather technical use in classifier networks, the instinct to ally it with Darwinian evolution is not misplaced. Philosophically, the Darwinian vision of many simple animals interacting in competition and small differences between them being selected for is a clear example of a dynamical system. The revolutionary nature of Darwin's suggestions is widely acknowledged and Prigogine, a central figure in the philosophy of complexity, has made explicit his debt in this regard. See Prigogine and Stengers, *Order Out of Chaos*, 215.

147 Maturana, and Varela in a series of works, make this internal relationship central, emphasizing the "structure determined" nature of biological organisms interacting with their environment. The current work, because of its emphasis on human learning, on change, emphasizes how the structures that determine change—in part through interaction with "the world." Where we start a hermeneutic analysis of the relationship between organism and environment (or even starting from this opposition) is largely a matter of pragmatic purposes in view. See: Humberto R. Maturana and Francisco J. Varela, *The Tree of Knowledge, The Biological Roots of Human Understanding* (Boston: Shambala, 1987). and Francisco J. Varela, Evan, Thompson, and Eleanor Rosch, *The Embodied Mind: Cognitive Science and Human Experience* (Cambridge, MA: MIT Press, 1991).
feedback—networks can map only one “thing” at a time: the total field. While a simple network can map multiple items—the Sharks net examined above demonstrates this—it can only actively map one item at time. With recurrence multiple areas of stability which map multiple contingent items in the field are possible. It would be possible, for instance, to simultaneously hold both the child’s mother and the child’s aunt as possible “faces” and resolve this ambiguity on the basis of subsequent input.

The issue of multiple, semiautonomous mappings is new enough to be little discussed theoretically. Our interests lead us to solutions to these problems which focus on networks that learn to partition their solution space. One class of solutions, applicable chiefly to perceptual learning, is to incorporate activation functions for the hidden units which make each hidden unit sensitive to a limited range of input units and sets up a competition, similar to that within Sharks pools discussed above, for adjacent hidden units. In this way the network is driven to have semiautonomous internal structures, each of which “represent” only a limited range of input. This strategy greatly reduces the amount of retraining and representational degradation which occurs when a network needs to be retrained on subtly but significantly different input patterns. In effect only the changed portions need to relearn to associate the new input with appropriate output. Another, related strategy is to set up competing networks rather than competing nodes in the hidden layers. In such networks all inputs are fed to all the available hidden units in the initial layer. There is no interaction between mini-nets on the same level. A referee network receives input and compares it to the desired output, determines which

148 The literature that I here interpret to reflect on this issue is dispersed and is often the product of narrowly drawn problems and equally narrowly drawn solutions. One useful if technical summary is Churchland and Sjenowski, The Computational Brain, 125-130.
semiautonomous mini-net is closest to the correct output, and sends backpropagation error messages only to that network. The referee network is trained concurrently with the other mini-nets and its performance improves over time. This managed competition results in a partition of the instances between differing mini-nets on the basis of global similarity with each subnet specializing in a particular subset of instances. The Jacobs et al. paper (1991) from which this description is drawn is a speech recognition system which learns to recognize vowels uttered by multiple speakers. Its networks, depending on the particular history of cases which they are presented, learn to specialize in voice types such as children's or men's voices.149

While the institution of such semiautonomy as is illustrated above demonstrates the possibility and the power of semiautonomous networks, very little is yet known about just how, computationally, semiautonomous regions appear. One response might be to claim that leaning on preexistent structure, as the above example implicitly does, is not neurological unreasonable.150 Another would be to lean to some degree on activity in the world. Neural nets as currently implemented are, necessarily, completely passive. That is, they are unable to slightly change the angle at which they view an item to bring it into closer alignment with a previously seen example. Both solutions seem likely, but

---

149 While technical, such networks may be commercially invaluable, as they are the basis for much of the connectionist-based work that is being done in the area of speech recognition. See R. A. Jacobs, M. I. Jordon, S. J. Nowlan, and G. E. Hinton, “Adaptive Mixtures of Local Experts,” Neural Computation 3 (Fall, 1991): 79-87.

this remains an unsettled area of connectionist research which bears watching by the educational community.

In summary, recurrent networks make possible a degree of complexity in which a wide range of "higher" cognitive functions can be modeled. This is possible without abandoning the properties of distributed representations and the associative learning that first attracted the attention of learning researchers. While much work remains to be done in this area, there is little reason, in principle, to think that network architectures cannot subserve a very broad range of human activities.\textsuperscript{151}

\textit{Summary: The Cognitive Plausibility of Connectionist Networks}

In this section we have moved from a brief examination of the classical account of learning based in prepositional logic and deduction to an extended examination of the cognitive characteristics of network architectures with a special emphasis on the implications of these architectures for learning. This review has viewed relatively simple and universal features of networks such as emergence and distributed representation against a backdrop of worldly complexity. Connectionist architectures show a pattern of success and failure roughly analogous to the broad human pattern and in distinct contrast to the cognitivist approaches which grow more directly out of traditional approaches to knowledge. Learning is treated in some detail, with our understanding of

\begin{footnotesize}
\textsuperscript{151} A recurrent call, from both sides of the divide that separates connectionist and symbolic modelers of cognition, is for so-called mixed networks. On the surface such a "compromise" may appear eminently reasonable: let each style of modeling do what it does best. Unfortunately, it is not evident that such a division of labor can be maintained or that those suggesting such a division of labor actually agree about what would be demonstrated in mixed systems. Symbolists tend to see connectionism, grudgingly, as implementation routines for troublesome subsystems such as perception. Connectionists mean something very unlike what the classical symbologists mean when they talk of symbol manipulation. See, for instance: Paul Smolensky, "On the Proper Treatment of Connectionism," \textit{Behavioral and Brain Sciences} 11 (March, 1988): 1-74.
\end{footnotesize}
what is possible with network architectures growing as we examine more complex examples.

We turn now to brief examinations of the neural and computational plausibility of network architectures. While important and useful sources of constraint on theorizing and deep sources of inspiration for further modeling that bear on the central issues of this work issues of neural and computational are less directly applicable to educational concerns than the cognitive modeling discussed at greater length above.

The Neural Plausibility of Connectionist Models

The neural plausibility of the connectionist model is one of its chief rhetorical strengths. Connectionist theorizing is significantly constrained and deeply inspired by what is known about the brain. This work is will endorse a rather weak version of the neural analogy. I take it that what is of interest to educators is that networks model effectively the cognitive ability to learn that humans share. In taking this position I view neurology as providing only a set of constraints to work in rather than as providing the phenomena to be modeled and understood. The cognitive modeling which is done should not violate what is known of the brain and its organization.

Most critically for the purposes of this work, the learning processes described above should be interpretable as credible simplifications of actual neural mechanisms. A brief review of the similarities and differences between actual neural networks and connectionist networks will serve to provide the reader with a general basis for making a judgment on this issue.

In broadest strokes the connectionist vision is plausible. That is, the brain is composed of large numbers of simple neurons which interact simply and locally. Unless we take a dualist position, human capacities must, finally, be
explained on some basis which does not contradict the elemental facts of neurology. As was discussed in chapter 1, there is neither enough time nor enough space to accommodate an image of the mind based on the von Neumann computer analogy derived from the traditional, logicist approach. What processing is done in the brain must be done in a distributed parallel fashion on the evidence that we currently have.

At the lowest level many details of the networks instanciated here are plausible. Neurons do appear to be quite simple units which do connect to many other such units. Neurons do exhibit the threshold phenomena where they spike only after they are sufficiently stimulated by input. The sigmoid activation function introduced above as part of the modification to the basic network model (which in combination with hidden layers allows the modeling of crucial logical functions) is based on what is known of the sensitivity of actual neurons. They, too, show nonlinear activation sensitivities.

On the larger scale of physiology, two-dimensional mappings of motor and sensory units corresponding to the device of hidden layers in connectionist networks is one of the more fascinating and well-established discoveries of modern physiology.\[152\] Similarly, the mini-nets instanciated in certain recurrent networks discussed above are similar in concept to the well known phenomena of cortical columns in which much of the interim processing of sensory input seems to take place. The development of these cortical columns in the visual system is a fascinating case of self-organization that occurs only during an animal’s experience with the world.\[153\] Such columns are relatively isolated,


being much more densely interconnected within the columns than they are connected to other adjacent tissue. Recurrence, or looping internal feedback, is a massive fact of neural interaction.

In light of evidence of this type, it is apparent that a convincing case can be made that connectionist networks are a simplified model of neural function. Indeed, there are those in the field who take just this position. Our continued focus on learning will lead us to take a slightly more nuanced and cautious position.

Computational neuroscience has its roots in the simple proposition of Hebbian learning. Hebbian learning, in turn, is based upon the assumptions of the classic McCulloch-Pitts neuron. It is not certain that complete analogs to either concept actually appear in the brain; where they do appear to be fairly solid instances, it is not clear that these instances can be generalized beyond the specific, restricted region in which they appear. Tracking this research in detail is beyond the warrant of this work, but a restricted review of the problems here will be useful in grounding the claim that connectionist networks do model the brain in ways that are interesting and useful for a theory of pedagogy. This claim requires a fair amount of unpacking. It will call for an outline of the principles upon which connectionist learning architectures are based, a review of the results of attempts to locate material processes which implement such architectures, and an explanation of the philosophy of science.


155 The interested reader, however, is enthusiastically referred to Churchland and Sjenowski’s excellent book The Computational Brain, and especially chapter 5, “Plasticity: Cells, Circuits, Brains, and Behavior,” for a lively if technical general exposition of the ongoing research in this area.
which supports the particular style of research which computational neuroscientists and connectionists support.

Connectionists assume the validity of the basic process of associative learning. In particular, they endorse various species of Donald Hebb's original formalization of the principle, in which he says:

When an axon of cell A is near enough to excite cell B or repeatedly takes part in firing it, some growth or metabolic change takes place in both cells such that A's efficiency, as on the cells firing B, is increased.\footnote{D. O. Hebb, \textit{Organization of Behavior} (New York: Wiley, 1949), 62.}

So stated, this is a \textit{principle} rather than a proposal of a specific mechanism. There is a great deal of collateral evidence that some mechanism satisfying this principle must be at work. Workers in neurology generally assume that a mechanism (or, more likely, mechanisms) exist which satisfy this principle. It is this principle (and not a specific mechanism) which connectionists model when they design networks whose relations include weight changes in response to the input patterns taken as experience. As usually interpreted, this formulation assumes the existence of a McCulloch-Pitts neuron which either fires or does not fire. We have already encountered this on/off interpretation in our discussion of the basic logic gates which support serial computation. In general, research has not supported the existence of a McCulloch-Pitts neuron. Instead, most neurons appear to signal difference on the basis of \textit{rate} of fire rather than on a binary basis. This means that the brain is better regarded as built of analog units rather than digital ones.\footnote{Or at least I so interpret it. This interpretation is not controversial but does depend crucially upon accepting that the synaptic junction is the fundamental computational unit of the brain. While there is real argument about this, most that adopt larger arrays, such as cell assemblies, as basic units are more, not less, committed to an analog interpretation. (See Edelman, \textit{Bright Air, Brilliant Fire} for a clear statement and review of research.) Even so this leaves open the possibility that at some level some neural computation takes place on a basis best characterized as digital.}
Further, there are apparently several Hebbian mechanisms which change the relative efficacy of the synaptic connections in response to changes in input activity. None of these mechanisms are securely established in the sense that all the gaps in the experimental evidence have been closed. Even in areas where considerable and convincing evidence exists—for instance, the role of the hippocampus in enabling the transfer of memory into long term storage—it is not immediately apparent that identical principles will apply elsewhere in the brain.

Such a gap at the center of an active and respected program of research is disconcerting and deserving of some reflection. How much confidence is reasonably shown in the conclusions drawn from working inside this program? An answer must be based on our shared philosophy of science. To some degree, it will be argued here, our hesitancy is based in a misunderstanding of the relationship between empirical evidence and theory in science. Regardless of the wide dissemination and acceptance of alternate, more socially constrained perspectives, science is still widely regarded as properly proceeding from empirical evidence to theory. A more sophisticated approach, and a more historically accurate one, would be that the concepts of theory and evidence are a product of a certain coevolution. Let us consider one of the most influential scientific theories to illustrate this point: evolution. Evolution has a similar gap at the heart of its endeavor. It depends crucially on Darwin's principle of descent with modification. For a hundred years there was almost no idea of the

158 Churchland and Sjenowski, The Computational Brain, 250-254.

nature of the mechanism which supported the principle. Even following the
discovery of DNA the crucial linkage between the presumptive genetic
structure and the environmentally selected phenotype is opaque. By and large,
while we do not doubt that such a connection exists, we simply do not know
exactly what it is.160 In the history of seeking such a mechanism, we have
altered both our theory and our idea concerning what the mechanism might
be—all without abandoning the principle of descent with modification (though
we mean something a good bit different and certainly more specific these days).

This suggests that a principle may be useful in a theory or model if it breeds
fruitful research, and that a good principle, such as descent, may survive—even
if with modification. The general principle of the backpropagation algorithm,
Hebbian learning, appears to be biologically realistic. Long term potentiation—
the technical term for changes in the “irritability” of a neuron—is an observed
phenomena; the details of exactly how this occurs have yet to be worked out.

While connectionist networks of the type we are interested in are
constrained by their commitment to neural realism, their reason for existence is
to generate a plausible theory or abstraction of the way the brain produces its
results. Slavish adherence to reproducing the brain in silicon would not
simplify and would not serve the purpose of theory-building. A model of an
historical event which merely reproduced every detail would not be a theory in
any useful sense. A simplified model which captured the phenomena of interest
would, on the other hand, be very interesting. The connectionist hope is to
identify the principles which govern network generativity without simply

160 This is a very exacting area of research and has more than passing relevance to the
networked, holistic approach advocated in this work. See: William Bechtel and Robert C.
Richardson, Discovering Complexity: Decomposition and Localization as Strategies in Scientific
reproducing the phenomena in question. A model which does so constitutes a very specific and testable theory.

This is not to say that the particularities do not matter. The existence of sigmoid activation curves in actual nervous cells inspired the development of nonlinear activation functions in network architectures that helped overcome specific learning problems in multi-layer networks. That we do not understand the details leaves room for many more fundamental discoveries to emerge out of neurological research. Among the largely unmodeled features of the human brain is the brain's analog, not digital, operation. It is also uncertain that the neuron is the basic computational unit or that it is as simple as connectionists assume. Most networks are much less internally differentiated than the brain and are much, much less complex in terms of sheer quantities of connections. Most networks pay no attention to nonneural chemical changes which affect the rate of neural fire across broad areas. Recurrent networks have modeled recursive, cross-level feedback very minimally. All of these features of brain processes are ones which could conceivably strongly shape the emergent features of a network. For example, "smarter" rather than "simple" units of computation, hierarchy, structure, and sheer quantity are all ways in which the rapidity of learning has been improved in particular network implementations.

Current connectionist network modeling is essentially betting on parallelism, distributed representation, and Hebbian learning as constituting

161 Indeed, it seems likely that cell assemblies, groups of cells that react in concert, are the computational entities described by nodes in most network models. See Edelman, *Bright Air, Brilliant Fire*, 81-98.


163 Churchland and Sjenowski, *Computational Brain*, 117.
the appropriate abstractions which will enable adequate theory building. This may not succeed, but the process of exploring our current network paradigm should be revealing. In the end a useful match between levels of abstraction and particular applications such as education will be a matter of empirical work which maps the phenomena of interest onto particular network models and then maps the activities of such networks back onto our observations of real-world phenomena.

This analysis, while generally affirming the neural plausibility of network architectures, gives us reason to show some caution in assuming that the current abstractions will map directly onto the sorts of learning that occur in the classroom. We can conclude that, while we may be finding some general principles which govern learning in networks (and that the brain is a complex instance of such a network), we should not assume that general models will map directly onto a particular human task. We will not be able to test a learning method on a machine. But the real promise is that we may be able to devise a learning method with a better chance of succeeding with actual students by using the principles that are discovered via modeling brain processes.

The Computational Plausibility of Connectionist Models

What would it mean to say that a model is computationally plausible? Computational theory is a very general framework which draws boundaries around what can be said to be possible and impossible. On the positive side it defines a general class of mechanisms which can compute any well-defined problem, given certain conditions. On the negative side it lends us a set of tools with which to examine particular implementations and to define computational architectures which can reasonably be expected to do particular jobs. A computationally plausible approach would be one which is powerful enough to
solve the complex problems with which it is faced and which will get the job
done in a reasonable amount of time on the available hardware.

Before we begin such an analysis, however, it is worthwhile to note that
network architectures operate most usefully outside of the boundaries that it
defines. A well-defined problem is, in a nutshell, a problem which can be
transformed into a statement in formal logic. As we have seen above, one of the
crucial preconditions of formal logic is that the categories be discrete.
Additionally, the axiomatic basis for the statement examined must be free of
contradiction. A very large part of the attraction of network architectures is that
particular implementations can be designed to work passably well when
neither of these conditions is true. What is lost in the trade from the traditional
point of view is certainty. Networks which tolerate such conditions well do not
necessarily find the best answer to a problem—they can become “stuck” in a
poor solution and it may be very difficult to recover from this state. This was
discussed briefly as “hysteresis” earlier in this chapter. From the point of view
of education, it is precisely the ill-defined problems of the world that are the
most interesting, and computational theory in its more formal instantiations
has little to say about such conditions except to point out the inappropriateness
of formal tools for the task at hand. Given the historical predilection toward
mathesis in our culture, however, such a service is of no small import. It is
rhetorically useful to be able to say that there is a rigorous, well-defined, logical
reason to reject formal logic as a basic framework for understanding our
interaction with the world.164

164 It should be noted that I cannot close a particular logical hole in my argument. To accept
the position outlined in this paragraph, one must accept both that the world itself does not
conform to the forms of formal logic and that the network structure of the brain which supports
the mind does not provide a suitable basis for formal logic. As one might suspect from my
exposition on the analytico-referential discourse, not everyone will accept these conditions. The
unity of the mind, the world, and logic on the basis of forms is the crucial identity upon which
The Strength of Network Architectures

Are network architectures strong enough to solve the problems we would set up for them? With the caveat noted that even a positive response to such a question would apply to only the small range of the phenomena that interest educators, the answer appears to be yes. This restricted question amounts to asking if networks can solve the full range of logical problems or only a restricted subset. Logic, and the ability to check our reasoning using logic, is an important scholastic task and one of which humans are capable—even if that capacity appears much more limited than we once assumed. If it were the case that networks could not be configured to solve such problems, it would constitute a strong argument that networks were at best an incomplete solution to the problem of what processes subserve cognition, and at worst that it is simply wrong. As was discussed above concerning whether networks could learn to solve the full range of logical functions, the answer is yes. Learning sets a higher bar than the simpler question asked here. For learning to solve logical functions we needed multilayer networks and a nonlinear activation function. For properly solving a logical problem we need only a three-layer network. There are a number of different combinations of connection strengths and activation thresholds (mini-nets) which will satisfy the most difficult of the logical problems: XOR. By properly arranging these mini-nets hierarchically, logical problems of arbitrary complexity can be solved. In formal terms such

---

165 The claim that network architectures were restricted to a limited number of logical functions was the central argument of Minsky and Papert’s influential book *Perceptrons.*

complex networks can be seen as a special type of Turing machine—a computer as powerful as any other.

However, network architectures are less efficient computationally than alternate architectures; many more calculations must be done to arrive at the answer computationally than with the contrasting serial architecture. In effect networks find it harder to do logic than do standard computers. But we are not directly interested in finding the most efficient solution to a particular computational problem; instead we are interested in modeling human cognitive capabilities. It is more important to our purposes that a model mimic the successes and failures of humans—and humans notoriously find logic more difficult than perception or other cognitive functions. Computationally we only need to establish that networks are capable of logic to say that they are computationally plausible in regard to this issue.

Time and Space Constraints in Network Architectures

If network architectures are strong enough to handle the job posed for them it remains to be seen whether it is sensible to expect that such architectures will be able to complete these tasks in a reasonable time, given the constraints that the brain imposes on how these tasks may be done. The preceding section, which reviewed neurological plausibility, will have prepared us to see, for the simple reason that network architectures were in part developed to explain how the brain's particular structure could give rise to cognitive capacities, that there is a plausible correspondence between the brain and connectionist networks. Given the capacities of network architectures that we have examined and the tremendously more complex, layered, and differentiated nature of the brain—all of which increase computational power—it seems entirely possible, even likely, that the brain is best described as a network processor.
Computational theory, by its very generality, can say little that is positive about whether human capacities do arise from network processes. It is limited to demonstrating that it is not impossible that this is so. Perhaps more interesting for our purposes is the possibility of demonstrating that other, competing architectures may be considered computationally impossible.

Chiefly, of course, we are concerned with showing that there are formal reasons for rejecting the logicist position discussed in chapter 2 and identified there with Descartes and the further development of the analytico-referential discourse. An attempt to show this turns upon a closer consideration of Turing machines and the role they play in computational theory than has been pursued to this point. By calling a mechanism a Turing machine, one refers to both a powerful general concept in computational theory and a moment in the intellectual history of mathematics and logical philosophy which seems to seal off the possibility of the mathesis of which Descartes and Liebniz dreamed.

The moment in history that Turing is in part remembered for centers on the logical project to formalize mathematics that was known as the Hilbert program. This project had inspired Russell and his friend Whitehead to write a book which attempted to restate mathematics in a bi-valued logical format. Initially the task was viewed as a tedious but necessary chore to establish a firm basis for the arithmetic functions which would support a more rigorous philosophy—a philosophy in the Liebnizian tradition of mathematizing a full description of the world. While this work was still in progress, the Austrian logician Gödel restated the problem of paradox in formal terms. The classical form of the logical problem of paradox is the old riddle of the Cretan liar in

---

Turing is also remembered as the leading light of the code-breaking team that allowed Britain to survive the darker days of WWII with an intact navy, and as the originator of the famous Turing test of machine intelligence. See Andrew Hodges, *Alan Turing: The Enigma* (New York: Simon and Schuster, 1984) for an excellent biography.
which the listener is asked to judge the truthfulness of the statement uttered by the Cretan: "All Cretans are liars." Such statements lead to a dizzying spiral of self-reference and, as Gödel formally demonstrated, are logically "undecidable." Gödel's contribution was to press the point that all logical systems make possible such self-referential statements and, specifically and distressingly, that arithmetic was one such system. Systems which make possible statements that they cannot solve are known as "incomplete"—and no incomplete system could be expected to adequately model the world.

This was a real blow to the Hilbert project, and various protagonists in that play reacted in different ways. Chiefly, an attempt was made to rule out self-reference. This was the basis of Russell's famous theory of logical types, which simply issued a rule against particular sorts of reference. In the case of the Cretan liar, for instance, some the confusion was deemed to result from a poor linguistic habit of referring to the class of all Cretans by the same label as was used to describe a particular member of the class. More fundamentally, all reference of this sort was deemed out of bounds precisely because it led to nonsensical "undecidable" propositions. Russell's defense, however, was only a way station on the road to salvaging the hope for a more rigorous, rational, and ultimately mathematized philosophy. To make Russell's rule solution for the general case, you had to have a general way of deciding what were and were not decidable propositions.

Turing bent his talents toward achieving this end. If he could find a general procedure for determining which statements were and were not undecidable, he could, theoretically, salvage the project. In a nutshell, if there was a procedure for deciding which propositions were not decidable, then they could be safely segregated from "real" mathematics. If they were not decidable the attempt to make a complete logical description of the world must fail. A logical
formalism would be either too weak or incomplete. In exploring this hope he
developed an abstraction of computation which reduced logical procedures to
its essentials. His abstraction consisted of two states (on/off, true/false), the
ability to move and read a serial tape, and the ability to replace the first symbol
with the other symbol. This was all that was needed to compute what was
computable.

This abstraction served as the concrete avenue to explore the consequences
of Gödel’s demonstration. The general question was, “Can we, through the
offices of logic, determine which propositions are and which are not
decidable?” The more specific form it took on Turing’s dream machine—a
machine in which any formally logical system could be emulated—was to ask if
there existed a system implementable on such a machine which would
determine which propositions were or were not decidable. Crucially for the
story which we want to tell, this imaginary machine had no merely physical
limitations; it had infinite memory and infinite time. Turing showed that a
logical system for deciding logical propositions was a logical system like any
other and that it fell prey to the limits Gödel had intimated. This took the
particular form of not being able to decide at which point to stop a
computation. If a theorem is undecidable, a calculation set up to solve it will go
on forever. If, on the other hand, it is merely difficult, it can go on for an
arbitrarily long time. There is no principled way to tell the difference

168 Strictly logically, of course, this is all about possibility; even in the best case it might be
empirically true that all logical formalism might be both incomplete and too weak. It is a
measure of the strength of the analytico-referential discourse in these communities that this
possibility, eminently reasonable for Toulmin’s humanists, is seldom actively considered.

169 Philip J. Davis and Reuben Hersh, Descartes’ Dream, The world according to mathematics
(Boston: Houghton Mifflin, 1986), 139-141.
There is no way to secure the axiomatic basis of logical systems, and therefore no certainty in the conclusions that are drawn from them, or, assuming with the discourse that the world is self-consistent, to map a logical system accurately onto the world. Descartes’ dream was only a fever-dream.

This broader story, as important as it may be for our larger concerns, also forms a backdrop for understanding the emergence of time and space constraints as a central feature in computational theory. Turing deployed his universal machine in the context of demonstrating the limits of logic. Computer science as a practical field (as opposed to computational theory) took relatively little note of this result but found great value in the abstract formulation of computation that Turing put forward while showing these limits. Turing’s “tape” became computer memory, his two symbols became off/on computer gates and, with the addition of moving the input inside the computer in the form of a program which controlled the series of inputs, this image of computation became the von Neuman serial computer that we are familiar with today, the foundation of a huge practical enterprise supporting hardware and software producers on a scale unimaginable in the 1950’s. Time and space constraints, throwaway items in Turing’s strictly theoretical construct, have become central issues in the new science of computing and in its computationally oriented offspring, artificial intelligence and cognitive science.

The practical issue can be concisely stated and is the foundation for the software industry: given the hardware that is available, is it possible to do a given task in a reasonable amount of time? If so, how?

The simplest sorts of human cognitive capacities such as perception, for all practical purposes, would be impossible to implement if the brain were a serial,  

---

digital, site-addressable computer in the von Neuman tradition. The two classic features of practical computation, time and space, can be used to demonstrate this problem.

Time is strictly a limited commodity. You can’t take all day to decide whether that thing coming at you is or is not a car. The basic speed of the computer “gates” in the brain, the synaptic junctures, and the propagation rate of electrical potentials along the neural axon, are very slow by silicon standards: neurons work on a millisecond basis while computer gates work on a nanosecond basis. This is slow enough that the observed speed of a broad class of the simplest human capacities including recognition and motor skills must be done in 100 or fewer “cycles.” A lot can be accomplished in 100 cycles by a parallel processor, but in a serial machine each instruction must wait in line for the previous instruction to execute. The typical recognition program takes, at a minimum, tens of thousands of cycles. This goes a long way toward explaining, given the disparity in basic gate speed, the fact that tasks involving pattern recognition can be completed in milliseconds by the human brain but take hours to be completed by a computer. At a minimum we need parallelism; the basic serial model is not plausible on the basis of time alone.

But this does not mean that we must necessarily discard a strictly logical structure on this basis alone. A properly segmented logical problem can be solved in parallel if there is an appropriate central executive to segment the problem and reassemble the parts. We have already discussed the problems with admitting such a “homunculus” into our explanations in chapter 2, but

---


here we are more concerned with showing that even if we were to grant the
possibility of such central control the homunculus will not allow us to escape
intractable computational difficulties.

Space constraints led us to finally discard the logicist position in regard to
the process which supports human thought. This follows from the simple
observation that there are not enough neurons or even synapses to encode all of
the recall humans can display at the available number of discrete sites. The
large ratio of sensory to cortical neurons, coupled with the observed fact that
virtually all of the brain is active during sensory input, eliminates the
possibility that there are any neurons left over to be dedicated to serving as
long-term memory sites. Thus the site address cannot be seen as a viable model
of the way the brain is organized. The only viable model is one which encodes
multiple memories over many separate units: distributed representation.

Distributed representation solves the space problem by increasing factorially
the amount of potential representation. Consider a simple four-object matrix: if
each “site” is a single address, there is room for four representations. If all
possible combinations of off/on are considered, you have sixteen. When one
realizes that each neuron in the cortex is said to average 2,000 synapses\textsuperscript{173} and
that the Purkinjie cells—implicated in information processing—have as many
as 100,000\textsuperscript{,174} the impact of this progression in the expansion of address space is
quite staggering. Distributed representation—at least in the forms which
purchase the requisite degree of storage space\textsuperscript{175}—is the bane of logicist

\begin{flushleft}

\textsuperscript{174} T. Kohonen, \textit{Associative Memory: A System-Theoretical Approach} (Berlin: Springer, 1977).

\textsuperscript{175} Noninteractive distributed memory is possible, but only at the price of making each
discrete pattern completely separate. This allows many fewer units of storage.
\end{flushleft}
systems because of the nondiscrete nature of representation it allows. The law of the excluded middle fails and categories are no longer cleanly separable.

Simply put, given what we know of both the brain and computation, the abilities that humans show are not plausibly the result of a strictly logical architecture.

Connectionism's Inadequacies as a Theory of Instruction

This chapter has been largely devoted to explicating connectionism and exploring its plausibility. Largely, it has been contrasted with the dominant logicist account made most explicit in branches of cognitive psychology but deriving from broader discursive structures. A case has been made for viewing connectionism as plausible against this particular backdrop. Objections made from within this dominant tradition have been discussed as they arose in the exposition.

But these are not the only objections possible. Connectionism has been explored with an eye toward its possible utility in informing an educationally useful theory of learning. But adopting a connectionist stance toward learning leads to a very different understanding of what learning might mean. Generally learning theories have not been theories which concerned themselves, as connectionism does, with the material process of change of which learning is actually comprised. This has been possible because of the strong discursive assumption that this was already and almost trivially known. Thought, like the world and language, was assumed to be governed by the forms of logic.

Theories of learning, by and large, have actually been theories of knowledge acquisition. They are about the conditions under which particular atoms of facts are acquired. Some of these theories focused on developmental readiness and others looked at cultural conditioned factors. In education, theories of learning
styles, classroom conditions, and examinations of teaching methodology echoed the larger emphasis on determining the preconditions of knowledge acquisition. There has been a void in the space where learning, understood as a material change, actually occurs. It has simply been assumed that learning occurs naturally under the correct conditions and that the task was to determine and reproduce those conditions. Insofar as theories of knowledge acquisition speculate about the nature of thought and learning, they fall back on the forms of logic and the sorts of factual objects that logic requires. Innatist or constructivist, individualistic or social, child-centered or disciplinary-based, situated or psychologistic, the traditional arguments have been about knowledge: the degree to which it is acquired, whether it is best learned in the company of others or alone, whether the most appropriate forms are disciplinary, and where it is located. It is not by accident that the classic curriculum question is, “What knowledge is of the most worth?”

A connectionist approach replaces the default assumption that learning takes the form of logic with the position that learning is an emergent property of certain kinds of network architectures.

But if connectionism lends us a more robust theoretical position concerning learning as a material process, a closer examination reveals that this valuable advance is inadequate to support a theory of instruction.

The problem here is that understanding how learning occurs does not tell us what is learned. Connectionism, unlike theories of knowledge acquisition, does not make a priori assumptions concerning the nature of the object acquired and does not assume that knowledge exists prior to its deployment. Simply adopting a connectionist position regarding learning, however plausible that position may be when viewed in contrast to its alternatives, does not put us in a position to answer the questions that educators need to answer in order to serve
their students' needs. We need to know what knowledge is and how it exists outside the person in order to use the insights of connectionist learning. With situated knowledge as well as connectionist learning, we may be able to understand not only how the child learns to recognize her mother but also the sorts of regularities in the child's interaction with the world that can constitute "motherness."

In chapter 4 we will turn to a possible candidate for a companion theory of knowledge which may help us gain a handle on these problems: situated cognition. Situated cognition will be represented as a theory of knowledge which helps fill the gap created by adopting a connectionist approach toward learning. Equipped with these tools, we will in later chapters turn toward an exploration of research and instructional design which is suggested by the intersection of connectionist learning and situated knowledge.
CHAPTER 4

Connectionism and Theories of Practice

As it turns out, the answer to the question of how many patterns the machine can learn has as much to do with the structures of the world as it does to do with the abilities of the machine.

—Jim Jubak, 
_In the Image of the Brain_, 1992, p.58

Knowledge, as we usually conceive of it, is difficult to locate in connectionist networks. An understanding of connectionist networks may support a theory of learning but not, directly, a theory of knowledge. Connectionist theorizing constrains our approach to the issue of knowledge by proposing a surprisingly different conception of representation. Representation is usually understood as a technical matter central to understanding how we can be said to possess knowledge but on my account is better regarded as a position adopted in consequence of a theory of learning. By proposing a particular way of understanding how representations are learned and how that learning shapes the qualities that representations can possess, connectionist theorization constrains our consideration of knowledge theories to ones that can accommodate such representation.

The process of category formation which connectionists propose would radically alter our understanding of representation were it to be accepted.

Neither the traditional approach to categories as involving the common

---

176 In fact this has been a way to criticize connectionist networks as not being a useful model of cognition because it does not have proper symbols. See: Jerry A. Fodor and Zenon W. Pylyshyn, “Connectionism and Cognitive Architecture: A critical analysis,” _Cognition_ 28 (March, 1988): 3-71.
possession of necessary and sufficient features nor the strong dissenting view that categories are organized around prototypes is in full accord with the implications of a connectionist position. This is no small matter; what is at stake is what our culture takes to be the basic organization of knowledge and how knowledge is acquired. Nothing could be more fundamental to the practice of education.

The Representation Debate

The question of just what constitutes a representation has loomed large as an issue in cognitive science and philosophy. By and large the field has divided between those who understand representations as symbols and those who claim that representations are better understood as images. The strongest argument on the part of those who endorse the symbolic position is that symbols and the syntactical relationships that connect them are the only way that the generativity of human thought can be explained. Their position gains strength by being implemented by artificial intelligence workers in programs—programs that actually run and lend the verisimilitude of presence to the position. This position is in part inspired by the Chomskyian analysis of human language and is the basis for the language of thought (LOT) conjecture. In the Chomskyian account, thought is structured as language is structured, with a combinatorial syntax which enables the individual to generate the unlimited possibilities of productive speech. Viewed against the background presented in chapter 2, this can be seen as an explicit formulation of the foundational

discursive claim that thought, language, and logic share the same form. The LOT proposal differs from the way the issue has generally been depicted in this dissertation by emphasizing language and the combinatory generativity of language as foundational instead of logical form. Still, the LOT proposal only emphasizes an alternate leg of the unity between thought, language, and logic, and the idea of syntactic generativity relies on the formal, recursive relation between objects of knowledge rather than any semantic or meaning-based conception of the richness of language.

The opposing camp endorses the thesis that memory consists of images. These representations refer to a much larger scale of objects and events than those postulated by the proponents of the symbolic alternative. They can consist of whole events, temporal sequences of events, objects in the world, and configurations of objects. The strongest arguments for this alternative come from observations of the process of human recall. Human memory is clearly associative, and its ability to guide action in the world is predicated on this associative quality. Experimental findings concerning the priming effects of prior, associated experience buttress this approach. If one assumes that images are the basis of memory, associative recall is easily explained. The person who recalls one item of an event is assumed to also gain, for free so to speak, access to other items of the event.


179 What is generally poorly appreciated in the presentations of those that endorse such an alternative is the degree to which they too fall into the trap of postulating objects. By postulating objects at a large scale they avoid some of the absurdities of the symbolic position, but they do not finally avoid the issue of how different objects can be related associatively.
Those endorsing the imagistic alternative are conscious of the way their proposals challenge the traditional view of cognition. They explicitly reject the traditional understanding of categories that proposes that we know a category by its primitive features—the things that members of a category have in common. They point out that this position is taken by both the folk theory and the dominant technical theory of representation. That the common understanding and the technical formulation are similar in this way can be explained on the grounds that both operate within the assumptions of current analytico-referential discourse. The imagists have believed that the only way to explain the observable characteristics of human recall is on the basis of a stored prototype or prototypes which constitutes each category. In contrast to the LOT proposal, we will call this the prototype theory of categorization.

**Representation in Connectionist Networks**

Connectionist approaches to the process of category formation, and the implications of this process for representation, are generally endorsed by those embracing the prototype theories and vehemently rejected by those who endorse the language of thought hypothesis. Connectionist approaches

---


actually differ radically from either of the two earlier approaches, but the commitment of prototype theorists is to a body of empirical findings which leads them to infer an imagistic basis for memory. They are able to treat connectionism as an alternate way to account for the phenomena that are at the center of their concerns. LOT theorists, on the other hand, are centrally committed to a formal mechanism, syntax, which is incompatible with connectionism's process of distributed representation.

In chapter 3 I discussed distributed representation as a form of representation which, by blurring the distinction between content and structure, enabled the pattern completion capacities of networks. These pattern completion qualities of categories are associated with distributed representation's ability to support categorization, generalization, and inference in ways which are learning-like. Distributed representation was also noted to have the computationally important capacity to store many more representations over a set number of nodes than alternative, localist schemes of memory.

To briefly recall this latter issue: distributed representation is computationally plausible in part because it is able to store many more representations than there are individual neurons or even individual synaptic connections. Distributed representations show this characteristic because they are not objects, but relational patterns. Each representation is a pattern of relationships spread over many nodes in the network. Thus if we assume that each node in the network has two states (off and on), a localist representation could encode the presence of only four "pieces" of information, whereas a fully distributed network could carry sixteen different possible configurations of off/on patterns. The larger the number of participating nodes, the greater the advantage in storage space networks gain.
This quality of network representation is known as superimposition: many representations can be imposed on the same set of nodes in a network. Recall here the Sharks network, which encoded a surprising amount of information in a useful associative array of 72 nodes. The associative qualities of this array arose as a direct result of superimposition. Superimposed memory of this sort is interactive. Each memory is, in effect, adjusted away from the set of strengths that would be the characteristic result of the input if there were no other memories, and toward the aggregate of memories already stored in the net. The new memory must "compromise." As the network moves beyond a minimum number of patterns that it can store without interaction, the characteristics of the process of storing these memories leads some to converge around the same patterns of activation. Similar patterns of input will result in very similar patterns of configuration in the inner, hidden units which are the analogs of memory. Different but similar input patterns will elicit the same output array, and we will be inclined to say that the network has categorized these as the same. This is what underlies category formation (and inference) in connectionists networks. This phenomena has been analyzed in terms of the interaction of nodes rather than at the level of the representation itself as I have done, and in that context it is usually referred to as a subsymbolic analysis.\footnote{Smolensky's influential article popularized the use of the term "subsymbolic" though others used the term before him. Paul Smolensky, "On the Proper Treatment of Connectionism," \textit{Behavioral and Brain Sciences} 11 (March, 1988): 1-74. For the reader who wishes to follow out the implications of interactive representations, the literature is likely to refer to this phenomena as "subsymbolic" or "microfeatures" (Rumelhart and Clark respectively). See: James L. McClelland and David E. Rumelhart, "A Distributed Model of Human Learning and Memory," in \textit{Parallel Distributed Processing, Volume 2: Psychological and Biological Models}, ed. David E. Rumelhart, James L. McClelland, and the PDP Research Group (MIT Press: Cambridge, MA, 1986): 170-215; Andy Clark, \textit{Microcognition} (Cambridge, MA: MIT Press, 1989), 108-114. This work places a particular emphasis on the interactive pattern of activation that constitutes a connectionist representation and for this reason prefers the term "superimposition" to those which focus on the role of constituent parts of the representation. It also avoids defining a central term using variants of the terms which it attempts to displace. Hofstadter uses the term "active symbols" to describe a related conception. Douglas R.}
Recall that it took the Sharks network longer to distinguish between individuals who shared many of the same characteristics and that the network, at its limits, was likely to incorrectly infer that a person in their forties was in their thirties on the basis of the preponderance of those in their thirties who shared the other characteristics of the individual sought. We saw in the Sharks example that such confusion could actually be an asset, as the network was able to generate plausible responses which had not been explicitly coded into the network. It would, for instance, settle on the name of a Shark who was most “typical” of all Sharks if its only input was the clue that the person referred to was a Shark.

As has already been emphasized, the associative qualities that give distributed representation its power are also the qualities that make it unsuitable as the representation that formal logic or syntax can use; the categories that constitute objects of knowledge are not separate. Contra Aristotle, connectionist models imply that humans cannot eliminate the hard-to-categorize examples in the middle—and logical operations require separation. But the differences go deeper than just this. Most crucially connectionist representation is indissolubly tied to current experience, real or imagined. Rumelhart and Norman say, “Information is better thought of as

---


185 Even if we could eliminate the middle we would not want to. Connectionists claim that eliminating difficult examples retards learning. If this is true, the idea that curriculum is content simplified for instructional purposes is a dangerous one. Chapters 3 and 4 explore the implications of using non-central examples in the context of materials design.

'evoked' than 'found.' This can be made clearer by referring to the Sharks example. The network's structure contains many possible patterns of activation corresponding to possible output arrays by virtue of the patterns of relationships between nodes. But no information is evoked without external input. That is, unless we turn on the "Sharks" input node and let its effect propagate through the network over several iterations, we do not get the usable information concerning the identity of the most typical Shark at the output layer. This is not just the trivial case of not getting an answer until a question is asked. The configuration of internal weights and propagated influences that is expressed at the output layer as, say, "Neal," does not actually exist until the question is posed, and the differences this posing causes propagates through the network. The information is simply not yet there to be "found" even though it is latent in the structure of the network. Further, as we have already seen, the particular configuration the networks settles into is also dependent upon the sequence of input, not just its presence or absence of certain features.

If we tie this to the human brain and the manifold sensory inputs which constitute its world, we can see human categories as a strange combination of stability and plasticity. In this way of approaching the problem all recognition is category re-cognition. The child's recognition of her mother is never recognition of an object as such, but is the re-cognition of the category Mother. The child has never before seen exactly the same face that she sees tonight. But she "recognizes" her mother by re-calling the category—*and she knows the appropriate response as part of that recognition*. She holds out her arms and is softly gathered in. This perspective indicates that every recall is the contingent

---

creation of the moment and necessarily differs from all other instances in which the child has recalled her mother. It is disturbingly malleable. But such representation is also, because it is plastic and not brittle as are logical programs, very stable. Mother in all her guises—and maybe some other adult women who act appropriately—gets the same treatment, and most often the child gets the result she desires.

Reflection on the characteristics of connectionist representation reveals why some researchers have found it so appealing in the context of the ongoing dispute about representation. Connectionist representations, for all their strangeness to one steeped in the Western tradition, have some very desirable characteristics. For those whose interests lie in a fascination with the generative qualities of human speech and thought, connectionist research into the emergent properties of networks reveals a different way to achieve the generativity formerly attainable only through the iterative processes of deductive logic. To this group connectionism also offers a glimpse of how one could go beyond the axiomatic mindset that leads some to insist that much of our linguistic and cognitive abilities are innate and that our development is essentially deduced. To artificial intelligence workers who demand that the theory be implementable, that it actually run and have presence in the world,

---


189 In linguistics the “deductive” stance is the classic Chomskyian position. Deep structure (grammar) is innate and sets up the range of possibilities—the particular language that is “acquired” is only one deduction. In the psychology of thought the best known example is in the Fodor-Piaget debate, ironically between two logicists, where Fodor insists that nothing develops from experience, that experience only triggers development. Piaget, while arguing that logic is what develops, continues to insist on interaction with the presumptively logically structured world as the path to that end. See: Massimo Piatelli-Palmarini, ed., Language and Learning, The Debate Between Jean Piaget and Noam Chomsky (Cambridge: Harvard University Press, 1980).
connectionist representation also offers an alternative way of organizing their efforts. The connectionist alternative offers a more flexible means of implementing many characteristics important to their projects.

Similarly, those who have endorsed the imagery as an approach to representation now have a way to explain how the associative phenomena that interest them might be realized in the material world. The superimposition which is characteristic of connectionist category formation and reformation,\(^{190}\) with its inherent generalization and inference, is a valuable tool in understanding how experience can be encoded nonpropositionally and still be generative and accessible. But connectionist representation does not simply endorse imagistic intuitions. It throws real doubt on the doctrine of prototypes. It also entails not regarding objects in the world as given primitives—a stance which those working with images as a metaphor for representation have generally accepted.

As valuable as connectionist category formation may seem for solving the problems posed by representation research, it radically destabilizes the concept of representation. Connectionists hold that any particular representation is plastic and shifts with both the spatial and, more importantly, the temporal context within which it is deployed. I have emphasized that this malleable representation has large advantages in overcoming problems that more traditional approaches founder on, such as the brittleness problem examined in chapter 3. But it creates new problems of its own. Most notably, postulating such plasticity, and indeed, making this quality explanatorily central, shifts us

\(^{190}\) Ideas of re-cognition and re-formation recall suggestively the Deweyian approach to cognition: “reconstruction.” This is a particularly cogent idea placed against the current field in education in which “constructivist” ideas are important organizers of new approaches to teaching methodologies. John Dewey, “My Pedagogic Creed,” in John Dewey on Education, ed. Reginald D. Archambault (Chicago: University of Chicago Press, 1964), 427-439.
from attempting to explaining the apparent flexibility of knowledge to needing to explain the apparent stability of knowledge.

One tactic would be to make the claim that knowledge is only apparently stable in the sense that is usually claimed. The standard story concerning the recognition of the mother's face is that the child possesses a stable representation and that the similarities between the mother-of-the-instance and that representation render the conclusion that this particular person is Mother trivial. As we have already seen, attempting to actually implement this triviality has led those working within that tradition to the profound and apparently insoluble problem of perception. It turns out to be a very difficult problem, one that is not trivial at all but is central to the capacities that we understand as most human. The competing connectionist story involves representation which is "custom-created" and which takes into account the ongoing activity and condition of the child along with her location in space and time. The connectionist representation is a response to the child's situation. The child is always already in a situation which shapes her ongoing perceptual/cognitive framework. It may very well be that the child has a history that leads her to anticipate the perception of her mother in a context in which she experiences an uncomfortably cool breeze. She has quite reliable experienced the presence of her mother whenever she has encountered such a situation in the past. The perception of the mother in such a moment is colored by this expectation and by accompanying expectations for being held and made warm. Considered from this angle, representations were only apparently stable, and the problem of their newfound plasticity is not a problem at all, but a solution.

But this will not, in the end, be enough for educators. What educators are expected to teach, and students to learn, are particularly the things which are not learned in an immediately useful context. Educators teach that knowledge
which the community values but whose context of use is dispersed or in the future. Thus reading or arithmetic is generally learned to a fluid degree of expertise only in schools; the context in which they are encountered in the child's everyday world is too dispersed to typically foster the engagement that allows for fluid accomplishment. Other subjects are taught in order to be useful in a projected but not yet present future.

As seekers of a distinctively educational approach to the problems of learning, an approach which can inform the processes of learning that take place in institutions which we call schools, we need to look further than the relatively easy answers that accompany the sorts of contextually appropriate recognition that takes place in the context of immediate use. The child's story, as well as similar stories told by the advocates of apprenticeship, point us toward what we need to consider to understand learning in an educational context, but they do not, in themselves, constitute such understanding.

191 I take whole language instruction to be one attempt to close that gap by focusing on contextually appropriate instruction.


193 They are similar in that the context of learning is the context of use—a situation which is exactly what distinguishes schooling from apprenticeship.
Representation Outside Connectionist Networks

Connectionist representation's dependence on the world for the stability of its representations is not limited to the immediate context of use. In the connectionist recounting that we will explore here, there remains a need to understand something analogous to what we have been discussing as symbols. The connectionist account postulates that symbols exist in the world, that the world is its own representation. In this interpretation the mother is herself the source of the stability in representation that the child experiences. The mother reliably acts as she has been known to act. These actions, her appearance, and the contexts in which she appears are all the basis for the stability that cannot be located in connectionist architecture itself.

Connectionists, while not focusing on this issue, have noted the problem and begun the rough outline of what would be necessary to move in this direction. Most cogently for the purposes that this chapter pursues are suggestions made by Rumelhart et al. which rely in part on the work of Vygotsky—a theorist whose social approach has also influenced situated cognition. Rumelhart et al. suggest, “People seem to have three essential abilities which together allow them to come to logical conclusions without being logical.” They say that people that “are especially good at pattern matching . . . are good at modeling the world . . . [and] are good at manipulating our environment.” Their


195 Dreyfus discusses this idea as “concrete representation.” See: Dreyfus, What Computers Still Can’t Do, 61.


197 Rumelhart, Smolensky, McClelland, and Hinton, “Schemata and Sequential Thought Processes,” 44.
discussion also suggests a special role both for language and for what they call external representations.

"Roughly speaking," they suggest, "the view is this: We are good at 'perceiving' answers to our problems." That is, the apparatus that we use to locate things in the world (perceptual pattern matching), to anticipate changes in our situation (modeling), and to act (manipulating our environment) can be used to simulate what we usually understand as logical reasoning.

Three-digit multiplication is one of the tasks taught to students which can be seen as one of the "displaced knowledges" that we concern ourselves with in this chapter. Consider the way that this is actually taught in American schools. Two three-digit numbers are placed in a column and a repetitive pattern of manipulating the number signs as individuals, usually with no reference to their meaning within the larger numbers, is practiced. The last number on the bottom is paired separately with each separate digit on the top and multiplied. Then the second digit is handled in the same way but the answer is displaced one unit to the left. The third numeral repeats this pattern. Then the newly produced numerals are added together, preserving their leftward displacement. The final addition is considered the answer to the multiplication problem. This repetitive process builds on earlier practices of multiplying multiple digits by one digit, which in turn is built on the practices involved in multiplying two one-digit numbers which are placed one beneath the other, with the multiple digit answer displaced to the left. In the context of a mathematics lesson, multiplying single digits by each other and displacing the answer leftward is a well-established habit. Seeing numbers placed one on top of the other, preceded by a large "X" and with a suggestive line placed beneath, is perceived as an

---

instance occasioning the activity described. To object that this is not real understanding is to miss the narrower point made here: by perceiving a problem as one which calls into play specific habituated practices, people turn a forbiddingly difficult problem in abstract reasoning into a tractable task. It is worth noting that it is only in the child who we have taught to do this explicitly that we are concerned that there is no “understanding” of the event. When adults engage in this activity in the course of daily life, we are not tempted to assume that they are acting without understanding. Very few of us can multiply 334 by 783 without resorting to pen and pencil or an imagined analog—a model—of this activity. The qualities of connectionist representation, pattern recognition, sensitivity to immediate context and history, and inclusion of habituated activity, are crucial to making this rendering fully sensible. Such an approach makes the world and our manipulation of it a part of our cognitive apparatus. Rumelhart et al. are worth quoting at length on this point:

These dual skills of manipulating the environment and processing the environment we have created allow us to reduce very complex problems to a series of very simple ones. This ability allows us to deal with problems which are otherwise impossible. This is real symbol processing and, we are beginning to think, the primary symbol processing that we are able to do. Indeed, on this view the external environment becomes a key extension to our mind.199

A key to understanding the power of this approach is the realization that our ability to anticipate changes in our situation, effectively to model the world, makes it relatively easy to see how one could internalize these practices. So far in this work I have tended to emphasize the specifically representational qualities of connectionist networks and recounted both historical and intradisciplinary debates in which connectionist theorizing can be seen as an act

opposing certain elements of these traditions. In this I follow the literature and the engagement of the participants, but this approach may actually do an injustice to the intuition which underlies connectionist models. By tacitly accepting the debate's assumption which separates representation and action, I may be misleading the reader (as may be the emerging literature) about the actual nature of the connectionist relaxation patterns that have been held to be connectionist analogs of representations. The tools that connectionists have at hand are computers, and computers have been constructed to be and are interpreted as passive representers of the world, just as the canonical tradition would suggest. But this is not the case with the biological beings from which connectionism draws its more interesting inspiration. For such beings relaxation states are not only the basis for representation but are also the physical basis for the ensuing activity in the world. A connectionist model does not provide any in-principle basis for separating the cognitive activity of representation from that of action. Following this line of reasoning, it is fair to say that the activity of perceiving the Mother includes reaching out in certain circumstances, such as when the child is uncomfortably cool. Similarly, the presentation and appropriate perception of the pattern of a three-digit multiplication problem can reasonably be construed as including the anticipation of the activity described above. This is an important advance because it eliminates a persistent conceptual bottleneck in our thinking about thinking. A large part of the tradition has been spent on trying to find a principled way to move from experience to representation and then from representation to action. The embodied account presented here attempts to avoid the initial, mistaken, division. The act of perception is also an act of representation and representation appropriately includes a readiness for context-specific action. The same relaxation state may subserve all three of the
functions that we separate discursively. But this strong linkage of the world and connectionist networks\textsuperscript{200} carries a price—when interpreted within the analytico-referential discourse with which we are engaged, it brings into question the separability of the person acting from the context of action. For a culture which values the independent learning of the radically free individual above all other types of learning, this can be seen as a questionable advance. The role of independent learning is much reduced in this interpretation of our cognitive abilities.

Not surprisingly then, connectionist architectures have often been criticized for requiring a teacher\textsuperscript{201}. I have argued in chapter 2 that this critique is without much force in an educational context, where the presence of the teacher is not understood as a problem. In more profound related criticisms, connectionist networks are taken to task as being \textit{architecturally} passive in the face of experience.\textsuperscript{202} Both critiques stem from considering the qualities we are presently discussing as a disadvantage. In the more sophisticated of such accounts, the discussant notes that without input from outside the system, the network settles into a rigidly stable state and would presumptively be incapable of activity. This actually understates the case. Connectionist models almost always include decay parameters, and certainly any model which has pretense to deeper biological plausibility does so. Decay refers to the gradual,

\begin{footnotesize}


\footnote{202}Such more sophisticated critique is most often leveled from within connectionism by workers active in the field. A nice example is: Rumelhart, Smolensky, McClelland, and Hinton, “Schemata and Sequential Thought Processes,” 39-40.
\end{footnotesize}
proportionate reduction in the strength of connections between nodes that takes place as part of the calculations which determine connection strengths during each iteration. In such models a network which operates continuously without new input does not reach a stable state representing its response to the last input cycle and simply stay frozen there, as these critiques would suggest. The effect of decay is to wipe out the organization of differences in weights which was the network's analog to memory. It "forgets," and this has consequences which are graver than mere passivity: whatever has been learned from experience is wiped out.

So connectionist architectures buy a partial solution to the perennial philosophical and primal educational problem of separating the person from the world, but at the dual price of appearing to move much of what we considered rational thought out into the world and also making continued engagement with the world the price of maintaining what is learned.

**Abstraction, Embodiment and Displaced Knowledge**

We have been slowly building a rough outline of an alternate account for the cognitive abilities that support the competence that we hope our students will gain. In chapter 3 we discussed very basic emergent properties of connectionist networks in relation to learning and began an account of what representation, abstraction, and category formation might be taken to mean under an interpretation based on network theorizing and experimentation. There we discovered that connectionist representations could not, alone, be expected to support the full weight of this enterprise, since they did not indicate a way to

---

203 Most often networks operate in a strange sort of analog to time wherein "time," understood as successive cycles of iteration accompanied by relaxation to a global state, occurs only when input is present.
decide which of the many regularities in the world would be selected as the basis for category formation, which is connectionism’s way of understanding the objects that the older tradition gives us. Out of concern for educational relevance, we directed our attention to the displaced knowledge that schools usually emphasize, passing lightly over the physical regularities that some have made central to their theorizing. Here we have pursued the thesis that socially ordered practices constitute the regularities that we seek. In doing so we have reached deeper into the properties of connectionist representation, emphasizing particularly the instability that is the darker side of connectionism’s valued flexibility. The particular forms of instability exhibited by connectionism representations have led us to consider the possibility that much of what we call symbol processing is actually manipulation of an environment which we have created for such purposes and brought into our social and individual worlds as practices. Three-digit multiplication has been briefly explored as an example. Our embodied activity in the world has been seen as a key to understanding how the stability of symbols can be sustained in the face of the constitutional instability of connectionist architectures.

We have perhaps pushed connectionist insights as far as they can take us at this point. Establishing the need for a socially organized theory of practices is the most that a connectionist approach to cognition can aspire to. Connectionist theorizing continues to constrain our approach to the material that follows, but it cannot offer the same positive guidance that it did when we were discussing the relationship between representation and learning.

---

204 Piaget (object acquisition) and Gibson (affordances) are the two most prominent in this regard. See: Barbel Inhelder and Jean Piaget, The Early Growth of Logic in the Child, Classification and Seriation (New York: Harper and Row, 1964) and J. J. Gibson, The Senses Considered as Conceptual Systems (Boston: Houghton Mifflin, 1966).
In what follows I will rely more explicitly on social theorists whose frameworks seem especially well-suited for the enterprise and context of this work. The path we have followed to this point has led to seeing *practices* (which are learned as patterns and integrated into both activity and perception by connectionist principles of representation) as the key to understanding the sorts of displaced knowledge which the community values and the schools are established to teach. Knowledge built on such a basis could be expected to exhibit certain peculiarities. It would, for instance, be context dependent in particular ways. Thus we would expect that simply learning the practice of multiplying three-column figures would not transfer to a situation in which the problem is laid out horizontally. Such a presentation would truly be a different problem to the uninitiated student. Staying at the level of practices, we can see how a new practice made up of specific patterns of multiplication and addition could be mapped onto the new, horizontally organized problem. One could especially easily imagine a practice which transforms the horizontal problem into a vertical one. But none of this activity produces the kind of understanding which the school generally strives to achieve. Again, it will be too easy to simply say that the abstract, "principles of mathematics" knowledge that schools value does not and never did exist in the forms that the tradition gives us. While both connectionist and situated cognition researchers might for their different research-based reasons endorse such a suggestion, this conclusion will not be sufficient to support an educational theory that deals adequately with the issue. That conclusion, while it may be well supported, fails to deal with the existence and utility of what has been called abstract knowledge. Connectionism and practice theories do not deny the *existence* of a more broadly powerful way of knowing. They deny that it takes the logical form that the tradition describes, and that it is learned in the manner that would be
appropriate to that form. The community's evident desire that students learn a powerful, generalizable form of knowledge is not somehow obviated by new research. More powerful understandings of this sort do exist and seem as inordinately valuable as the community apparently understands them to be. The task before educators is not to dismiss such abstract, displaced knowledge, but to explain it on a different, more educationally adequate basis.205

This will be clearer if we return to the concrete example of three digit multiplication that we examined earlier. Educators generally agree that the mechanical, rote production of the answer, even the correct answer, which is produced by the practice I described does not constitute the sort of learning that is finally sought. Coming to understand the role of practices in the production of the student's competent response is an advance. But it is not all that we wish to understand.

As math teachers will testify, simply labeling the various positions that the numerals hold as "place value" does not result in the broader understanding that is sought. The student may be able to label the "tens place" correctly without being able to use this concept in any other situation. Given the story that connectionism tells, this is not surprising. The student has learned to label a part of a practice—and does not possess what we would call an abstract concept. The label signifies only within this practice.

It is just at this point that our usual apparatus for explaining learning and knowledge most obviously fails us. A student who can repeat the definition and apply it appropriately is believed to understand. He or she is said to.

possess the concept but apparently fails to "transfer" or "apply" it to new situations. This problem is generally understood as a failure to access the conceptual object stored in memory. Within the standard framework, a framework which posits objects (defined in terms of necessary and sufficient features) as the atoms in a logical calculus of thought, there is seemingly little that can be done. In this story the apparent context dependence of our retrieval mechanism is simply a source of failure. The most likely route to more accessible recall is to somehow strengthen the memory—a task to which repetition and practice are thought to be well suited.

In the view we are developing here the generally powerful abstraction that is sought is not inaccessible, it simply does not yet exist, regardless of the ability to repeat a definition or label a portion of a practice. There is therefore no failure to transfer or apply the concept and context dependence is not the source of this nonexistent failure. But, granting that powerful generalizations do exist—even if the traditional definition is inadequate—how are these generalizations achieved? It is at this point that the developing position that we take here has something radically different to say about the practice of teaching. It would suggest that this generalization is but another specie of category formation and is learned in the same way as other categories are learned, through the socially organized practices which led us to recognize difference as the same. Recall the child and her mother. The mother is different at every new

206 The prominence of application problems within the traditional research literature of education implicitly supports this claim.

207 Situated Cognition as an educational movement is motivated by the research based on this conception, which consistently finds context to be the most reliable way to predict whether the competent behavior with which the researchers are interested will be produced. See: John A. St. Julien, “Explaining Learning: The Research Trajectory of Situated Cognition and the Implications of Connectionism,” Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA, 20-24 April 1992.
perception of her and yet the child learns to recognize each as her mother and
to act appropriately. The pedagogical issue suggested by this is one level of
abstraction higher. The child has the label of mother (admittedly
prelinguistically) but does not have the next level of abstraction which we may
suggest is that of parent. Similarly, the label "tens place" located in the context
of a single practice is not the abstraction which educators seek. The labels
"mother" and "father" as set against other adults define for the developing
child the category of parent. But, importantly, there are not really labels which
pick out the mother or the father against the broader background of the
prelinguistic child's life is.

At this point we return to the intuition of Rumelhart et al. that their work is
consistent with Vygotsky's understanding of the relationship between concept
formation and language Vygotsky argued in Thought and Language.208 In
discussing their understanding of the way the same network which produces
perception and activity could also produce the anticipation of activity and
thereby effectively emulate a formal model of the world which they call mental
simulations, Rumelhart et al. remark:

There are a number of ways of controlling that sequence. One way
involves the running of "mental simulations." Another way involves
recycling linguistic inputs.... Suppose that the interpretation that led to
the production of the internal speech was much richer than the linguistic
forms could possibly suggest. Thus, the linguistic forms pick out aspects
of the entire interpretation to emphasize. Once this emphasis has been
taken place and the new input has been processed, the next state will be
strongly affected by the new words we chose to express our first idea.209

While this formulation captures Vygotsky's suggestion of a pivotal role for
language and particularly his presentation of the role of the formal definition of

208 See especially chapters 5 and 6: Lev Vygotsky, Thought and Language, trans. Alex

209 Vygotsky, Thought and Language, 44.
the word in scientific concept formation,\textsuperscript{210} it misses a central point in Vygotsky’s formulation: that this meaning is established intersubjectively through activity. The word does not merely limit the plenitude of meaning given by experience but also plays a creative role in directing attention within the world. Rumelhart’s formulation also misses the hermeneutic character of Vygotsky’s position regarding the word. In Vygotsky’s view the word’s meaning develops over time. Initially it is very concrete and may refer to what an adult would judge an inappropriately broad range of objects or activities. The child’s “Ma-ma” may refer to all moving warm things that she likes—including her father and the family cocker spaniel. The child’s experience and spontaneous categorization based on that experience take place with a socially structured framework. Calling her mother “Ma-ma” does not interrupt the process of being made warm, but calling her father by that term very likely will result in an interruption while he repeats “Da-da”—much to the child’s chilly annoyance. We may fantasize that the term “parent” can emerge only on the basis of the previously developed complex surrounding mother and father in which the word and the social practices associated with those words played a central role. “Parent” may be considered a tertiary development which would be impossible to achieve without the earlier separating out of mother and father through language and other social practices. The development of the concept parent, however, in turn lends a very new meaning to the earlier labels father and mother.\textsuperscript{211}

\textsuperscript{210} Vygotsky, \textit{Thought and Language}, 146-148.

\textsuperscript{211} This is far too schematic to actually be the case. The simple hierarchy suggested here is bound to be much more complexly configured in any real child’s life.
By analogy, what is needed to create a more powerful abstraction for the label “tens place” is another member of the same category, “tens place,” complete with appropriate distinctive practices which embody the relationships that more sophisticated adults would understand in terms of place value. This would allow an abstraction similar to that of “parent”—an abstraction which would alter the original meaning of the original label in the more abstract and powerful direction that is desired. What is wanted is another practice using place value and, crucially, the word label “place value” used by the teacher to isolate the particular relations that comprise the new abstract object of place value. The two different social practices associated with the single term “place value” would help to isolate for the student the intersubjectively standard meaning.212 One way to do this would be to teach a second practice for multiplying three-digit numbers but for these practices to be embedded in a problem presentation that is organized horizontally rather than vertically and whose mapping onto the already established method of doing such problems is centered around what the teacher understands as place value.213 In such an event the term “place value” can be disembedded from its context in a particular problem presentation and progress made toward a more powerfully general conception of the term.

212 The “generalization” captured through this process would be another categorization, another instance of “seeing difference as the same.” This is not a way of introducing “feature extraction” at a higher level.

213 In a remark which is interesting and suggestive for educators, Rumelhart, Smolensky, McClelland, and Hinton, note (p. 47) in “Schemata and Sequential Thought Processes,” that it appears to be extremely difficult to invent novel and useful new external representations, saying: “It may be that the process of inventing such representations is the highest human intellectual ability.” Heady stuff for the methods teacher who takes this task as a central element in his or her practice.
Such an approach would serve to “displace” the formerly concretized label “place value.” If we are to understand the sorts of “displaced” knowledges that education primarily deals with as learned in a connectionist fashion, it will be on the basis of this socially situated symbol/practice.

To engage in a bit of displaced abstraction of my own, I offer the following: the process that I have described is one which establishes a particular practice and a very local competence based on that practice prior to the attempt to teach the abstract meaning that is attributed to the practice. This is in notable contradistinction to the path that the tradition lays out for us. Generally we teach the “atoms” of meaning first and expect that drill and practice will set these meanings in the proper context. The activity of introducing a unit with rigorously defined vocabulary words is common across the curriculum and reflects our underlying assumptions about knowledge and its acquisition.

In such a view the tradition has simply inverted the actual order of learning. Learning a definition, whether of the term “place value” or “eukaryotic,” is simply not how meaning is established. This joins a dissenting tradition exemplified by the later Wittgenstein, Heideggerian phenomenology and hermeneutics which reject the claim that meaning is established through necessary and sufficient features (whether innate or induced) but that the meaning of a symbol is grounded in its history of use. The current account, informed by connectionist theories, grounds these claims in an understanding

---

214 It is reasonable to note that this is not really “concrete” in the sense that we usually think of the term that opposes it to abstract. On the conception put forward here, the distinction is one of degree rather than one of quality. That is, the term “place value” is still tied to a particular context and practice—we have simply enlarged the area and made the object more available through language.

of the material process which supports an appropriately associative form of representation. The Wittgensteinian observation that there simply are no necessary or sufficient characteristics which define the term "chair" as it is used in our language can be seen as a condemnation of language—a claim that language is not the tool it should be in that it does not accord with the presumed purity of thought and logic. But according to the understanding developed here the "ill-defined" nature of representation is a quality of all human signification—and language is only one area that exhibits these qualities of thought. Indeed, far from being a source of imprecision, language in part acts to constrain the perfusion of meaning that our individual experience leads us to. It is not a source of error but of necessary intersubjectively derived stability.

Again we are faced with a too-easy opportunity to simply declare the dominant tradition wrong. Education has long been divided between advocates of a formal approach to learning centered around the memorization of abstract knowledge and those that advocate an experiential approach to knowledge acquisition. Beneath a long and bitter dispute, both agreed, in general, that what was acquired were the classical objects of knowledge and the transformational rules that connected those objects. What is offered here is the possibility that the classical objects of knowledge simply do not exist. This alternate account offers in the place of necessary and sufficient features that form the essential basis for classical category formation the possibility that abstraction is done on the basis of our embodied practices. Multiple practices,

216 Wittgenstein, *Philosophical Investigations*, 216. Interestingly, this analysis of the chair as defined by its use was earlier used by Dewey, see: John Dewey, "Experience and Thinking," chap. in *Democracy and Education* (New York: Macmillan, 1916), 143.

217 From the progressive movement to the more recent back-to-basics movement, such an opposition has been a staple of educational discourse. For a Deweyian rejection of this dichotomy, see: John Dewey, "The Child and the Curriculum," in *John Dewey on Education*, ed. Reginald D. Archambault (Chicago: University of Chicago Press, 1964), 337-358.
mediated by the teacher’s discursive labeling of similarities between the patterns in particular practices, form the basis for increasingly abstract category formations, a basis for what we usually call concepts. This tempts us to reject the explicit teaching of definitions and rules which have been assumed to form the basis for competence.

Here, again we are restrained largely by our dedication to a distinctly educational enterprise. Even if we grant that definitions and rules (the educational expression of objects and laws) are not constitutive of our competence, it does not necessarily follow that definitions and rules are not one path toward that competence. Caution in concluding that definitions and rules are useless are suggested by the endorsements, albeit limited, of both Vygotsky and Dreyfus. The dissent of both of these thinkers from the logicist tradition informs this work and the value they find in rules is explored below.

Vygotsky, whose strong social emphasis is the basis of much of the blossoming interest in social constructivism and other socially grounded approaches to learning, endorsed, in chapter 6 of his book *Thought and Language*, an approach to learning what he called scientific concepts that was based on providing a definition prior to experience. Similarly, Dreyfus, whose phenomenologically inspired critiques of the received tradition have powerfully influenced this work as well as the ongoing reevaluation in artificial intelligence, has put forth a description of human expertise that begins with the teaching of explicit rules of engagement.

Accounting for the power of such approaches and the experience of teachers that methods based on definitions and rules can be useful is one of the implicit tasks of any work that hopes to offer a usefully different understanding of learning. How can the effectiveness of such regimes be accounted for if we give up the traditional story which has supported them? Having accounted for their
effectiveness, how does the account being developed here differ from or augment the alternate Vygotskian and phenomenological accounts? Is the difference such that education would be well served by preferring the interpretation based on connectionist representation and situated practices being developed here? To be able to say that the developing account should be preferred, one would have to conclude that this framework explains the phenomena more usefully. One way that an explanation might prove more useful would be if it yielded a more complete account which would make useful correctives to earlier descriptions.

**Vygotsky on the Word**

Vygotsky, in Chapters 5 and 6 of *Thought and Language*, projects one developmental path from poorly formed categories to concepts reorganized on a strictly rational basis during adolescence. Like Piaget in this one way, Vygotsky assumed that the ideal final state of development was logical and opposed this style of thought to the childish or primitive. In light of research done more recently, such an assumption seems naive. Connectionist representation allows us to see how Vygotsky's insight into how "scientific" categories are learned remains valid, and how the techniques based on this insight remain effective even if the basic mechanism which subserves cognition is not rationalized.

In *Thought and Language*, Vygotsky discusses two ways to come to mature or logical conceptions. The first path, the path "spontaneous concepts" take, is to move from the sorts of broad, concretized labels exemplified by the word "Mama" to rational categories characterized by their abstract features. The second path, the path of "scientific concepts," begins with a definition based on necessary and sufficient features and is characteristic of learning in schools.
Vygotsky writes that this second sort of learning “plays a leading role in the development of school children.” In light of the research done in intervening years—particularly in attempts to extend Piaget’s similar intuition about the logical endpoint of development—this formulation must seem questionable. Piaget’s project had been constructed in a way which made the developmental continuum leading to logic essential to the program which developed from his work. With the empirical failure to sustain a general, stage-like move to a logical mode of thought, the whole project was thrown into doubt. Vygotsky’s work, however, has been seen as centrally concerned with the social aspects of development. Logic is not essential to the Vygotskian project in the same way that it was for Piaget.

Can we account for Vygotskian successes without making his assumption that what is constructed through the processes he describes, the formal, logical categories which we now doubt, can serve the fundamental role he assigns them? If we are to do so, what then unifies the two types of learning, spontaneous and scientific, if their telos—logical activity—no longer serves this purpose?

Delving a little deeper into Vygotsky’s typology we notice that the word plays a central role in both of his forms of concept learning. In spontaneous learning the word label is originally concrete and is not, for Vygotsky a full example of a sign but is only its functional equivalent. In his view young

---

218 Vygotsky, Thought and Language, 148.

219 In view of the fact that Vygotsky emphatically rejected the then-dominant idea that logical categories were formed via the schema suggested by formal logic (see: Thought and Language, 142), it may prove useful for the interested reader to review pp. 138-140, where Vygotsky makes it clear that true concepts, for him, are composed of abstract traits. In part Vygotsky’s work remains fresh to educational readers because he effectively resists the notion that logic is achieved logically—an approach that is as dominant in educational practice today as it was in Vygotsky’s time.
children pass from thinking via categories which refer to “unorganized congeries” to thinking in terms of “complexes” to “pseudoconcepts” to, finally, “true concepts.” The passage that is described is one which moves progressively toward abstract, logical categories. More interesting for the present purposes is the role Vygotsky sees for the word:

...words take over the function of concepts and may serve as means of communication long before they reach the level of concepts characteristic of fully developed thought.

We are interested in this long period in which words may serve as a means for communication before they are said to become logical. During this period words are what I have called a practice and what Vygotsky labels a “functional tool.”

Words and other signs are those means that direct our mental operations, control their course, and channel them toward the solution of the problem confronting us.

This role of the word is Vygotsky’s distinctive contribution to our understanding; his work removes the word from its all too often disembodied and abstract status and remakes it as an exemplary practice.

---

220 Each functional stage is described in reference to this endpoint. See, for example, the discussion of complexes on p. 113: “Since a complex is not formed on the plane of abstract logical thinking, the bonds that create it, as well as the bonds it helps to create, lack logical unity.”

221 Vygotsky was aware that more than just words in language could function as a sign and his text makes this apparent, but the central role logic plays in defining a true concept leads Vygotsky to reserve full, logical signification for the word.


225 And the role of the word is what he saw as his own contribution to the discussion in which he was engaged. See the early pages of chapter 5 for a review of the field and p. 106-107 for his conclusion.
reflecting human activity. Vygotsky's clearest examples and the thrust of his experimental studies lie in a fine-grained examination of the way the use of words is socially managed and how the word comes to have a more and more mature meaning by the policing of its use by adults. By "mature" Vygotsky meant that the word came closer to the meaning and usage which an adult would display—an activity which he assumed was logical.

But Vygotsky revealed some ambiguity in this formulation. When he discussed pseudoconcepts, he noted, with an air of disappointment, that it was impossible to tell the difference between pseudoconcepts and true concepts in their use226 and that children are unaware of the difference in their own thinking when they use true concepts.227 Vygotsky also stresses that the final, logical development is never total, and that much everyday activity remains in the realm of complexes and pseudoconcepts.228 If pseudoconcepts are functionally identical to true concepts, what is the motivation that pushes the developing person to finally form true concepts? In contrast to the rest of Vygotsky's carefully laid out progression toward a mastery of word usage based on the practices of the social milieu in which the activity takes place, this final step is curiously unmotivated.

He gives the example of separating groups from a set of blocks, each of which has a unique combination of color, shape and size traits. Children progress from putting together maximally similar blocks to grouping on a

226 This is what Vygotsky is pointing to on page 121, when he says, "... separating a pseudoconcept from a real concept is not easy, and this task is positively beyond the capacity of phenotypical analysis."

227 Vygotsky, Thought and Language, 123.

228 Vygotsky, Thought and Language, 134-135. In this passage Vygotsky is moved to invent yet another category between pseudoconcepts and real concepts: the generalized representation. The generalized representation seems to be a pseudoconcept used by an adult.
single trait such as color. Any grouping means foregrounding some attributes and backgrounding others. For Vygotsky this is the beginning of the creation of abstract concepts, "true concepts" in his terminology. For a true concept to develop, these traits must be abstracted from their context, appropriately labeled by a word, and the original complex (pseudoconcepts) which functioned as a concept in the child’s daily life must be reconstituted as a true concept. This is just the sort of logical object that was described as emerging from the analytico-referential discourse in chapter 2 of this work. In seeking to link his penetrating observations on the actual practices associated with children’s categorization, Vygotsky is attempting to show how this sort of privileged knowledge can result from the process that he describes. Vygotsky describes a process built on practices or tools which move a learner toward increasingly more powerful abstractions continuously built on practices and the learner’s activity in the world based on those practices. The Mother emerges from the background and is progressively abstracted from the child’s earlier category which included all people and dogs (from our perspective) to a category which even in Vygotsky’s explanation functions in just the way of true concepts. What makes this terminal abstraction different from any other act of abstraction? Is it somehow not based on social practices? If this distinction is a practice-based one, what distinguishes it from earlier abstractions, say the one where the term “Mother” included all women but no longer any children or dogs? What distinguishes it from further abstractions such as “nurturer?” What prevents Vygotsky from seeing logic as a category of practice, with all the vagaries that attend categories organized on the basis of practice?

The perspective being developed here throws profound doubt on the existence of logical organization as a possible basis of thought—be it in the child or the adult—and logic is instead understood as a practice among other
practices arrayed within particular communities to ends conceived within those communities. That the scientific and academic communities have valued practices which enable their everyday activity should be no surprise and realizing this should in no way throw doubt on the value of those activities. Viewed against the discursive background, Vygotsky appears to have had to struggle, within the analytico-referential discourse, to get to "real," abstract, formal logic grounded in formal categories composed of necessary and sufficient features or fail to locate true concepts and true thought at all. We need not be so constrained. Vygotsky does not, and perhaps cannot, say at what point the concrete, material attribute whose similarity is the basis of a pseudoconcept becomes a disembodied feature which is the basis of abstract, logical reason. This sleight-of-hand transition is compelling nonetheless because the tradition implies that logic is achieved, since that is the only way we have to explain our competent action in the world. The irony of Vygotsky's stance is that he went far toward showing just how it might be in that the socially organized milieu of language and other practices that pervade our lives are the functional equivalents of logical concepts. These pseudoconcepts—concretely formed and organized—are perfectly serviceable tools to get us through the daily routine. He balked, though, at understanding true concepts and especially "scientific concepts" as continuous with spontaneous ones. It is to this area, and a further elaboration of Vygotsky's understanding of the role of the word, that we now turn.

An analysis of Vygotsky's understanding of the leading role of definitions in acquiring scientific concepts is suggestive for the point being developed here. On Vygotsky's account scientific concepts and their acquisition in schools in formal terms are major factors in the achievement of abstract, logical categories during development. He understands this as a case of starting from the
scientific, logical social product of a concept and arduously filling in the concreteness that is necessary to make it useful. Vygotsky strongly stresses that the development of scientific concepts moves in the opposite direction from that of spontaneous concepts. It moves from the abstract to the concrete, and the student—and the teacher—encounters the greatest difficulty in making the concept appropriately concrete. What is revealing for the present purpose is how this concreteness is achieved. Roughly, what Vygotsky finds is that though the child can repeat and even rephrase scientific concepts well, he or she does not, at first, have any deep appreciation of their meaning; in fact these are better characterized as images in many of the concrete ways that spontaneous concepts and pseudoconcepts are understood. What the child first acquires is the proper use of a word in the academic context of schooling. The practices of schooling continuously reorganize what this term represents to the student in a way which is strongly reminiscent of the process of moving to the more socially appropriate uses of terms like “Mother,” which we discussed above.

A closer examination, an examination which Vygotsky himself enables, reveals a single process and a single line of development in the acquisition of appropriate conceptual apparatus. The child starts with concretized, imagistic categories embedded in social practices which are a part of the child’s social environment. The child learns to wield these tools, generally in the form of words, more and more appropriately as he or she is inducted into practices in which usage is further constrained. Eventually the child is able to wield these tools in a way that is appropriate to the social milieu and has achieved functional mastery of the concept. This path is the same whether the tool in question is understood as a spontaneous concept or a scientific one. These

---

229 Vygotsky, *Thought and Language*, 193. “This oversimplified development of the concept of serfdom looks more like an image than a scientific concept.”
concepts can only be differentiated from outside the analysis that Vygotsky performs. Concepts conceived as logical objects composed of abstract features are assumed prior to analysis. The unquestioned assumption that they exist at any point as the basis for our in-the-head thought (as opposed to being a valid practice in our discourse) is the basis for assuming that such concepts either develop from readily demonstrable concrete categorization or that preexisting, abstract, scientific concepts can be filled with concrete content.

Note that in all this I am not claiming that abstraction does not exist or that Vygotsky's work is not valuable to teachers. On the contrary, his work demonstrates in a fine-grained way that language and particularly words are valuable practices whose power lies precisely in their ability to be manipulated as external representations that encapsulate increasingly powerful abstractions—including abstractions constructed hierarchically out of earlier words. Working from the traditional understanding of what constitutes useful knowledge education has conceived of its role as teaching powerful abstractions. This remains a noble and appropriate goal, even in a situated connectionist accounting. But educational practice has not been as effective as we might wish, in part because we mistake the nature of abstraction and associate it with logical objects and objective definitions. Vygotsky brings us to the point where we can see words as effective practices and can imagine how they and other practices might be designed to help our students take useful possession of the abstractions that they represent. The issue is not how to encourage the construction and/or memorization of logical objects of knowledge, but how to build practices with the greatest generality, accepting that all representations-in-the-head and practices-in-the-world are dependent upon context for their realization. A central task becomes building practices that allow the student to perceive previously differentiated contexts as the
same, and therefore to see particular, previously established practices as appropriate in the new context. Certainly language and the word that Vygotsky invests with so much power is a chief avenue toward that end.

_Dreyfus and Rules_

Rereading Vygotsky through the lenses of practice and connectionist representation helps us better understand the role that formal definitions in particular and language in general play in development. Such a reading provides a way to understand the continued value of seeking powerful abstractions in education while changing our image of what such an abstraction might be and how it is to be learned. But this involves only one side of the traditional framework as put forward in chapter 2. In addition to the logical objects of the tradition, defined in terms of their features, the tradition also points to the power of rules.

Again we are faced with a situation in which there is a temptation to say that the emerging perspective endorses the position that there are no rules. But as Bereiter has noted, rules are too valuable in the practice of education to abandon.\textsuperscript{230} Bereiter, in his article "Implications of Connectionism for Thinking About Rules," makes the educationally salient point that in a connectionist interpretation rules are best understood as elements of public discourse rather than the substance which constitutes rationality. As he puts it:

> What about the explicit teaching of rules? Once we recognize rules as part of the public discourse rather than as lines of mental program code, we can afford to be entirely pragmatic about their use in education.\textsuperscript{231}


\textsuperscript{231} Bereiter, “Implications of Connectionism for Thinking about Rules,” 15.
But Bereiter gives little specific guidance on how to be pragmatic in this area; as educators we will need to listen to a more formed voice on the specific ways in which this might apply to education. Dreyfus and Dreyfus suggest that rules play a crucial role in the development of competence that is compatible both with the line of reasoning we are developing here and the mission of teaching displaced knowledge with which the school is charged.\textsuperscript{232} On the brothers Dreyfus account the traditional story has the expert developing expertise by moving from specific cases to a higher, abstract understanding of the problems and a competence based on this understanding. This understanding is familiar to educators and is the source, in one guise or another, of the claim that there are wide swaths of competencies that schooling cannot prepare the student for precisely because schooling takes place in the context of school and not in those places where the actual building blocks of competence can be acquired. This understanding is made oddly more credible by the persistent inability of experts to articulate the rule-based basis for their expertise—granting the assumption that rules must be at the basis for their expertise, the experts' inability to verbalize their rules makes it all the more necessary that students join them in the field. The Dreyfus analysis, based on phenomenological considerations, triumphantly concludes that experts do not use rules at all—rather, they recognize many independent situations as calling for appropriate action.

This conclusion concerning the basis of an expert's competence dovetails well with the connectionist vision of representation that we discussed above; connectionist representation seems tailor-made to combine both perception and

action in the way the Dreyfus thesis demands. On one point, however, a
collectionist would differ from their interpretation of the empirical data: the
expert—or the competent actor in an area of human endeavor—would not
recognize many thousands of separate instances but would recognize in many
thousands of instances the relatively few categories which would call forth
appropriate action. The Dreyfus brothers are willing to trade many thousands
of particularized, unconscious rules for many thousands of separate, vivid,
prototypical memories. A fair trade, they think, based on the evidence. But
experts are as unaware of comparing the current situation to a stored
prototypical memory and reacting appropriately on the basis of this perceived
similarity as they are of following rules. Rosch’s own later work in prototypes
(the earlier version of which these two writers draw on in the work cited) has
led to a much weaker version of prototypes without the strong, central,
remembered instance. The two authors’ own phenomenological roots would
indicate the need for a process of representation and action that operates
smoothly when uninterrupted. A phenomenological approach supposes a
process in which perception brings, as a part of the act of perception itself,
regularized activity in its wake. This activity is itself part of the reorganization
of the field of perception and prepares the way for further, appropriate
perception and activity. Any interruption, in the anticipated perception or in
the activity, disrupts this process and calls for more conscious, considered
regard. The connectionist proposals regarding human perception and activity

233 See the discussion in their section: “Stage 5: Expertise,” 338-340.

234 See the discussion by Lakoff of Rosch’s changing position and its implications in Women,
Fire and Dangerous Things, 39-57.
describe just this sort of process. Vivid (but unconscious), independent prototypes are not a comfortable part of the story.

This difference should not blind us to the way the Dreyfus’s account aids our project. While they reject rules as the basis for the experts’ competence and endorse a broad range of experience in the area as an antidote, their account is not one which calls for simply placing students in the context where their desired competence is practiced and hoping that they will induce the unconscious rules which lead to full mastery. Instead they believe that competence begins—not ends—with rules. Novices are taught via rules. Their own formulation is brief enough to quote without paraphrase:

Normally, the instruction process begins with the instructor decomposing the task environment into context-free features which the beginner can recognize without the benefit of experience. The beginner is then given rules for determining actions on the basis of these features, like a computer following a program.235

While we may choose to doubt the existence of “context-free features,” the thrust of this formulation is clear. The teacher draws the attention of the student to particular parts of the overall situation. The student is then offered a set of rules which abstract skeletal practices common to the domain. The brothers Dreyfus use the example of a chess game in which each piece is assigned a value, and a rule is given that calls for an exchange at any point in which the value of the exchange is in the actor’s favor. From the point of view offered here, explicit features and rules based on those features’ presence or absence are pedagogically useful when they enable the student to begin to engage in the problem domain with even the roughest approximation of the fluid competence of the expert. Considered as teaching practice such a program orients the student toward the parts of the domain which the more competent find

significant. This begins to engage the student in the type of activity that would follow from the perception of experts—the use of decontextualized features woodenly emulates such perception. Taking this position allows us to judge when rules and abstract features are likely to be useful—and to judge when they become a hindrance. Viewed from this position, the traditional story of abstract features and unyielding rules, one suspects, is but an unfortunate reification of valid teaching practices.

Vygotsky and the Dreyfus brothers advance our understanding of the abstract, powerful and displaced knowledge which education is charged to teach. An examination of their work reveals that the combination of connectionism and practices being pursued here can make meaningful contact with the “higher” functions that schooling pursues. While both Vygotsky and the Dreyfus brothers advance our understanding of how social learning take place, neither focuses on the practices which have emerged as central to the account of learning being developed here. The interests and subject matter of Vygotsky and the brothers Dreyfus are especially congenial to the interests of educators, but to gain a firmer grip on practices, we will need to examine the work of other social theorists.

Theories of Practice

There are historically many strands which may inform our understanding of how symbols exist in the world. Broadly we can divide these earlier attempts into two categories: the asocial and the social. Asocial approaches are typified by those who treat experience and the world as important determinants of what we learn but focus on this to the exclusion of the social organization of experience. This is archetypally represented in an educational context by the
close-grained work of Jean Piaget in the child’s acquisition of certain basic conceptual abilities.\textsuperscript{236} Such an approach is also apparent in the work of the perceptual gestaltists.\textsuperscript{237} Perceptual ecologists who descend from Gibson also have tended to work in a world in which “affordances” were simply there in the environment to be picked up.\textsuperscript{238} Such work can be mined for its important insights into the identification of concrete symbols and a long tradition of working with such concepts but, finally, they are not central to education’s task because the sorts of symbols (invariances, gestalts, or affordances) that they investigate are closer to the embedded knowledge that, like apprenticeships, are seldom appropriately taught in schools.

Our interests are more focused on those who have acknowledged the complexly social nature of learning and have struggled to understand how knowledge might exist as a result of people’s interaction with each other and the world. This is an immensely more complicated task, and were our mandate other than the education of our children, we would be wise to avoid the prospect. Attempts to understand this connection have been many and the prudent educator will make use of a wide array.

The diversity of approaches is evident in a simple listing of the areas in which work has been done that is of interest to those trying to understand how knowledge might be embedded in the world. Pragmatism, phenomenology, semiotics, continental sociology, and a complex of Russian activity theorists

\textsuperscript{236} For example, see Inhelder and Piaget, \textit{The Early Growth of Logic in the Child}.

\textsuperscript{237} This was true of Kohler’s generation, but it is not true of their thoroughly socialized descendants who recognize the central importance of social shaping (for example, Laing or Rogers).

\textsuperscript{238} Again, not all who follow are as limited in their approach as the originators. See: Donald A. Norman, \textit{The Psychology of Everyday Things} (New York: Basic Books, 1988).
have all been fertile frameworks for research into this area. Each discipline has its particular strengths and weaknesses.

Pragmatism, descending from C. S. Peirce, gives us triadic semiosis, Deweyian educational philosophy, and Mead’s understanding of the formation of the self and the role of institutions. Much of the power of these approaches can be traced to the Peircian concept of sign activity in which the object that is in the world, the representamena that is in the mind and the interpretant that the person produces (for example: a word, or an habitual action—reaching out) are irreducible parts of the sign in the process of human signification.239

Viewed through the lens of semiotics, the present work is centrally about the representamena, how it is learned and what its qualities might be as a material entity. This chapter, in focusing on the social aspects of learning, acknowledges the Peircian insight into the importance of the interpretant in its socially organized guise. It is this interpretant that others, and we ourselves, interpret and act on the basis of. The dynamic quality of triadic semiosis, with its insistence that sign activity is a continuous participation in a complex network of relationships which together shape the human outcome, makes semiosis especially congenial to an interpretation of cognition based on network architectures.

Heideggerian phenomenology, which also emphasizes the indissoluble relationship between activity in the world and meaning, holds potential to help us explicate connectionism’s external symbols. The instructive history of phenomenology, which moves from Husserl’s attempt to describe explicitly the

---

239 For a fuller explanation, see: Whitson, “Cognition as a Semiotic Process.”
predicates that he says compose the world (noema) to Heidegger's devastating critique of this project, reflects on artificial intelligence's similar project.\textsuperscript{240}

The complex of Russian theorists who struggled to understand this area remains underappreciated. Vygotsky, of course, is now familiar to an educational audience, and his work—particularly ongoing research into the concept of the zone of proximal development—is possibly the most relevant work to issues of practice now being pursued in educational research.\textsuperscript{241} But other work, particularly that of Luria, who investigated the boundary between the social and the neurological\textsuperscript{242}, and Bakhtin, who explored language and particularly genres in a way that will prove productive in later chapters, are also important sources of ways to understand how symbols can take on a stable external existence. Even less well known are the activity theorists such as L'ontiv who are heir to the richness of Russian social thought.\textsuperscript{243}

Similarly, a reexamination of the classical sociological concept of role, of Bourdieu's approach to habitus,\textsuperscript{244} and Foucaultian senses of practice will surely prove fertile ground for those who wish to recast the symbols of cognition externally. To even begin to adequately explore these opportunities would be the subject of a much larger work than is possible here.

\textsuperscript{240} Dreyfus, \textit{What Computers Still Can't Do}, 34-37, 310n.

\textsuperscript{241} Perhaps the most useful example is: Denis Newman, Peg Griffin, and Michael Cole, \textit{The Construction Zone} (Cambridge: Cambridge University Press, 1989).

\textsuperscript{242} While Luria produced works which could be judged predominantly social (\textit{The Mind of a Mnemonist}) or predominantly neurological (\textit{The Working Brain}), a particularly inspiring book which cannot be so categorized is \textit{The Man with a Shattered World}.


Here we must resist the temptation to exceed the current mandate. Our purpose is limited to showing why a connectionist approach to learning is incomplete without a complementary social approach to knowledge and demonstrating that this connection can be fruitfully made. The author is indebted to a wider array of theories which bear on practice than are examined in this chapter, as will become more apparent in the two chapters which follow. Here this larger body of work will be slighted in favor of examining in more detail the possible connection to situated cognition.

Situating Situated Cognition

Situated cognition\textsuperscript{245} is examined in some detail here on a number grounds. Situated cognition is an active part of the current literature in education. Interestingly, situated cognition enters the discourse of education most forcefully through the offices of some of those associated with the cognitive science discourse that has been a background to the work of connectionism.\textsuperscript{246} Situated cognition is also, frankly, as needy as is connectionism of a partner in making its case. Many of the approaches mentioned above are actually fuller theories which include a sophisticated approach to describing the self. The lacks of situated cognition and connectionism neatly mirror each other.

\textsuperscript{245} Situated cognition, like most young movements, is difficult to accurately delimit. Here it will refer most centrally to Jean Lave’s work, the work of Cole and Scribner in anthropologically oriented approaches to which her work owes much, and the work of such as Brown, Collins and Duguid which attempt to use their formulations.

\textsuperscript{246} Brown, Collins and Duguid have all been associated with the cognitive science tradition critiqued in this work. For their latest position, see: John Seely Brown, Allan Collins, and Paul Duguid, “Situated Cognition and the Culture of Learning,” \textit{Educational Researcher} 18 (January, 1989): 32-42.
Perhaps most important, though, is situated cognition's undeviating focus on student competence and its central appreciation of the role of practices in constituting such competence.

The key insight of situated cognition is that practices are constitutive of competence in the world. Because practices are relational, social activities situated in the world, situated cognition's approach to knowledge and learning is a thoroughly social one in which the particular actors are secondary to the socially formed and sustained practices that they enact. The originary unit of analysis is the community which sustains these practices.

Situated cognition is in part motivated by the attempt to account for the infamous problem of the failure of "learning" to transfer from the context of its acquisition to the context of its application.247 More precisely, knowledge acquired in one situation, usually schooling, fails to be appropriately used in a situation where its application would seem obvious. In education, examples of this sort have traditionally been referred to, following Bloom's usage, as problems of application. Situated cognition cuts the Gordian knot of tangled explanations which have followed upon education's acknowledgment of this problem by moving the analysis away from a focus on knowledge contained in the head of individuals toward practices embedded in particular situations in which community members participate.

This solution, while elegant and exciting, brings in its trail new problems. The basic move of situated cognition is to take what constituted the heart of the problem—that students' competence were impaired when they moved out of

247 For an overview of the trajectory of situated cognition, see: St. Julien, "Explaining Learning."

the context of acquisition and into the context of use—and make that very
problem the solution by claiming that such competence as is available is
embedded in the situation itself. This turns the explanatory world inside out.
Accepting situated cognition’s vision, we should no longer expect the student
to show any transfer of competence between situations. In Lave’s words, the
learner will need to be a “legitimate peripheral participant” in the community
of practice in order to acquire true competence.249 The trouble here is that while
situated cognition gives us a strikingly plausible way of understanding the
intractable problem of the failure of knowledge to transfer usefully, it leaves us
completely unable to account for instances in which knowledge does transfer.
At its basis this problem arises because situated cognition has built a theory of
socially situated knowledge leading to competence on the part of participants
in their communities. It is the community, finally, which is the unit of analysis.
The student, as a separable, multiply constructed subject acting across
situations, does not exist—only participants with varying degrees of legitimacy
in ongoing communities of practice. Any explanation of successful transfer will
have to, at least in part, be built on the basis of the person who acts across
situations. Without a theory of the-person-learning, situated cognition will
remain incomplete in regard to the purposes of education.250

This pattern is strikingly reminiscent of the connectionist resolution to the
problem of representation. There the basic move was to take the problem that

249 Jean Lave and Etienne Wenger, *Situated Learning; Legitimate Peripheral Participation*

250 It should be noted that Lave, at least, is strikingly unconcerned with such purposes. On
her account, education as currently conceived is simply misguided and must fail because it
encourages a form of learning which is bound to be systematically distorted. See, in one
example, Lave, *Situated Learning*, 94-100. I have commented on this problem and the
implications of apprenticeship that follow from Lave’s position in: St. Julien, “Situated
Cognition, Apprenticeship, and Schooling.”
representation in use seemed altogether too plastic, changeable, and slippery to support the kind of rationality that our cultural background leads us to understand as essential to reliable thought, and invert the meaning of that plasticity by making the plasticity itself the basis of reasonable thought and action. Connectionism, I have claimed here, buys a theory of learning but at the price of being left without a theory of stable knowledge—a stability which must be located not in the mind but in the recurrent patterns of our interaction with the world. Dovetailing with connectionist insights, situated cognition purchases a theory of situated knowledge which locates stability in the world but at the price of losing the subject which acts across situations.

A theory of learning and knowledge adequate to support educational aims must do both. For our relatively narrow educational purposes we have come to focus on connectionist representation and situated practices as the locus of our examination. This has been motivated in part by our understanding that connectionist representational modes are dependent upon recurrent experiences which are the constant basis for our ongoing re-cognition of situations and things in the world.

Conclusion

This chapter has traced a path through the implications of the social forms of our lives for a learning theory based on connectionist principles. Against the backdrop of an ongoing debate about the nature of representation connectionism presents a startlingly new alternative. But it was less than clear that the learning that schools were mandated to encourage could be supported by an understanding of representation built on connectionist approaches. Constrained by a commitment to education and particularly to focusing on the
fine-grained practices that constitute successful teaching, we then examined the work of Vygotsky and Dreyfus. Their work encouraged us to believe that there were alternate accounts that admitted the value of sharply defined words and domain-specific rules which did not conflict with, and indeed which could be enriched by, a connectionist account. We then briefly discussed other theorists who could inform the approach to practices that the earlier explorations had led us to emphasize. We finally settled on situated cognition as a potential partner for connectionist approaches in part because the strengths and weaknesses of these two exciting new approaches were complementary.

In the two chapters that follow the tenor of this work will change. To this point I have been occupied with laying out a framework for understanding learning in a new light without losing sight of the purposes of schooling and the needs of students and teachers. Having laid out one particular way of integrating connectionist insights into the universe of education—and having pointed to connectionism's very real deficits as a learning theory that would be useful to education without a social component based in practices—it remains incumbent upon this writer to show how such a proposal could be fruitful for both educational research and instructional design. Chapter five will explore a research project based in part on the insights developed here, and chapter six will discuss an instructional computer program inspired by the research discussed in chapter five and exemplifying the combination of situated practices and connectionist learning developed in this chapter.
CHAPTER 5
Eukaryotes or Prokaryotes?

The locus of the mind is not in the individual. Mental processes are fragments of the complex conduct of the individual in and on his environment.

— G. H. Mead,
The Philosophy of the Act, 1938, p. 372

This dissertation has proposed an unfamiliar way of conceiving learning and knowledge. It suggests that representation is best understood much more broadly than is currently the case. Both perception and readiness for action are integral parts of the material cascade of differences that compose a connectionist network's analogs to traditionally conceived memory and thought. I have suggested that these qualities and more technical qualities related to networked architectures make connectionist models of learning powerfully explanatory when viewed against the background of difficulties arising from received explanations. But the valuable qualities of contextually sensitive plasticity which make connectionist forms of representation attractive bring in their trail questions about how the observable stability of knowledge can be accommodated by connectionist models of learning. The account given here turns radically away from trying to correct such deficits within the model of connectionist theorizing251 and toward finding the necessary stability in

251 Though such an approach is possible and is the main path those working in connectionist paradigms are following. Most such work, in my interpretation, buys stability at the cost of biological plausibility or by giving up some of what makes connectionist representation valuable, such as interactive representations. A little explored possibility and one which would seem to have much potential is the possibility that patterns composed of patterns would be more stable than individual ones. For a brief exposition of this idea see: John A. St. Julien, “New
external, material symbols. Such symbols, tied to human activity through practices, make good use of the pattern recognition and completion qualities inherent in the superimposed representations characteristic of network architectures.

This image contrasts so vividly with the traditional understanding of learning and knowledge that it was not apparent just how connectionist explanations could connect with the usual understanding of the role of the school and the practices common in schools except to condemn them. Chapter 4 dealt with some of these problems on a theoretical level by attempting to support social and phenomenologically inspired accounts of such “higher” activities without discarding the valuable portions of received practice. This chapter and the chapter that follows will attempt to put such ideas into play and to serve as a test ground where the reader can judge the potential practical value of the approach developed earlier in this work for research and instructional design.

The Background of the Study

The study discussed in this chapter grows out of an interdisciplinary microscopy teaching project with a research component. This project, and the exploratory research study this section focuses on, were rich in implications which have been reported on elsewhere.252 One element of the project brought


The microscopy research discussed in this chapter was supported by NSF-LASER grant (1991) HDR 01: “Exploring Microstructures: Introducing Biology Students to the Images, Tools,
students from groups that are traditionally underrepresented in scientific professions into the university to explore the use of the university's electron microscopy laboratory. Finding a way to teach students to make sense of the other-worldly images produced there was a central challenge for the developing project.

The long-standing interest of Dr. Jim Wandersee (the principle investigator) in the use of graphics and imagery in scientific texts along with his pedagogical experience and that of others in the project suggested that making sense of scanned electron images was not a trivial problem. As research into transfer might lead one to suspect, it was the common experience of teachers that students who are capable of doing well on a traditional test on the features which characterize this distinction do not thereby gain the ability to recognize the cells on a page or in a laboratory setting.253 The use of micrographs in this project pointed to a central conundrum in biology education: even though much of the knowledge base in biology was originally gained by looking at objects and images, surprisingly little understanding is conveyed through this medium in most textbooks. A body of research suggested that the simple availability of accurate images in textbooks was insufficient—students seldom gleaned from the textbook images the knowledge that the designers had intended. A wide array of reasons for the disconnection between image and textually-based understanding was suggested in research group meetings and included such factors as testing methods, a functional rather than analogical or

and Applications of High-Tech Microscopy." Other members of the research team, without whom this chapter would not have been possible, were, from Life Sciences: Drs. Becky Demler, Cindy Henk, Sharon Mathews, and Marion Socolofsky, and from Education: Drs. Catherine Cummins and Jim Wandersee.

253 See Hugh Gladwin, "In Conclusion: Abstraction Versus 'How It Is.'" Anthropology and Education Quarterly 16 (Fall 1985): 207-213.
narrative textual style, and leaving implicit the imagery in the Greek and Latin-derived technical terms. The larger research project pursued all these possibilities, but the author’s attempt to turn to connectionist representation is the current focus.

The particular problem the research team decided to investigate suggested a strong contrast with connectionist approaches. This problem was teaching the taxonomic distinction between Eukaryotic and Prokaryotic cells using locally developed micrographs. Taxonomies are archetypal examples of traditional, classical category formation. Linnaeus built his seminal work on the basis of necessary features, and his grand project was based on the assumption that one could reliably categorize the entire living world into discrete categories based on these features. Linnaeus’s design (Systema Naturae, 1735) was a systemization of the natural world within the analytico-referential discourse. The Eukaryotic/Prokaryotic distinction, however, is based on the latest widely popular taxonomy of life: the five kingdom system. This system, in line with modern biological theory, emphasizes ecological and evolutionary considerations in addition to Linnaean features. Even so, this distinction is typically taught as though it were based solely in Linnaean assumptions; our local text, for instance, listed features of each group and produced an extensive chart comparing the essential distinguishing features.

Even more suggestive was the struggle that our life sciences colleagues went through in trying to convey the proper basis for making the distinction. It became apparent that they—experts all, who could make this distinction

\[254\] For example, “eukaryote” derives from a Greek phrase meaning “with nucleus.”

themselves at a glance—could not easily tell us on what basis they made this distinction. They struggled, disagreeing, to give us a list of definitive features. Rather than take our experts as inarticulate, I took them as experts . . . if they did not know how to respond, then it was likely that we were asking an inappropriate question. Like the experts the Dreyfus brothers discussed, it seemed possible that their expertise in this matter lay more in their experience, and the perception that appropriate experience engenders, than in a knowledge of any set of features or rules. The contrast between their perceptual competence and their analytic difficulty focused my attention on perception and its relation to category formation. So, the focus shifted toward what these experts did best: making perceptual rather than analytic distinctions. Because the explicit purpose of the project was to bring students from underrepresented groups into the scientific community as participants, it seemed appropriate to attempt to teach the competencies that actually support that community.

The contrasts between the observational grounding of taxonomy and the way it is taught and between the experts' practical competence and their inability to articulate a list of definitive features were striking and suggested a role for the perceptually based pattern completion of connectionist models.

Connectionist Constraints, Practice-Based Design

It would be straightforward at this point to simply recap the analysis already laid out in the preceding pages, to apply that analysis to the design of elements

\footnote{The assertion that experts are inarticulate about the grounds for their own rule-based knowledge is a chief presumption of "knowledge engineering" and expert systems. The present account assumes that the question which demands a culturally canonical rule to cover all cases is "inarticulate" in that it misses the perceptual basis for category formation—and that rules are not the basis for such learning. (For a connectionist-based discussion of rules in an educational context, see C. Bereiter, "Implications of Connectionism for Thinking about Rules," Educational Researcher 20 (April, 1991): 10-16.}
of the research instrument, and then to analyze the results using this same theoretical apparatus—straightforward, and a necessary task, but less than the whole story. Such a discussion would obscure the fact that the theoretical exposition laid out in the preceding chapters was critically shaped by my encounter with the empirical problem of designing teaching and research materials for this project.

The practical project of designing learning materials for this problem made clear that while connectionist principles strongly constrained the sorts of representation that might be considered useful—leading to a rejection of usual ways of organizing the material and especially the images that were my central concern—they did not lead to much in the way of positive suggestions for design. As discussed above, connectionist approaches explicitly reject the standard, historically sedimented understanding of knowledge and how knowledge is acquired. Connectionists believe that objects of knowledge modeled on the discrete categories of classic logic do not exist. Because category formation is not a matter of necessary and sufficient features, a connectionist account would reject the traditional assumption that categories can be acquired through the memorization of features or that such features are what is induced during experience. On the other hand, the chief competing model of knowledge, the prototype account, is implicitly rejected because connectionist models are not based on and do not imply that representation is based in persistent, unitary, central instances against which all other members of the category are compared.

Interestingly, taking either of the two older competing approaches could reasonably lead to the design of curricular material similar to that found in the standard textbook. Regarding images and making the categorical distinctions which are central to taxonomy, both would imply that one should start with an
ideal image which exemplifies either all the relevant features clearly or presents
the perfect example clearly. Both cases begin by presenting the unity of the
category to be learned as unproblematic. One simply bases that unity in a single
set of features while the other bases that unity in a vividly experienced, holistic
image. In both cases the unitary object represents the ideal of useful knowledge.
They differ chiefly in how this ideal image is apprehended. Neither truly
challenges the cultural givens of objective knowledge upon which the current
discourse depends.

Connectionist modeling, on the other hand, indicates the danger of such an
approach. In connectionist experiments it has proven crucial to present the full
array of possible members of a category to a network in order that the network
learn to generalize properly. If a narrow range of examples is presented the
net will undergeneralize and treat only the exemplars as appropriate members
of the class. In fact, a crucial element of network training is to determine the
range and the total number of different examples that will, effectively, overload
the memory space and force generalizations. Too few will allow the net to settle
into a configuration in which each of a small number of similar exemplars is
treated as a distinct case. In solving some problems, workers in this area have
even resorted to reducing the number of hidden nodes to force
"overgeneralization."

The implication of this for the design of educational materials is that the
discursively "natural" tendency to present only perfect examples during the

257 Patricia Smith Churchland, Neurophilosophy: Toward a Unified Science of Mind-Brain,

258 A widely quoted and interesting example of both under and overregulation is found in:
Rumelhart, James L. McClelland, and the PDP Research Group (Cambridge, MA: MIT Press,
initial learning stage is misguided. For category formation at least, a connectionist approach would argue that the impulse toward simplicity exemplified by the phrase that "curriculum is content simplified for pedagogical purposes" is misguided. It may—and in connectionist models it does—result in faster and more reliable reactions to the particular exemplars in question. But it does not result in competent learning of the category, only in reliable responses to essentially memorized instances. Outside of those specific stimuli, the network reacts in a way reminiscent of the brittleness of symbolic systems: essentially no coherent response results. The flexibility of distributed representations is lost.

This is the sort of difference in theoretical implications which is testable. It is also a difference with immediate pedagogical implications: in teaching for the future competence of our students, is it most effective to teach a new category as if it were a single, central unity or is it most effective to teach that category, from the beginning, as if it were a dispersed, practical, category? In the former case, we choose one or few ideal exemplars and focus quickly either on making the features explicit (in the classical case) or on making the holistic image vivid (in the case of prototype theory). A distinctive connectionist approach would, instead, present a wide range of examples, and would include "outlier" instances in an early attempt to create a category which would encompass a range of instances similar to the final shape of the category in the use of the discipline from which the distinction is drawn.

The eukaryote/prokaryote distinction presents itself as an almost ideal point at which to test this conjecture. First, teaching this distinction is easily understood as enabling a perceptual distinction. Perceptual distinctions are usually thought of as difficult to teach in schools; enabling students to make such crucial practice-based distinctions is understood as a property of
experience and a strong reason for moving toward the apprenticeship models which have developed out of situated cognition.\textsuperscript{259} While much of connectionist theorizing is understood as minimizing the in-principle distinction between perceptual and so-called higher forms of thought, categories with clear real-world referents, such as chairs or birds,\textsuperscript{260} are the sorts of categories about which people seldom disagree in their lived lives and which support their shared interaction. The sought-after competence is a matter of seeing an object as a member of a socially agreed upon category.

The particular case of distinguishing eukaryotes and prokaryotes is valuable because it is unlikely that most people outside of disciplinary communities have much, if any, actual experience with the distinction. Eukaryotes and prokaryotes do not exist in most people's lived experience and they are brought into the lives of biologists and microscopists chiefly through photographs taken by inaccessible and expensive electron microscopes. For most people it is a truly new distinction made between unfamiliar objects. There is less prior meaning and association to potentially confuse and bias our analysis of learning than in most instances of new learning.

But if connectionist approaches question the wisdom of the traditional idealist approach to category formation and suggests that a wider set of examples is appropriate in teaching categories, they do not suggest concrete practices that replace the use of single ideal examples. This problem emerges clearly when we realize that no student is at all likely to take a stack of


\textsuperscript{260} These were favorite examples of Wittgenstein and Rosch, respectively.
micrographs and spontaneously divide them into eukaryotic and prokaryotic piles when asked to bifurcate the original pile. Immediately the issue becomes: how is an unorganized mass of materials to be made into a coherent presentation which will signify a difference between the two groups and lead to a perceptual reorganization that causes the student to see one set as eukaryotic and another as prokaryotic. On the surface it seems that one must invent signifying practices. Unfortunately, by their very nature no single individual can invent such a practice. There must always be another to whom the practice signifies. Practices are preeminently social and are usefully established only through a history of concerted use. Obviously, this is an unsettling conclusion if the current project is to create new curricular materials. It is not immediately clear how one can design a new way of presenting materials without using the already established practices which signify a particular, mistaken idea of what constitutes a category. How does one establish the unity within a category and the opposition between categories that bifurcations such as the eukaryotic/prokaryotic distinction rely on?

Struggling with this design problem led to a fuller realization of the necessity for taking into account the socially organized regularities of the world. I found I could not bring into play the ideas of connectionist representation and the power of pattern completion that it offered without a better grasp of the social practices and regularities that preexisted my attempt and, it became obvious, were essential to it. I was perhaps lucky to have the austere discipline of typography and graphic design in my background. This field frankly counsels that design elements are conventional rather than natural and, somewhat ironically, that the hallmark of good design is that it appears transparently natural to the reader. Apt use of convention in the service of textual meaning defines the field. The practitioner learns to see the text in terms
of visual conventions and the possible manipulations of these conventions to convey meaning.\footnote{See, for instance: Jan V. White, Editing by Design: Word-and-Picture Communication for Editors and Designers (New York: R. R. Bowker Co., 1974).} I fell back on such an approach without, at first, fully realizing its appropriateness.

 Luckily for the designer, there are practices in graphic design which signify to the reader opposition and unity, and these practices can be used to point to and, more profoundly, create relationships between objects. These practices are usually not consciously understood by the reader and are hence felt to be natural, but the designer must be able to see them as discrete practices to be manipulated in the pursuit of the craft. Unity, for instance, is conventionally signified by placing things to be grouped “uncomfortably” close together in an organized fashion. The six members of a management team are shown by placing their separate photographs in two rows with very little space between photos in a row or between rows. They reader “sees” a unit, easily, fluidly and, to all but the designer, naturally. Opposition is pointed to by placing the two opposing sides literally on opposite sides of the page. “Dueling” pitchers in a feature story on an upcoming baseball game are often placed on opposite sides of the page, a hostile relative positioning which every fan can read even if both images are stock, smiling publicity stills. Practices such as these, which seem natural but which actually have been established through long usage, are as available to the designer of educational materials as they are to any corporate designer. It is to such practices that I turned in the design of materials for the study. On reflecting about why a particular arrangement was appropriate for the meaning that I intended to convey, I came to fully understand that these common design principles were best described as social practices deriving their power from their previous history of use; as such, they were available to me to
signify or point to particular qualities of the images I hoped the students would learn to distinguish. I now turn to the study which occasioned these reflections.

The Study

In the microscopy project study we developed four one-page examples of textbook treatments used to introduce the eukaryote/prokaryote distinction. The first, representing a traditional approach, was drawn from Biology, a major college textbook that uses the five-kingdom system. The alternate treatment used two groups of five transmission electron micrographs developed by the team and a text that was organized around a metaphor of community—town and city—that emphasized complexity and differentiation. Two further treatments mixed these conditions; one associated the traditional treatment of the text with the newly developed micrographs and the other joined the traditional micrographs with the newly developed text.

The Experimental Materials

The traditional text exemplified the conventional practices of textbook publishing. It used a descriptive introductory paragraph, a single prototypic image of each cell type, and a cutline with the photographs pointing out the definitive features of the cells. This treatment is typical of introductory biology texts and of the genre of science texts. The text was expository and did not refer to the accompanying micrographs, the micrographs were prototypical examples of their class, and the cutline to the micrographs referred to the features exemplified by the micrographs. The alternate treatment used ten images, five micrographs from one category on the left and five from the other.

262 Campbell, Biology.
category on the right, with the text running down the center. The images were chosen to represent a wide range of members of the category, some of them "outliers." Social conventions involving symmetry, grouping, and opposition were employed to design a page that would lead the students to perceive a field in which similarities were to be sought within examples grouped together and differences to be sought between groups on opposing sides of the page. Our hope was that this could come, over a history of interaction with such texts, to be recognized as an occasion for category formation based on the differences suggested by the written text. We hoped that such practices could be added to the storehouse of tactics available to textbook designers and would constitute a more effective use of images in support of the activity of category learning, thereby extending the textbook genre.

The Conditions of the Study

The experimental materials were distributed with an accompanying protocol to the teachers in the biology department of a large junior college participating in the grant activities. 257 students completed the activity and accompanying questionnaire. The pre-test portion consisted of the four different combinations of traditional and newly developed imagery and text discussed in the previous section. After the initial materials were taken up a questionnaire was distributed which asked the students to recognize elements of the cells as described in the text and to categorize newly presented images as eukaryotic or prokaryotic.

While a number of interests were pursued in this portion of the work, including work on developing another genre element introduced in the post-

\[263\] White, Editing by Design.
test portion of the protocol, the portion of the project which is relevant to our current interests is the students' responses to a question asking them to categorize five previously unseen micrographs as either eukaryotic or prokaryotic. By grouping the two conditions which used the traditional imagery and the two conditions which had utilized the newly developed practice of placing two groups of five images in opposition, it was possible to isolate the independent effect of the imagery in the pre-test from the effect of the textual treatment in the pre-test.

Statistical Measures

Because the distinction that the learners in our study make are basically categorical, albeit a bifurcating nominalism, the more common parametric statistics are not appropriate. The nonparametric measure chi square ($\chi^2$) was chosen to reflect this constraint. This measure was used in two ways. The first, following Cohen, is known as the "goodness of fit" test and tests to see if the treatments, considered separately, successfully "taught" the distinction between prokaryotes and eukaryotes. In the language of statistics the null hypothesis is that, for each image categorized, the treatment did not significantly improve the students' score over chance. Were this conjecture true we would expect that on the average students would answer correctly fifty

---

264 We explored the value of a consistent underlying grid in the production of the perception of relative size in making the categorical distinction between eukaryotic and prokaryotic cells. Prokaryotic cells are usually much smaller than eukaryotic cells, but the traditional conventions of publishing and the practical economic constraints of printing make direct representation of this difference impractical. A reference grid would be a partial solution to this problem.


percent of the time and this statistic will allow us to characterize how often the observed result could have been achieved by chance alone. (See table 5-2) Here, and below, the level of significance is set at .05—indicating that I am willing to accept the risk that 5 times out of 100 a conclusion that there is a difference between the two treatments will be wrong.267

Table 5-1: Chi Square test comparing traditional and alternative treatments

<table>
<thead>
<tr>
<th></th>
<th>Traditional</th>
<th>Alternate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q11a</td>
<td>Q11b</td>
</tr>
<tr>
<td>observed</td>
<td>expected</td>
<td>observed</td>
</tr>
<tr>
<td>Euk</td>
<td>12</td>
<td>57</td>
</tr>
<tr>
<td>Prok</td>
<td>102</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euk</td>
<td>99</td>
<td>56.5</td>
</tr>
<tr>
<td>Prok</td>
<td>14</td>
<td>56.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euk</td>
<td>51</td>
<td>55.5</td>
</tr>
<tr>
<td>Prok</td>
<td>60</td>
<td>55.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euk</td>
<td>52</td>
<td>56.5</td>
</tr>
<tr>
<td>Prok</td>
<td>61</td>
<td>56.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euk</td>
<td>97</td>
<td>57</td>
</tr>
<tr>
<td>Prok</td>
<td>17</td>
<td>57</td>
</tr>
</tbody>
</table>

267 There are numerous special forms of chi square. To avoid problems in comparing numbers drawn from differing algorithms, I have taken the basic formula from Steele and Torries and have implemented it in a Microsoft Excel worksheet. The tables that follow were all constructed from that worksheet.
If the first use of the chi statistic was to test whether or not either given treatment (and an undetermined prior experience) would allow the student to correctly categorize the images, the second use tests whether one treatment can be said to be significantly better than the other in teaching each image. There are two statistical approaches using chi square that can aid us in making this judgment. The first involves comparing the total number of right and wrong answers across questions in order to determine whether there is a significant difference between the effects produced by the two treatments. The results of this approach are shown in figure 5-3. It demonstrates a significant difference between the two treatments at the .05 level. (Indeed, this result is significant at the .005 level.)

Table 5-2: Chi Square test comparing traditional and alternative treatments

<table>
<thead>
<tr>
<th></th>
<th>observed</th>
<th>expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>right</td>
<td>481</td>
<td>419</td>
</tr>
<tr>
<td>wrong</td>
<td>107</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The second statistical approach to making this judgment relies on the chi square characteristic of additivity. Because the chi square statistic is additive, table 5-4 can be used to contrast the overall difference between the two treatments by simply adding the separate chi square measures for each of the five instances and figuring the significance based on five degrees of freedom. The null hypothesis is that there is no difference between the two populations that received the alternate and traditional treatments of the images. Again the

found statistic, 25.637, with five degrees of freedom, is significant at the .05 level. Like its sibling it also is significant at the .005 level.269

Table 5-3: Chi Square test of homogeneity; comparing traditional and alternative treatments via the addition of individual statistics

<table>
<thead>
<tr>
<th></th>
<th>comparing Trad &amp; Alt Qa</th>
<th></th>
<th>comparing Trad &amp; Alt Qb</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>observed</td>
<td>expected</td>
<td></td>
<td>observed</td>
</tr>
<tr>
<td>Euk</td>
<td></td>
<td></td>
<td>1.333</td>
<td>Euk</td>
</tr>
<tr>
<td>Prok</td>
<td>110</td>
<td>102</td>
<td>0.627</td>
<td>Prok</td>
</tr>
<tr>
<td></td>
<td>1.961</td>
<td>3.794</td>
<td></td>
<td>2.571</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>comparing Trad &amp; Alt Qc</th>
<th></th>
<th>comparing Trad &amp; Alt Qd</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>observed</td>
<td>expected</td>
<td></td>
<td>observed</td>
</tr>
<tr>
<td>Euk</td>
<td></td>
<td></td>
<td>3.843</td>
<td>Euk</td>
</tr>
<tr>
<td>Prok</td>
<td>79</td>
<td>60</td>
<td>6.017</td>
<td>Prok</td>
</tr>
<tr>
<td></td>
<td>9.860</td>
<td>2.120</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>comparing Trad &amp; Alt Qe</th>
<th></th>
<th>Additive Chi Square: 25.637</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>observed</td>
<td>expected</td>
<td></td>
</tr>
<tr>
<td>Euk</td>
<td>111</td>
<td>97</td>
<td>2.021</td>
</tr>
<tr>
<td>Prok</td>
<td>7</td>
<td>17</td>
<td>5.882</td>
</tr>
<tr>
<td></td>
<td>7.903</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(1df, .05=3.84)

Knowing whether the differences in the data are significant at the .05 level allows us to say whether the differences could reasonably be interpreted to have occurred by chance alone. While such statistical significance is valuable, it does not directly give us a sense of how important these differences might be. Power statistics can be used to provide a sense of the relative size of the difference. Such statistics are underutilized in educational research with policy implications, where confirming a difference does not necessarily imply that the difference is large enough to warrant the investment necessary to change ongoing practices. The effect size index “W” is adopted from Cohen for this purpose.

Table 5-4, Effect Size: \( W \)

<table>
<thead>
<tr>
<th>Power Statistic (( W ))</th>
<th>observed</th>
<th>expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>right</td>
<td>0.818</td>
<td>0.742</td>
</tr>
<tr>
<td></td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>wrong</td>
<td>0.182</td>
<td>0.258</td>
</tr>
<tr>
<td></td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.030</td>
<td></td>
</tr>
</tbody>
</table>

\[ W = 0.175 \]

Discussion of Statistical Results

The statistics used above can be conceived of as attempts to answer the following series of questions. 1) Does this teaching tool work? Do students do any better than chance on the post-test after exposure to the pre-test instrument? 2) Does the tool do its job any better than the usual alternative? Will students do significantly better after using the novel design than similar students do after using a teaching instrument adopted from a traditional text? 3) Supposing a significant difference is found, just how important is it?

The "Chi Square Test of Significance" table 5-2, reassures us that the method does work—students are able to do significantly better than chance after exposure to the pre-test instrument (see the column "alternate treatment"). The table also shows that for three of the images the traditional treatment was also successful, although at lower levels of confidence. What is most interesting is the two images for which the alternate treatment was successful and the traditional treatment was not (images C & D). Both of these image are of prokaryotic cells and were images which were included in the data set during development as "outlier" cells. College instructors and former biology teachers identified these as images which would, in their opinion, be among those in the total set which were most difficult for students to classify correctly. The intuition of these teachers was borne out under both conditions. These are the
images which students in the traditional treatment failed to classify successfully at better than chance. Less dramatically, while the students who experienced the alternate treatment did better than chance, these are the two images in which we have the least confidence that random error has not fooled us: the other three images have confidence ratings that indicate that only 5 times in a 1,000 could these findings have occurred by chance.

These two images were difficult, at least in part, because they had a discernible "center." For those students who were looking for a feature, the only candidate provided by the pre-test was the nucleus. On this interpretation those students who were looking for a nucleus but who had insufficient experience to correctly identify the nucleus—or more pointedly, had experience of the wrong kind—were tempted to interpret any central body as a nucleus. The students who were exposed to the alternate set of images were presumably better able to disambiguate nuclei from other organelles.

While the results of the first set of statistics encourage us to regard the alternate treatment as successful and while it is apparent that the alternate treatment tends to be more successful than the traditional treatment, we need to be careful not to make the unsupported claim that the pattern that we see could not be generated by chance alone. The difference in the degree of confidence that is shown in rejecting the null hypothesis in table 5-2 does not mean that we can assume that the two treatments differ significantly from each other; we know only that they differ significantly from chance. To answer the question of whether the student using the alternate imagery does significantly better than the student using the traditional imagery, we use the chi square in different way. Tables 5-3 and 5-4 test the differences of each sample from each other rather than from the presumption of a random distribution. If no significant
difference is found the samples, regardless of any perceived pattern, they are understood to be homogenous with respect to the variable measured.

Both measures find the alternate treatment to be superior to the traditional one in helping students to distinguish the eukaryotic and prokaryotic cells found in this example. We can conclude from these statistics that the two treatments can reasonably be seen to have different effects.

Interestingly though, the overall pattern cannot be read down to individual cases; not all of the differences between individual instances achieve significance. In two of the cases (C and E), the alternative treatment is associated with significantly better results, another case, b, comes very close; and the other two produce so little difference that there is less than a 5 in 100 chance that repeated measures would not show the result to be achieved by chance alone. Such results are difficult to interpret and are the usual basis for concluding that more research is necessary to determine the grounds for the perplexing differences found.

On one level, one can simply say that this uneven result at the level of individual images is the way that statistics work and the chief reason that we value them as tools. That is, small differences that are insignificant or of unreliable importance in individual instances can be shown to be related to a much more reliable overall pattern that can guide our practice. What is being demonstrated here is more on the order of concluding that classrooms of students taught in the same manner would benefit by instructional materials similar to those in the alternate treatment. But educational research is in the position of attempting to draw conclusions about what will be best for the students that come to us as individuals. In such a situation overall trends do not translate easily into the conclusion that particular students will be helped by the treatment.
The most straightforward interpretation is that the positive effect is relatively weak against the background of prior knowledge and experience that account for most of the difference from pure chance, and that the effect shows up only in the aggregate for that reason. On this account some measure of the power of the discovered statistical significance to account for the effect measured is called for.

Table 5-1 above details a power measure (W) developed by Cohen (1977) from the chi square statistic. Cohen's formula results in a W of .175 which is by Cohen's statistically defined standards a small effect. This represents a statistical conclusion which tends to confirm the idea that the failure of some of the individual images to achieve significant results is because the differential effect of the alternate treatment regime is small relative to large background factors and the traditional treatment.

This conclusion should give us cause for pause. If the demonstrated effect, however "real" in terms of significance, is small, should we advocate changing our practice based on such finding? Is the difference large enough to justify our changing our practice? I suspect that we should not be too quick to conclude that a statistically small result is necessarily an educationally insignificant result. Cohen himself cautions that his standard is statistical and that an "ad hoc" sense of magnitude "for a particular problem in a particular field" is called for when considering the practical use of results.

An argument for considering these results practically important can be constructed on two grounds: first, that the small statistical result is actually, within the particular field of education, a relatively large one, and second, that

---

270 Steele and Torries, Principles and Procedures of Statistics, 216.

271 Cohen, Statistical Power Analysis, 224.
this “instantaneous” measure does not adequately reflect the potential power of the strategy over time.

A reexamination of Table 5-1 reveals that those learning from the new, alternate treatment averaged approximately 82 percent correct, while those exposed to the more traditional treatment classified the cells correctly only 74 percent of the time. In a classroom this would translate into a situation where the average grade would be a B rather than a C on an identification test involving cells. This, I submit, would be an educationally significant result.

But this places the emphasis on a single test taken at a single time. Classrooms are places with a history. The recent emphasis on competency-based education, regardless of other problems it had, gained its credibility largely through the insight that those left behind at one level of instruction were likely to get further and further behind as they tried to build on concepts and learnings of which they had only an incomplete grasp. Any method which promises to bring along a greater percentage of the students at any one moment will, like a higher rate of compound interest, pay out disproportionately greater dividends at a later date.

Both of these rationales remain firmly within the standard justifications of testing and achievement. And this is as it should be; MacIntyre’s advice that a new theory must be judged, at least initially, by the standards of the theory it seeks to displace can be applied to instruction as well. To again cite:

The best theory so far is that which transcends the limitations of the previous best theory by providing the best explanation of that previous theory’s failures and incoherences (as judged by the standards of that previous theory) and shows how to escape them.272

---

The failures and incoherences of the current approach to instruction, as is well recognized from within that viewpoint, center around the problem that educational discourse discusses under the rubric of application. Students are not able to turn their school knowledge into a reliable basis for action in the world. The most common descriptive explanation has been that school knowledge is “about” the world while the more useful forms are “of” the world. By its nature this dichotomy bodes ill for the institution of schooling. Accepting such a dichotomy, educators either accept the general lack of utility of material learned in schools for the daily activities of life or attempt to erase the differences between schooling and the world. Neither road is, finally, palatable, and the choice offered is a desperate one.

The full benefit of the approach offered here, as has been argued earlier, lies in the ability to teach something different rather than the ability to teach the same thing better. It offers the possibility of teaching for the sort of perceptual competence which supports our lived experience without abandoning the institution of schooling. It offers an opening toward making full and conscious use of the perceptual mode of learning that supports the child’s easy and fluid recognition of her mother. This is the same ability that supports the expert’s recognition of a eukaryotic cell.

**Conclusion**

This chapter has served two functions: to demonstrate that the particular combination of connectionist and practice-based theorization suggested here has implications for the design of educational research, and to begin to show that practices based on this conceptual framework can be successful.
The design of the research tool itself as well as the analysis of the data that resulted from its use were intimately tied to connectionism and practice theories. A more standard approach to improving students’ the use of images would have presented prototypical images (as both classical, feature-oriented and prototype theories would suggest) the approach suggested by connectionism emphasizes multiple images, some pointedly marginal to the category. Similarly, the design practices drawn upon and the attempt to array them as a new genre of image usage would have been less consciously and less thoroughly arrayed outside of a theoretical context that emphasizes practice.

To the extent that this study has demonstrated that a treatment based on situated connectionism, which varies in notable ways from the traditional ways of presenting such material, can be successful this study indicates that such an approach is capable of generating new and successful approaches to designing teaching practices.

While these results are gratifying a short reflection upon the principles advanced in chapter 4 will reveal that at least two principles discussed early in that chapter have received short shrift in this example. Both activity and the temporal embeddedness of situated activity were stressed as essential to understanding the apparent stability of representation in a situated connectionist framework. These themes will be taken up in the explication of a more fully realized instructional treatment developed out of the emerging framework of this study in chapter six. There I will discuss a program I have developed which builds on the design features partially worked out in the present study, and will address the issues of temporal sequencing and activity in category formation. Developing a computer-based implementation was tempting for two reasons: to take advantage of the plasticity that the medium offers and to show that the computer medium need not be limited to the sterile
reproduction of older forms, e.g. electronic worksheets, for which it has become infamous. The most commonly cited example of this problem is the pervasive use of computers to implement drill and practice routines and to emulate worksheets. But this critique can reasonably be extended to even such an apparently unique application as hypertext which, as far as I can see, does not go beyond the tools available to the early French encyclopedists. Most hypertext links are simply more convenient implementations of glossaries and indexes and the more sophisticated (and rare) interdocumentary links are first cousins to the traditional reference footnotes.
CHAPTER 6
Designing a Programmed Environment to Induce Category Formation

In asking what computers can do, we are drawn into asking what people do with them, and in the end into addressing the fundamental question of what it means to be human.


The current chapter attempts to demonstrate that the positions developed in this work are generative. An all too often accurate observation is that some styles of theorizing make a “distinction without a difference.” Because one theme of this work has been to reconceptualize the efficacy of such educational practices as rule-based learning, repetition, and abstraction in light of the perspective being put forward, there is some danger that the reader may conclude that the theoretical position implies such a distinction without a difference. The computer offers itself as a particularly tempting medium for developing a demonstration of the potential difference that follows from adopting the perspective advocated here. It is generally agreed both to be under utilized and to have great potential in education. This potential flows chiefly from its flexibility, which allows the designer to develop theoretical


275 See particularly chapter 4 and the defense there of objects-of-knowledge reconceived as external symbols and rules as potentially pedagogically useful tactics which are, nonetheless, the product, not the producer, of competence.
implications which would simply be impossible in any other single media. In the present case both activity and time-based recurrence are crucial aspects of the position on learning and knowledge that has been built in this work. It would be very difficult to build both of these features into other media. The computer medium can be designed to provide "virtual" manipulation of external objects, which is analogous to the activities that form categories in our everyday experience. Similarly, temporal sequencing can be rigidly ensured, making certain experiences follow one another in time and ensuring the repetition of key activities that our earlier explorations indicated were crucial to building connectionist memories.

**The Challenge of Software Design**

Designing an educational computer program is a risky business. Much educational software has been roundly criticized for reproducing the errors already found in print materials. By and large this criticism has been fair. Both worksheets and programmed texts have been reproduced in this medium without any discernible improvement being achieved by the shift in display medium. Even more sophisticated software packages such as hypertext suffer from this critique. Most implementations of hypertext are more convenient forms of glossaries, indexes, and occasionally, the reference footnote. There seems to be little here that was not already incorporated into the dreams of the original French encyclopedists. Hypertext software that attempts to push these limits is usually considered "complex" and is condemned for not being "user-friendly." Indeed users do find them difficult to use and often find themselves lost deep in a string of interconnections that has led them irretrievably far from the text that occasioned the side trip.
But the computer remains extremely tempting to the designer. The underlying digital simplicity enables an almost infinite flexibility of possible uses: calculate \( \pi \), expand a fractal, lay out a newsletter, balance your checkbook, manipulate the tonal range of a photograph, or write a novel. The plasticity of the computer, its chameleon-like ability to be what the user wants it to be, is very seductive.

The contradiction is apparent: if computers are valued for their plasticity, why are they so conventionally used? I will argue that this happens for two interrelated reasons, both of which connect to themes in this work. First, designers, like others in our culture, mistake the nature of learning and hence of knowledge. They begin with the idea that knowledge is an object and that it can be cleanly transferred and unproblematically transformed by rules analogous to those of logic and grammar. Such a conception seriously constrains the possible ways in which knowledge can be represented to ways which have already been widely explored in conventional print design.

Second, software designers—in part because they unreflectively accept the idea that knowledge exists in this objective way—are unable to adequately conceive the role of practices in learning. Without an understanding of the role of practices in learning, they cannot grasp how signifying practices can, through connectionist styles of associative learning, be built into experience. Without any way to build new practices, they must fall back onto practices which are already understood among the literate population. This, I suspect, is why the possibilities of hypertext are so conventionally implemented—the typical user finds truly different practices confusing, and the designer cannot conceive of teaching the learner a new, more useful practice during the use of
the material since the designer does not recognize the role of such practices nor understand how such practices have lasting cognitive effects.276

The real task before the designer is to grasp the possibility of creating new, more useful practices and the larger structures within which such practices signify. Such larger structures will be characterized here as genres, following Bakhtin. Such practices and genres would need to be taught to the user or student; the designer would need to recognize that their meaning would appear only in the course of the users’ interaction with the material and design the experiences to facilitate this. Only on such a basis can truly different tactics which make use of the computer’s plasticity be successfully assayed.

A Theoretical Background for Category Learning

This program explores the possibilities for designing learning materials from a connectionist/practice standpoint. It attempts to teach students to bifurcate a series of cells into two categories: Eukaryotic and Prokaryotic. The problem is both apt and ironic. Apt because this is a category that is specifically scientific, logical and “hard.” If the Wittgensteinian and pragmatic characterization of categories, which emphasizes the formed-in-use character of the categories of our everyday use, is to be shown inadequate, the most likely candidate is a category which is self-consciously objective. However, if connectionist speculations are to be taken seriously, even such categories must finally depend

276 The original conceptions of hypertext suggested much more radical possibilities, possibilities, including mimicking the associative qualities of human memory. However, even here the basic image was that hypertext would be easy to use because it projected a “natural” human form on the world—no consideration is given to seeing the reciprocal relation between cognitive artifact and human learning. See: Theodor H. Nelson, Dream Machines (Chicago: Nelson/Hugo’s Book Service, 1974). Republished in 1987 as Computer Lib, Redmond, WA: Tempus).
on the external symbology of embodied practices in order to be learned and to remain stable over time.

The irony lies in attempting to design a program which teaches a user to experience the category as a bifurcation, an either/or, using design tools, semiotic and practice-oriented theories, and connectionist learning theories which all point toward the idea that our categories are not and indeed cannot be so discrete. But this is the sort of irony with which our culture has already become comfortable: witness the way in which semiotic theories of meaning find their most pervasive influence in the construction of largely meaningless advertising.

Practices and Genres

Genre, as used here, refers to Bakhtin’s valuable but somewhat unfamiliar characterization of the term. In Bakhtin’s usage, genre refers to a stable type of utterance used within, and to signify the presence of, a particular type of human activity.\(^{277}\) Such genres can be simple or complex; complex genres are those composed of simpler genres. Bakhtin suggests that complex genres are made up of simpler elements which can be used in various patterns to make up different genres. This leads Bakhtin to conclude that the elements take on their compositional character only as part of the whole utterance that determines their interpretation. Here my usage will diverge from Bakhtin’s, largely in the interest of clarity. Rather than talk of simple genres which compose more complex ones, I will speak of practices which compose all genres—including ones which Bakhtin discusses as simple. Interestingly in this context, Bakhtin

\(^{277}\) This usage is pointedly broader than the more common use of “genre” to denote a style of novel—an interpretation that Bakhtin specifically disavows. See M. M. Bakhtin, *Speech Genres and Other Late Essays*, ed. Caryl Emerson and Michael Holquist, trans. Vern W. McGee (Austin: University of Texas Press, 1986), 6061.
was trying to establish a taxonomy of genre types on a unified, elemental basis—a foundationalist project. I do not share this task and suspect that it obscures the utility of Bakhtin’s formulation. Establishing a formal boundary between simple and complex genres would be at least as difficult, I suspect, as drawing an uncontroversial line between eukaryotes and prokaryotes. I hope a more productive path will be to see each signifying utterance that we pick out as composed of practices—practices which in other times and for other purposes could be analyzed as utterances with their own constituent practices. What I will want to retain is the understanding that the meaning of each individual practice can be usefully understood only against the background established by the whole genre understood as a purposeful utterance. Thus the overall page layout can be considered the background "genre" against which we understand the efficacy of such practices as grouping and spatial opposition in explaining the page’s ability to enable learning. For other purposes the overall page layout itself might be considered a practice pointing to a bifurcating category formation in an attempt to design a new genre of textbook design. The chief implication of this point of view is that any actual genre can be considered as composed of a hierarchical series of practices. How we choose to describe any concrete practice—as a practice or a genre—depends upon our purposes in taking up the analysis.

The problem addressed in chapter 5 involved an attempt to array already established practices in a new pattern create a new genre for textbook images which would be more effective than the traditional single, prototypical illustration. As we saw in chapter 5 there is some hope that the design alternative we explored might be a successful candidate for this role.

278 See: Bakhtin, Speech Genres, 64, and note ‘a’ on that page as well as editor's note 4, p. 101, for a discussion of this project.
It may be helpful to recall that the practices from which the candidate genre discussed in chapter 5 were designed were ones which were already familiar to the typical user educated within Western conventions of print design. Unity was signified by placing images together closely, more closely than convention would allow conceptually separate objects to be arrayed. Equality of status was indicated by making the two groups the same size and placing them on the same level relative to the horizontal axis of the page. Finally, position between the two groups was established by placing the two groups on "opposite" (this convention is embedded in our language as well as our page design) sides of the page. These practices, it was argued, were not natural, but were the result of each individual's history of interaction with texts which assumed the meaning of these conventions.

Practices and practice-like concepts are both widespread and variously defined—as the varied background of theorists discussed in chapter 4 makes apparent. In the present context, viewed against a background of connectionist learning and looking toward their participation in genres as discussed above, practices will be particularly understood. Looking back at connectionism we see a practice as the sort of recurrent, patterned, relatively stable event which, when recognized, can make a difference in the ongoing activity of the perceiver. It is this sort of event, an event which recurs often enough against varied backgrounds to be perceptible as a relatively stable object, which is easiest for the human cognitive processes to register and recall. Looking in the other direction, we realize that such practices appear in the context of larger human enterprises which constrain and direct the sense that can be made and the range of reactions that are appropriate in a situation. Adopting Bakhtin's terminology, we can call these constraining contexts genres. It is this context-defining genre within which a particular practice signifies.
In this work we have returned periodically to the seminal example of a child learning to recognize her mother’s face. She repeatedly encounters the smell, the tactile softness of her mother’s clothing and touch, the shapes of nose and cheek, the soft brown eyes, and the motion of reaching out and the sensation of being picked up when she is uncomfortable, when she is cold. The recognition, the activity, and the result are indissolubly tied together. Each separable “practice,” be it recognition of her mother’s face, the activity of reaching, or the socially mediated result of being picked up and made warm are parts of larger genre activity for which we have no ready name. Let us call it a “comforting-when-cold” genre.

The constituent practices of this “comforting-when-cold” genre could be recruited to other genres where their contextualized meanings would differ with the different genre activity that they helped constitute. It is important to note that such genres, following a connectionist account of learning, are not classical schemata with the structural assumptions that so bedeviled artificial intelligence theorists. Genres, unlike schemata, are flexible, learnable patterns without definitive fixed boundaries or a fixed internal pattern which controls their expression. This is the type of learned regularity supported by the networked relations which constitute human cognition. Such a formulation also differs from more traditional approaches by drawing no in-principle distinction between the activity of the child and the activity of the mother in constituting the genre activity as a whole. As discussed in chapter 4, connectionist memory remains stable only in a stable environment. This approach eliminates any easy

---

dichotomy of world and mind. Such mental activity as exists must exist between persons engaged in activity as well as within the individual bodies of the participants. The approach to practice and genre outlined here gives substance to the pragmatist description of mind:

The locus of the mind is not in the individual. Mental processes are fragments of the complex conduct of the individual in and on his environment.\(^{280}\)

With a connectionist approach to learning, with the tool of practices that are learned through connectionist processes, and with genres which are composed of these practices, we can approach an explanation of the social nature of cognition that is not merely descriptive but which refers instead to a material set of processes, both inside and outside the head, that can subserve a theory of instruction.

Practices and genres as discussed here rely on two factors which were only marginally explored in the eukaryote/prokaryote research discussed in chapter 5. The research instrument was a largely passive and somewhat atemporal tool. There was little extension in time over which to develop the recurrence necessary to basic forms of connectionist learning and there was little opportunity for the activity of classifying the cells. Both of these characteristics are hard to build into the largely static print medium.

The plasticity of the computer medium makes it possible to build a much more interactive and temporally extended text. Both activity and temporal recurrence are crucial features of the position developed in this work. Demonstrating that they can be usefully integrated into the design of instructional materials is central to presenting a case that a learning theory

based in connectionist ideas can be educationally useful. It is to these issues that we now turn.

Activity and Temporality

The intertwined issues of activity and temporality which appear periodically throughout this work are brought together here. In chapters 2 and 3 we discussed the atemporal nature of the analytico-referential discourse as exemplified by the static structures of expert systems. The "brittleness" of such systems and their failure to perform as expected derive from their inability to learn, an inability to alter their fundamental structure as a result of experience. It was argued that this is not simply a technical problem but a profoundly conceptual one: the division of form and content is implicit in the discourse from which such work draws, and this generates intractable problems in practice. Working from the underlying metaphor of timeless, placeless structures of knowledge—the formal, logical forms familiar from geometry—particular content was understood to fill only the slots left by more foundational logical work.

The connectionist approach to learning presented in this work as a way of overcoming this discursive impasse is, we discover upon closer examination, profoundly dependent on the particularities of both time and place. In the guise of context-dependence, the implications of particularity in learning had been conceptualized as a problem. But a connectionist approach to learning makes context constitutive of competence and not an impediment to it. The connectionist understanding is that all learning is learning of just this particular kind.

The central property of network architectures in regard to learning is pattern completion. Pattern completion as an emergent property of network
architectures can occur only through a history of repeated interaction with regularities in the world. As we saw in the Sharks example, the order in which a set of information is revealed does matter; the sequence in which information is presented biases the eventual output of the network and may either facilitate or prevent the network's settling on the correct answer. This time ordered sensitivity to initial conditions is a key source of the flexibility and error resistance which recommend neural nets to many researchers. The passage of time and a history of repeated interaction with a category of events is essential to the sort of learning proposed by connectionist theorizing.

But the valued flexibility of network architectures is purchased at the price of unstable representation within the brain alone. The pattern of relationships which constitutes the connectionist analog to memory will degrade without continued, stable interaction with worldly situations which sustain the pattern. This led, in chapter 4, to a broad exploration of the ways in which the necessary stable categories of the world are largely socially ordered. The speculation of Rumelhart et al. that connectionist symbols were largely "external symbols" was taken up and discussed in the context of Vygotskian concepts and the practices of situated cognition. Sustaining order in the social world, and especially in the "higher" realms of abstract thought such as mathematics, depended upon manipulating material, external symbols. The example of three-digit multiplication accomplished through pen and paper manipulation of such external symbols—which are and remain difficult to manipulate in the head—was explored as an example of such external symbols.

The program discussed below attempts to extend the strategies developed in chapter 5 to program design. In this example, intended to explore the

\[281\] Though recursive networks were discussed as one way that time dependent comparisons of nonsimultaneous events could be accomplished.
possibility of creating a more effective practice of illustration for the textbook genre, the limits of the textbook genre needed to be respected. In the context of a computer program, however, where both active manipulation and a temporal recursion can be instituted, different possibilities appear. A learning material can be created which mimics more of the characteristics associated with everyday experience. To the degree to which instructional materials can be created which reproduce the elements of experience which make it effective a learning environment which incorporates such materials should share in that effectiveness. The program described below attempts, based on the theoretical positions developed in this work, to extend computer-based instructional design in this direction.

The Category Learning Program

In emphasizing temporal sequencing and the student’s activity in this discussion, a very real difficulty is introduced into the presentation. Succinctly: a text like the one you are reading now can only talk about time and activity—it is especially poor at demonstrating how these factors might be important. While this is often a problem in written texts, it is particularly poignant when a major point of the work under discussion is to point out the inadequacies of such texts and to suggest an alternative. Consequently, the reader is encouraged to make use of the disk found in appendix A: Category Formation: An instructional program. Running through the program a few times should provide an experiential background which will make the following description more meaningful.
The Program’s Design

This program is intended to demonstrate the utility of combining a connectionist approach to learning with situated insights into the particular practices that constitute knowledge. Design features linked to connectionist learning and practice-based knowledge broadly conceived will be discussed first.

Connectionism and This Program’s Design

Connectionist architectures are built around a broad conceptual model in which a system "settles" into a solution based on "pressure" exerted by a dynamic transformation template for correlating external data with internal response; eventually the system settles into a stable response pattern which corresponds to differences in the problem set it is learning to classify. The template has been conceptualized as a teacher and as representing environmental "reality." In either case each act of categorization by the net is coupled with immediate corrective feedback based on its last response. In a connectionist framework this is not a useful adjunct that makes learning easier; it is an integral and necessary part of each act of attempted classification.

The Category Learning Program reflects this imperative by making each classification an act which appears, at least from the user’s point of view, to contain its own confirmation or rejection. The user picks up the image that is to be classified and drops it on the group which the user thinks it most resembles. The program either rejects an incorrect categorization by bouncing the image back and presenting the choice again, or it accepts the correct classification by including the new image in the array used to define the category. The activity of categorization has immediate effects in the experiential world afforded by the program. That "world" either accepts or rejects the categorization.
Connectionist modeling indicates that features are not the basis of recognition—an assumption made by folk theory and many cognitive theories. A network that has learned to correctly parse its "sensory" experience does so on the basis of the overall pattern of relationships in the data. Connectionist nets do not discover a set of necessary and sufficient features by which they then correctly classify the input. They are holistic pattern completers. Generally, any feature recognition nodes or grouping of nodes that observers pick out to talk about are, in fact, not essential to the classification. (I would contend that this is true of human perception and discourse as well.)

The Category Learning Program implements this approach by deliberately avoiding giving the user the features traditionally used to teach the category. It also attempts to discourage our common habit, based on our intellectual practices, of searching for these discrete critical features.

Connectionist modeling also throws doubt on a alternate cognitive theory that recognition is based on generalization from prototypical (central or best) examples. The early use of a training set limited to exemplars makes it difficult to acquire outlier members reliably. The network encodes the pattern of relation that characterizes the exemplar and classifies all remaining members of the problem set by that limited pattern. Either over- or undergeneralization may result, depending on the particular problem set. The more difference that the category actually contains, the greater this problem becomes. The Category Learning Program does not use exemplars and chooses its examples from a broad range of taxonomic categories and degrees of difficulty.282

---

282 Differing degrees of difficulty were established during the design stage of the project discussed in chapter 5 by asking microbiology instructors and science education students to rank order the available pool of micrographs in order of their difficulty for the student.
The network models which are the basis of connectionist theorizing also depend on a recurrence of events over time. The mother's face reappears in various contexts and against multiple backgrounds but always means, among other meanings, warmth and comfort. The detachable quality which results cannot appear when the event is either a constant background feature—and hence cannot represent a notable difference—or is so intermittent as to fail to correlate with any useful difference. It is at this point that connectionist models force a look outward, toward the socially grounded regularities of practices discussed below.

In sum, the Category Learning Program is designed on connectionist assumptions concerning the nature of learning. This stance differs critically from a traditional view of learning implied by this stance is that category/object learning depends on direct confirmatory or nonconfirmatory linkages to the world and holistic pattern recognition which is not based on prototypes. Both of these positions turn our attention outward, toward the socially organized regularities that constitute the objects of our experience.

**Practices and this Program's Design**

Practices as used here are socially organized and signified regularities further constrained by the assumptions inherent in connectionist learning. This category includes objects and more temporally extended events which are perceived by the user as separable from the background in which they appear. They correspond closely to the design professional's concept of design elements. It is difficult to overestimate the importance of practices in the successful completion of the tasks set up by the world this program makes available. Even so mundane a task as “picking up” an object and moving it is mediated by clearly social, learned practices such as the “click and drag” feature common to all mouse-driven computer interfaces.
The Category Learning Program trades on a whole set of such already given social practices. These practices are not “natural.” They are simply so much a part of the person-acting that they are not noticed—but their effectiveness depends entirely on the history the user brings to the task.

For instance, representing unity by grouping things that we intend to represent as a category is an old and honored practice. Most folks looking at the initial learning screen of the program see three things represented, not ten, because our history of interaction with similar objects grouped together in a deliberate pattern has meant this in our past experience (see figure 6-1 below: “Initial learning screen”).

![Initial learning screen](image)

**Figure 6-1: Initial learning screen**
Or consider how this organization structures difference: the two categories are on opposite sides of the page, denoting opposition. Similarly the unnamed group is beneath and between them, and contains fewer members. We are not predisposed to assume that they are opposed in the same way that the screen design, and the user's history, encourages the viewer to take the upper two categories.

The plasticity of the computer also makes available practices using motion and sound that are not usually available for use in educational material. The practice of putting members of a group "in a pile" and of reacting to correct choices with sounds of approval and admittance to the pile, while reacting to incorrect choices with sound of disapproval and a rejection from the pile (forcing the student to try again) are equally practices, whose meaning is established in the user's history and only drawn on here.

At the point where the learner has to act in order to categorize, he or she is presented with one example of each of the two categories to be learned. This effectively means that each selection enacts the bifurcation that the program hopes to teach. Bifurcation is a familiar way to form categories—and the dominant academic practice.

These practices build knowledge into the world so that the learner will recognize a particular pattern in the material given. The activity of sorting in this way—without rules, exemplars or critical features—is crucial to forming a category. The practices discussed constrain the activity of choosing which image goes in which group by trading on knowledge already embedded in the situation in which the problem appears. Unity, opposition, bifurcation, grouping, and approval/rejection sounds all constrain the possible pattern—out of the plenitude of possible patterns—which the learner will settle on. It is
highly unlikely that a user given a screen full of unordered, uncommented images would discover anything like the taxonomic distinction this implementation attempts to teach.

Such a student would have only pattern recognition to rely on and pattern recognition, unallied to our socially constructed history of signifying practices, is not enough.

Constrained by what we know about both category learning and practices which situate knowledge in the world, an approach such as this enables the teacher to design a learning experience which makes available to the student the categories through which the community of use (in this particular case, biologists) structure their world. Perhaps the best way to demonstrate this in the current, textually limited, context is to lead the reader through a brief recounting of the experience of using the program which emphasizes the various theoretical constructions that are embedded in the program.

Interacting With the Category Formation Program

The program opens with a screen which briefly explains its purpose and asks for the user’s name and other useful data—versions for use in a particular context could ask for the class currently enrolled in or about prior experience in the area.

Starting with the initial learning screen, the student encounters a layout composed of two groups of four images (representing the category) and a group of two images (unknown cells to be classified) and text which calls for her or him to move one of the two unknown cells into either the eukaryotic or prokaryotic group by clicking on an unknown cell and dragging it to one of the groups (see figure 6-2: Moving an unknown to a category).

If the categorization is correct, an “ahh” sound is played and the successfully classified unknown replaces the member of the category that it was dropped
on, becoming itself part of the new set of images of the now reconstituted category which presents itself to the user during the next iteration of the categorization (see figure 6-3: A new grouping results). A new set of unknowns from which to select is presented.

Figure 6-2: Moving an unknown to a category  Figure 6-3: A new grouping results

If the categorization is incorrect, an "uhn" sound is played, the image is "rejected" by being bounced back to its original position, and the same choice is offered again.

This activity is repeated, seven times in the current instanciation. This extension in time allows the learner an opportunity to encounter this occasion to categorize as a regularity in the socially constituted world. But, as in the world we are used to, it is not encountered in exactly the same way each time. The program's design allows the user to pick up and categorize either of two unknowns each time a new set is created. The learner can put the chosen cell into one of four spaces. In categorizing the first set there are eight possible different outcomes. Starting with one of the two category sets altered by the previous choice in one of the eight ways, there are eight new choices for each of the eight branches established earlier—at this point sixty-four possible
outcomes. By the time the user has worked through seven sets, there are over two million possible different paths through the alternatives. From a user’s point of view, this is effectively infinite—no user will ever see all the possible combinations. While the number of possible valid solutions is very large, no single action faces the user with an unmanageable array of choices. One of two unknowns is selected. The selected unknown is dropped onto one of two groups; if the attempt to categorize fails, the user knows to try the other group. The activity is simple; only the result is complex. Within the constraints set up by the program, the learners can order their own experience and set up groupings of images that best aid their learning.

Finally, at the end of the set, the program calculates the percentage of correct categories, makes an encouraging remark based on the learner’s success, and allows the learner to exit this portion or work through the examples again. A history of the user’s interaction records each choice, how long it took to make the choice, and where on the screen the unknown was dropped. This entry is extended each time the user returns to the machine and is entered under his name with a different date and time. This record could be used for further research and for diagnosis and evaluation.

Following the analytical frame sketched above, the practices constituting an activity genre of categorization are such pre-understood elements as grouping the cells to form a category, denoting membership by moving a cell into the group which represents the category, moving by clicking and dragging, including particular unknown cells to denote success in categorization, and hearing an “uhn” sound to denote failure. This genre, once its constituent practices take on regular meaning through participation in the whole activity, can become a practice available to teach similar categorization, it can be treated as a practice of bifurcating categorization to be arrayed within a larger genre of
computer based instruction. For instance, the category "contains nucleus" or "contains vacuoles" can be established using exactly the same series of practices in a larger instructional genre of computer-based cellular learning.

Note that each successful act of categorization reconstitutes the category that the user experiences and that the final distribution of images emblematically constituting the category is under the control of the learner, who may choose to "keep" a series that retains the greatest variation or may try to focus on perfect examples or may simply replace the most similar member with the new nominee. This "reconstruction" is an externalized example of the connectionist thesis that each act of recognition is an act of recategorization. The act of categorization physically changes the category to which the particular example is assimilated.

Note as well that category learning, as exemplified here, is a unified whole ranging from perception to pattern-matching thought to action. Once well-learned, the activity is a seamless one. The unknown is recognized as a member of one group and the activity of joining it to the category fluidly follows. In a description of the fluid competence of the expert categorizer, it may appear that perception initiates the sequence, but historically, activity and the world's reaction to that activity—whether it rejects or accepts the choice—provide the ground for pattern matching between cases treated similarly; the perception of a particular cell as a member of a category follows. Finally the unit of analysis is the whole activity itself; further breakdown may be useful in discussions, but any particular perception, thought, or action only makes sense in the context of the whole activity.

In some ways this program presents a particularly constrained virtual world to the learner. Interestingly, it is through the constraints and not through the millions of choices that the program proves itself useful to the learner. The
current rhetoric of virtual reality follows the now-traditional interpretations of hypertext in finding value in the infinite possibilities that can be made available. The account given here challenges that assumption and suggests that constraint is more important than infinite choice in constructing a useful learning environment.

Conclusion

This chapter has attempted to demonstrate that the theoretical positions developed in this work lead to significantly different design principles which can be implemented in the new, plastic, computer-based media. It has focused on temporal extension and activity as two crucial elements in a reconceptualized approach to learning that are not available in the atemporal, passive textbooks which dominate our current learning materials. The implication being developed is that if this approach to learning is correct in any large measure, education needs to pay more attention to providing temporally extended, recurring opportunities to actively participate in creating the objects of learning in material ways in the world. This constructivism is a thoroughly social one: it claims that constructing a schema in the privacy of our own head is precisely what we do not do. Rather, we learn to recognize pattern though our interactions with a socially patterned world from which we adopt and adapt the objects of our experience.
CHAPTER 7
Situating Connectionism:
Summary & Implications

As long as our brain is a mystery, the universe—the reflection of the structure of the brain—will also be a mystery.

—Santiago Ramón y Cajal,

We are what we repeatedly do. Excellence, then, is not an act, but a habit.

—Aristotle

I open this chapter with two quotations, quotations which symbolize the two factors around which this dissertation has emerged. In the first, Ramón y Cajal, the seminal Spanish neurologist and philosopher, remarks on the connection between understanding the brain and understanding the world. Quite rightly he claims that we cannot understand the world until we are able to grasp how we grasp it, and that understanding the organization of the brain is central to this understanding.

In the second, Aristotle declares that habit, not will, is the foundation of the self. In this he is followed by the pragmatists who shared his understanding of the role of habit. Viewed from the standpoint of the community, habits held in common are the practices that have been central to the social portion of this work. Socially established "habits of mind" are the glue that allows us to coordinate our actions and are fundamental to our sense of identity.

Taking these in concert implies that to understand the universe—including ourselves—we will need to understand how we learn the habits of mind that
are formative of the world in which we live. This reciprocal relationship between how we learn and what we learn is one of the deepest mysteries of our humanity. Recognizing this relationship enmeshes us in a recursive, hermeneutic web that emphasizes human activity in meaning-making. It uncovers, as well, limits and possibilities for our own activity. We can begin to see that our habitual understanding of how we understand our selves and the world is a critical linchpin in our construction and our society's construction. Pull that linchpin and disorganization is likely to result—but a disorganization which holds the possibility of an emergent new order which we may more readily turn to our purposes. In this dissertation those purposes are educational ones; I have contended that the older habit of understanding ourselves as logical machines has not served our students well, that in fact, if too strictly applied, it limits their ability to become competent actors in their world.

Critiques of logicism are a feature of our intellectual landscape, as the continuing romantic reaction to Enlightenment rationalism demonstrates. But to stop at critique is both irresponsible and, more to the point, ineffective. We need to present an alternate, more humane and useful story for the one that we have grown used to telling ourselves.

This dissertation is largely dedicated to laying the groundwork for telling the newer story which will allow us to reconceptualize ourselves, to change how we understand the very way we learn.

As laid out in chapter 1, pursuing this goal entailed a triadic strategy: first, to focus on learning as a material process and to avoid descriptive theorization where possible; second, in view of the pervasive character of current habits of thought concerning learning, to explore a broad array of evidence across historical, biological, social, and even computational fields that support the thesis that the older position is unlikely or impossible and that the connectionist
position on learning better fits our current understandings; third, in order to
avoid the trap of offering a theory without clear implications for change, to
demonstrate some of the implications the emerging perspective would have for
educational research and practice.

**Reviewing the Path**

This dissertation has roughly followed the path of laying out a context,
examining the particulars of connectionist learning and situated knowledge,
and offering two examples of how adopting the new perspective could make a
difference in the way educators do research and design instructional material.
Along the way I hoped that a broad and convincing case would be built that
our received conception was untenable and that an alternate understanding
was viable.

In chapter 1, I contend that the descriptive style of theorization that
dominated learning theory does not, in fact, produce a theory of learning
because it does not attempt to account for the material process of change which,
by any non-dualist account, must be the site of learning. What passes for
learning theory is better understood as a theory of knowledge acquisition, and
confusing the two has resulted in a blind spot in educational theory. The
metaphor of a blind spot in our vision is a fairly precise one: one does not notice
a blind spot until a special attempt is made to demonstrate that something that
you don't normally question is actually not there. We usually fill in the absence
with the next nearest thing—in this case with theories of knowledge
acquisition—and fail to notice that nothing is actually there.

In chapter two I attempt to produce an awareness that the particular logical
image of learning with which we have filled our blind spot in not natural, that
it is the product of particular historical circumstances. Reiss and Toulmin’s account of the rise and elaboration of the current discursive structure focuses on issues of how people demonstrate, to themselves as well as others, that they know something. When reference is unclear, as Reiss claims it was in the 17th century, a crisis of representation results. Reiss shows us that the current discourse arose to settle the problems of representation which grew out the collapse of an older way of making meaning. The solution that evolved to cope with that collapse postulated a formal congruence between three systems: language, logic, and material causality. Toulmin’s attempts to understand how our cultural systems have developed make it clear that formal logic has come to dominate our approach to understanding. Descartes’ crucial introjection of logic into the mind and the accompanying abstraction of linguistically stated, geometric reasoning into the mathematical formalism of the calculus made possible the rise of securing certainty as a central western project, a project to which both Dewey and Toulmin object.

The latter portion of chapter 2 takes the modern example of cognitive science and traces the way in which the image of thought as logic has been given material instanciation in the development of the program of artificial intelligence. I point to the way in which artificial intelligence was born in a recursive relationship between the Western image of thought as logic and the development of computers as material instanciations of logic. Following our cultural history, the computer was built on the image of the reigning, logical conception of thought, and our understanding of ourselves was recursively affected by the presence of a machine that computes. By making the assumptions of the traditional position both material and explicit, artificial intelligence constituted itself as an un-selfconscious test of the discursive position that human thought is logical. If we contend that human thought is
logical then computers ought to be able to think as humans do, for as Turing demonstrated, anything that can be done on one computer can be done on another. As the experiment has developed it turns out that artificial intelligence actually disproves this contention. The logicist model of the serial digital computer is inadequate as a model of human thought. The human brain is neither fast enough nor large enough to be built on the serial model. Models built to mimic logic are logically capable in ways that no person is and fragile in practice in ways that even the simplest human is not. But the most telling critique is that such models are incapable of learning, and without being able to learn remain literal, dependent simpletons.

Connectionism arises in the attempt to solve the problems of the serial model. With very little consciousness of its potential discursive role connectionism posits an alternative account of meaning-making. It provides ways out of the difficulties produced by the previous conception—it is fast, capacious and robust. Most importantly for the educator connectionist systems shine at learning tasks, are able to perform recognition tasks far beyond the level of its serial competitor, and provide a ready model for content-addressable, associative memory.

Chapter 1 attempted to orient the reader and chapter 2 worked to build a background against which to understand the emergence of connectionism. Building on this chapters 3 and 4 are engaged in the task of laying out in some detail the nature of connectionist thought concerning learning, problems with connectionism considered from the point of view of instruction, and the need to bolster connectionist learning with a theory of knowledge based in social practices.

Chapter 3 explores the characteristics of the network architecture associated with connectionist models with an eye toward detailing the way in which
particular, material models account for associative memory, simple learning, temporal learning, and abstraction. The rather intricate exposition of simulations and particular learning architectures serves to ground the claim that connectionist models are plausible models of human learning. The exploration of this claim, however, exposes weaknesses as well as strengths. Network architectures, at least in the forms which yield the valuable characteristics of associative memory, fluid category formation, easy recognition, and learning, are inherently unstable. Although some researchers explore solutions strictly within the resources of connectionist architectures, I suspect that the stability of representation is finally a function of the patterned social interactions that sustain and construct individuals. The issue of knowledge, dismissed in earlier chapters as mere theories of knowledge acquisition, has returned to haunt a theory of instruction based on connectionism.

Chapter 4 takes up the problem of developing a theory of knowledge which is adequate to the needs of instruction. Attempting to understand knowledge brings up the representation debate in philosophy and cognitive science and the opposition between logical and imagistic approaches to human representation. Connectionist accounts propose a third path, offering the infinite generativity valued by logical accounts value and a material process by which the effects that interest the imagist camp can be generated. But connectionism asks both sides to give up something as well. Logicists lose, well, logic. They can no longer appeal to the sort of categories which support classical concepts of certainty. Categorization and human reason are painted as fully contextual. The imagists, on the other hand, give up the central, definitive prototype for a diffuse, contingent category.
Connectionist representation is so unintuitive that accepting its story tempts one with the possibility of simply declaring that stable knowledge such as the tradition gives us is illusory. But a commitment to teaching constrains us. Without a way to conceptualize knowledge, connectionist learning is useless in an education context. Situated perspectives are offered as a way to understand how stable symbols can be located in interaction with the world rather than possessed in the mind. The intersubjective habits which constitute practices are proposed as the location of the sought-for stability. Following out the insights of Vygotsky and Dreyfus, both the objects and the rules of the classical account of knowledge are reconstituted on a connectionist basis. Taking this path allows us to understand the effectiveness and ineffectiveness of teaching practices associated with defining objects of knowledge and rules of relation. By shifting the basis for understanding these phenomena from logical form to social practices, we gain the ability to predict the effects of our practice as teachers more accurately.

In chapters 5 and 6 the focus shifts to utilizing the framework built in the preceding chapters. This shift is motivated by the suspicion, raised in chapter 4, that connectionists perspectives only provide a different basis for justifying present practice. Were this the case, no matter how "true" the approach is, it would be ultimately sterile in practice. Chapters 5 and 6 attempt to show that the distinctions proposed do make a difference and develops some of the differences that flow from taking the stance advocated in this dissertation into concrete demonstrations.

The implications of adopting a situated connectionist stance for educational research and the design of instructional materials is the focus of chapter 5. In it I explore the implications of combining connectionism and practice theories in designing materials that will help teach students to distinguish between
eukaryotic and prokaryotic cells. The eukaryotic/prokaryotic distinction is interesting because it is outside the normal range of experience—most images of these cells are taken with electron microscopes—and is therefore relatively free of prior learning. It is also interesting because the ability to recognize these cells as members of either the prokaryotic or eukaryotic kingdom is a good example of the practical, perceptually-based knowledge that supports the diagnostic competence of experts such as doctors and medical technologists, and is precisely the sort of knowledge that some advocates of apprenticeship claim cannot be adequately taught in schools.

One place in which connectionist theorizing differs from its more traditional competitors is in the fundamental question of how we form categories. In connectionist accounts categories are based on multiple, concrete experiences which are structured to lead the individual to classify difference as the same—a perspective which is at considerable odds with the classical account, which assumes that categories are built on the basis of the pre-existing similarity of distinctive features. Connectionist modeling implies that using single “best” examples which exhibit clearly the features associated with classical category theory is not the best way to form categories which are useful in the fluid, if messy, way that experts display. Instead, simulations imply that categories are most quickly and accurately formed by using multiple members of the category which are deliberately weighted toward the inclusion of marginal members. This follows from the practice of emphasizing boundaries (rather than central membership) which is implicit in a connectionist understanding of how categories are formed.

The work of designing actual materials proved enlightening. In doing this work I fell back, almost automatically, on previous experience in graphic design. In graphic design there is the assumption that the role of the designer is
to present the meaning of the piece in such a way that the effect is perceived, effortlessly, as "natural" by the reader. To achieve this natural effect the designer, paradoxically, has to be quite conscious of how naturalness is constructed. The experience of finding myself almost automatically arraying social practices with particular meanings established in other, quite different contexts, in the service of teaching bifurcating distinctions based on connectionist assumptions, brought home with force just how important practices actually are in communicating meaning. Without these practices it would have been impossible to design materials which taught the distinction effectively. These practices were the prior learning upon which the successful presentation was based.

The final materials, which used multiple examples which included marginal members, employed the conventions of page design to indicate which elements were to be taken as similar and which to be read as opposed. A simple exploratory experiment was performed and I analyzed portions of the data for differences between the situated connectionist materials and those materials derived from a traditional text. Gratifyingly, since we were introducing new practices and unfamiliar ways of "reading" the images, the newer materials exhibited a statistically significant, positive difference from the traditional materials.

Although the design of materials for the experiment was inspiring and the result reassuring, it was also frustrating. I became acutely aware of the constraints of the one-shot, essentially passive paper-based medium on any design of materials inspired by connectionist principles. Connectionist principles include a strong emphasis on the repetition of patterns to be learned, a stress which places a special value on learning over time. Connectionist principles also lead to viewing the activity of the learner in altering the state of
the learning materials as important in its explanation of how the differences that characterize concrete, particular objects come to be recognized as single categories which we take as objects of knowledge (such as "chair"). Chapter 6 pursues these issues into the realm of the design of educational computer programs. The computer tempts the designer with its almost infinite plasticity. It is possible both to control the way in which the user repeats his or her experience of the material over time and to allow for, and even insist on, the active manipulation of images presented on the screen.

Reflecting on computer design, I was led to adapt Bakhtin's concept of the genre for its clarifying insight into the dependence of the meaning of any particular utterance on the activity-signifying genre in which it occurs. No experimental data is presented with this program; instead, this portion of the work emphasizes the ways in which the situated connectionist position developed in this paper can lead to the design of educational materials and particularly educational programs which do not simply reproduce the characteristics and assumptions associated with earlier media. The program developed in chapter 6 is proposed as a candidate genre for making the bifurcated distinctions common in educational practice.

In this dissertation, then, I have attempted to show the reader a path which will lead to a reconsideration of the way he or she conceptualizes learning. I have tried to show how learning theory is a problem in education and is not to be taken as an unremarkable background factor to our practice. I have emphasized that this is not a matter of simply exchanging a bad theory for a good one, but a problem which is built into the most fundamental ways in which the present discursive way of making meaning operates. The implications of discarding a discursive structure without having an alternate, more useful story to tell of our practice is daunting, and I have worked to
present at least the bare outlines of an alternate way of conceptualizing learning and knowledge that evades the problems with the current understandings of representation. Finally, I have made an effort to demonstrate that the newer, situated connectionist story that I tell can imply different and more productive approaches to educational theory and practice.

But this path is not as smooth as one would like; it remains unpaved and in need of further work in many places. There are also tempting side paths and potential extensions which are still unexplored. The proposed stance is different enough to have much broader implications than have been discussed to this point. Indeed, it is unlikely that all the implications can be seen from this vantage point. It is to these absences and possibilities that I now turn.

Absences

In offering an alternate understanding of learning I have outlined a story which stretches from physiology to sociology. This is a broader range of disciplines than have been understood as being relevant to learning. By emphasizing factors on either end of this scale I have implicitly ignored other points in the continuum. This approach was motivated by my determination to build a theory of learning which would be a useful component of a framework for instruction. By dismissing descriptive theorizing as inadequate to the task at hand, I was led to adopt a stand calling for an understanding of the material processes of change that constitute learning. Connectionism is the only realistic competitor in this arena. Adopting a connectionist stance leads very quickly to a search for external sources of regularity to support the unstable matrix that is central to connectionist representation and memory. Exploring the pedagogical implications of connectionism presses the researcher toward attempting to
understand situated knowledge. In a very real sense, the story told here is the simplest story that could be told given the assumptions and intent that guided its development. By adopting a position that favors material process over functional description, however, I have left little justification for ignoring other material conditions which undoubtedly effect what and how people learn.

The Individual

Most notably I have little discussed the individual. In other contexts I have critiqued situated cognition for not giving an adequate role to the person acting across situations, a critique I repeat in chapter 4. \(^{283}\) I do not here, however, present a theory of the social constitution of the self. In part that is because George Herbert Mead and John Dewey, in their respective social psychological and educational contexts, have already presented a position on the self and its basis for autonomous action that fits comfortably with the perspective on learning that I propose. \(^{284}\) Recall the quote from G. H. Mead which headed chapter 5:

The locus of the mind is not in the individual. Mental processes are fragments of the complex conduct of the individual in and on his environment. \(^{285}\)

Pragmatic theories have not lacked explanatory power but have been largely unable to resist the weight of cultural assumptions that militate against their


adoption. Among these cultural assumptions is the logical image of thought as the only justifiable basis for rational action in the world, an image which is basically incompatible with a pragmatic understanding of the basis for action. As I have discussed in relation to pragmatic theories of the individual, strongly bio-social perspectives are at a considerable disadvantage in presenting a convincing case because their view of the self is at odds with the socially sedimented self-understanding of those of us raised in Western traditions of radical individualism. This dissertation is a step toward a conception of learning which is not at war with pragmatic insights.

A situated connectionist position on the self provides support for visions of the self as multiply and contradictorily constituted in social interaction.\textsuperscript{286} It decenters our attention from a reified individual actor and moves it out into the world of social practices. If such social practices are what constitute us as social actors, it is easier to believe that individuals may be composed of different practices within differing social situations and that potentially contradictory practices, understood as intersubjective habits, will be instilled in differing circumstances.

One future task, particularly in the context of teacher education, will be to explore the possibilities of building explicit links between the learning theory proposed here and pragmatic concepts of the self. These linkages would be particularly useful in understanding how education students become teachers. What practices support the transition to “thinking like a teacher?” Can we build practices into our educational programs that encourage the teacher to see, as the phenomenological perspective would have it, “with teacherly eyes?” Such a

\textsuperscript{286} For an excellent rendition of the implications of such a position from an educational point of view, see: Cameron McCarthy, “Rethinking Liberal and Radical Perspectives on Racial Inequality in Schooling: Making the Case for Nonsynchrony,” \textit{Harvard Educational Review} 58 (August, 1988): 265-275.
sense of the self would allow us to relink the issues of habit, perception, and
cognition within the still-active pragmatic tradition of reflective inquiry which
supports a style of teacher education in which an ethical imperative for teachers
to choose to be open-minded, responsible, and whole-hearted is emphasized.287

A plural understanding of the constitution of the self and the role of social
activity in that constitution also lends a distinctly different understanding to
issues of multicultural education. Particularly in cases where school practices
are understood in opposition to practices developed in family, ethnic, or gender
groupings, a sensitivity to the ways in which successful students who are also
members of such groups are contradictorily placed can lead to a different set of
solutions for such students. Typically, following an individualist ideology,
schools encourage “positive attitudes” linked with “self-esteem.” Such
programs implicitly attempt to reconstruct the individual on the basis of values
approved by schools and serve to reinforce the opposition between schooling
and the practices of alternative groups. Schools seldom succeed in convincing
students to abandon the local groups which are the foundations of their identity
unless such students are already marginal in those groups. A more successful
attempt to serve the legitimate educational goals of schooling might be to
devise practices promoting scholastic excellence that can be interpreted from
within the perspective of one participating in oppositional groups as conferring
prestige. This would be a matter of devising a matrix of practices, a school
climate if you will, which encourages a multiple, contradictory self which
includes a successful scholastic self rather than insisting on a unitary identity
constructed on the basis of school values. Schooling and the scholastic attitude,

287 This is, of course, a reference to Dewey’s formulation. For a piece which emphasizes the
ethical dimension of the pragmatic stance, see: Carl A. Grant and Kenneth M. Zeicher, “The
Teacher,” in Preparing for Reflective Teaching, ed. Carl A. Grant (Boston: Allyn and Bacon,
1984), 1-19.
would be offered as one of many "ways of being" in the world—not as the master form.

_Institutions_

In making suggestions about the construction of the institution of particular schools on the basis of their interaction with the social groups which organize personal identity, I enter another set of issues which have been largely absent in this work. The social aspects examined here have remained focused on the shared, habitual practices which comprise individuals and which define social groups. I have not talked about those social groups. In sociological usage, or more precisely in a sociological usage which descends from G. H. Mead’s seminal work, institutions such as schools, classrooms, affinity groups, ethnicities, and families are central to understanding human sociality.

Attempting to design the practices which help shape such groups is particularly touchy since these institutions are precisely the sorts of organizations that have shaped those who are attempting to reconstruct them. Ethical questions and questions of power, a final, almost terrifying power to even shape what is understood as power, would pass into the hands of the designers of alternate practices. This poses an immediate ethical dilemma for my own understanding that the work produced here is most directly useful to teachers in the design of instructional practice through which to teach content and in the design of practices which organize the classroom as a learning community. It is precisely at the point where a "better" theory of instruction will allow the construction of effective technologies of control to replace the patently ineffective ones currently in place that the ethical issue of power arises for any teacher. This concern invests with an additional emphasis the
traditional pragmatic charge that teachers act on the behalf of their students and in the pursuit of democratic ideals.

The Body

While the work of this dissertation does lead to an embodied conception of human cognition, it focuses chiefly on practices and connectionist learning. Other elements of our bodily constitution are also clearly important.

Connectionist perspectives, conceptually built on the interaction of many simple parts, systematically ignore factors which directly affect its functioning at a global level. The best example of this is the lack of remark on advances in understanding the chemical component of human thought. The brain, in addition to being composed of many densely interconnected neurons, is also bathed in a continually changing flux of internally generated chemicals which affect mental function on a global level and have been implicated in mood.\(^{288}\)

Consideration of this absence makes it clear that connectionist theorizing is not to be identified with brain function as such but that it is more usefully understood as a neurologically plausible model of particular brain functions. It cannot be read directly down onto actual individuals in any simple way.

The work of Mark Johnson considers the body as a whole and the ways in which our particular bodily configuration constrains the way that we can understand the world and the metaphors through which we think. This, too, is well worth the study of teachers, for such work helps us understand the limits that may apply to our attempts to construct new practices based on differing metaphors. Such metaphors will be most readily accepted and used if a clear experiential analog can be found.

Implications

Some implications of the perspective offered in this dissertation have been explored in chapters 5 and 6. There I work on a few of the implications of a situated connectionist stance for educational research and the design of educational materials. In general, the task before the teacher or researcher is to investigate the ways that situated knowledge is appropriated by students who are using the practices that constitute such knowledge. The situated connectionist approach advocated here assumes that this appropriation occurs via the processes of connectionist learning. This approach does not, however, assume that what is appropriated is very much like what the classical tradition gives us as the objects of classical knowledge; not only are such representations unstable, slippery, and context-dependent but they are also not simply located in the brain and possessed there by the individual. In this view creating new knowledge is not the province of obscure researchers or writers of Ph.D. dissertations but is the journeyman work of teachers who create and sustain appropriable practices. In an interesting aside which is suggestive for educators, Rumelhart et al. note that it appears to be extremely difficult to invent novel and useful new external representations, saying, "It may be that the process of inventing such representations is the highest human intellectual ability." Heady stuff for a teacher who takes this task as a central element in his or her practice.

---

289 This is not always apparent in the writing of all those who work in the Vygotskian tradition who call on the phrase "appropriation." Like Vygotsky himself (see chapter 4), they often seem to believe that a logical object is formed in memory.

290 David E. Rumelhart, J. L. Smolensky, James L. McClelland, and Geoffrey E. Hinton, "Schemata and Sequential Thought Processes in PDP Models," in Parallel Distributed Processing,
Recovery and Reconstitution

Taking a view of the teacher as a creator of knowledge through the creation and sustenance of practices, and as an active agent in the co-construction of learning via his or her participation in those practices while teaching, leads to a different way to regard the history of teachers' work. It seems likely that many of the practices that teachers use casually are more valuable resources than are commonly realized. These, and not simply the content that they purportedly carry, are the "stuff" of curriculum. In an increasingly technocratic era many of these practices have been "rationalized" and largely lost to the everyday culture of teaching. Practices such as heterogeneous cooperative grouping and peer teaching are being reintroduced as the latest methodology after being earlier suppressed as inefficient and as relying on non-experts to teach. One is led to suspect that other practices could be recovered through historical research and that these would be a potential buried treasure in an archeology of education.

Other practices which remain a troubling feature of our school experience could be reconstituted under the perspective presented in this dissertation. For example, it is popularly understood that memorization is simply a poor form of learning. Yet memorization, in part because "possession" of facts remains the discursive sine qua non of learning, remains common. Teachers often experience a conflict between the need to teach basic factual knowledge in a new area before a contextually meaningful framework for learning is established and their distaste for memory work. Interestingly, especially in the elementary grades, such work is often accomplished through rhymes, songs, and nonsense patterning. These are precisely the common practices that classroom teachers

avoid when “showing off” their students to visiting student teachers. If we understand learning as connectionist pattern completion mediated by social practices, the common habit of employing patterning exercises to move quickly through largely meaningless material can be reconstituted as an appropriate learning practice in particular situations. The plaint of the social studies methods teacher that learning the capitals of the states is not social studies is often met with the reply that the state-mandated test for this grade level implies a different understanding of the nature of the social studies. An appropriate response might be to abandon the largely futile attempt to make salient the fact that Lincoln is the capital of Nebraska in a reasonable length of time and instead to spend three or four 15-minute sessions singing a nonsense rhyme about the states over a few weeks’ time. This could free time in the social studies portion of the day for a more meaningful approach to cultural differences or the decision-making powers of the students.

Intentionality

Indeed, meaningfulness is a difficult problem not only for the teacher but for the broader discourse in which teachers are embedded as well. Philosophers have discussed this under the rubric of intentionality. This goes back to the issues explored in chapter 2 as problems in representation. Somewhat simply put, for both teachers and philosophers, a problem in representation exists when it is not clear how a correspondence between an object in the world and the sign that people use to designate that object is warranted. The whole problem has been attacked by no less than John Dewey as a false problem based on an unwholesome Descartian dualism.291 The problem, for Dewey and for us, 

291 This was a major theme of Dewey’s. See: John Dewey, The Quest for Certainty: A Study of the Relation of Knowledge and Action (New York: Minton, Balch & Company, 1929).
is to formulate a coherent alternative monism that can substitute for the
discursively natural division between the sign and the thing signified. The
pragmatists, of course, implicitly relied on the insights generated by C. S.
Peirce’s triadic version of sign constitution which includes the object signified
as a part of the sign. But semiosis has remained marginal, perhaps in part
because it seemed a speculative formalism without an obvious vital connection
to the world. A situated connectionist approach has the virtue of providing an
explicitly material monism. As has already been dwelt on, meaning in this
approach is established through practices. Understanding these practices as
patterns to be completed by the learner and understanding representation to be
directly tied to the experience of that pattern, undoes the assumption that the
relationship between the sign and the thing signified is arbitrary. At least in the
experience of the student, the sign is directly linked to the pattern it represents.

More profoundly, the world is its own representation—connectionist
memory is directly tied to the continued experience of the world for its
sustenance; the link to the world is not merely originary but maintaining. The
categories of experience emerge from our active engagement with problems. A
dualistic perspective, however persistent it may be in our day-to-day language,
is avoided here, and the dualistic problem of justifying the connection between
symbols and objects is evaded.

Because in the alternate monist account all learning takes place through this
interplay of practices, it is not, strictly speaking, possible to understand any
form of learning as meaningless. But it does introduce differences not found
within the traditional account. Learning that fails to transfer to its appropriate
context of use is the sort of learning that we often label meaningless. It is either

292 Peirce’s style is notoriously difficult; for a readable rendition of triadic semiosis, see: John
forgotten or not recalled at the appropriate time. In the traditional story this inability is about a failure in the storage and retrieval mechanism. In the monist story built here the “unrecalled” knowledge just does not exist, and the problem lies not with storage and retrieval but with cognition and re-cognition. The practices which constitute our individual cognitive extension into the world are not present, are not recognized, or are not re-created. The appropriate teacherly response would be to work with new practices more closely tied to the context of use.

Take, for example, vocabulary building. One common practice is a spelling test. But simply because a child can spell “veto” does not mean that he can understand very much about the way that President Bush ran the country when faced with the short phrase “Bush governed by veto—and look where it got us.” muttered in daily discourse. A situated connectionist analysis would support the practices of whole language in which new vocabulary in encountered is the context of use. But it would go beyond the simpler versions of whole language and also endorse the use of sources in which the sense of a word is “stretched” to its reasonable limits. In addition to a consistent use of the term “veto” in its central political meaning, it would also be useful to extend its sense to situations where a father might “veto” his son’s plan or where a law of nature might “veto” the construction of a time machine.

Development

One area deserving further consideration from the point of view of the analytic being cultivated here293 is development. The integrated position suggested here can help to make sense of the often contradictory and confusing

data of child development. McClelland, one of the original workers in the field of connectionism, has advanced a model which shows how slow, incremental learning in neural nets can lead to the apparent plateaus and sudden dramatic changes in performance which teachers observe and which have lent a sense of truth to stage theories. McClelland’s position, however, is compatible with positions that locate the change locally—it is not the global change in ability which Piaget’s model postulated.

Stage theories of children’s abilities have remained a feature of educational discourse, despite convincing research which questions their validity, largely because of their heuristic value. They still seem to help guide a teacher’s practice. Teachers’ experience-based understanding that while children’s exact developmental path cannot be predicted, the broad outline of their changes can be safely assumed, is contradicted by the present discourse which understands such reasoning as unsubstantiated and teleological, as illegitimately attributing cause to the result. The educational language that sees a child as developing “toward” certain competencies such as language use is outlawed. But perhaps the persistence of such language reflects a failure of the current discourse to account for the regularities that teachers actually encounter. Just as the analysis of cognition suggested by the traditional, logicist discourse (discussed as the analytico-referential in chapter 2) has proven inadequate to the task of understanding cognition, so too, perhaps, will it take another approach to development to explain that phenomena to teachers in a useful way. One way to approach this would be to look to the larger field of complexity theory, which is developing in response to perceived isomorphisms in new and

productive approaches in many fields such as genetics, ecology, immunology, and not incidentally, connectionism.295

A Pragmatic Contextualization

Chiefly, I hope, moving to a new vision of the process of learning can help to sustain a move away from the technocratic mode of education and toward one based on embodied practices. In large part this is continuous with a long tradition emphasizing practical competence.296 Most recently and accessibly this tradition has been located in pragmatist approaches to education.

Interestingly, both John Dewey and G. H. Mead espoused an approach to education which affirmed the role of both the biological and the social. Mead, in particular, used the term "bio-social" as an adjective to modify his position. A series of remarks drawn from the pragmatists may serve to emphasize the compatibility of their approach with the approach taken here:

... based on what must happen in the brain. Every sensation corresponds to some cerebral action. For an identical sensation to recur it would have to occur the second time in an unmodified brain. But this strictly speaking, is a physiological impossibility. ... Every thought we have of a given fact is, strictly speaking, unique, and only bears a resemblance of kind with our other thoughts of the same fact. When the identical thought recurs, we must think of it in a fresh manner, see it under a somewhat different angle, apprehend it in different relations from those in which it last appeared. And the thought by which we


cognize it is the thought of it-in-those-relations, a thought suffused with the consciousness of all that dim context. 297

The proposition [that no two "ideas" are ever exactly the same] is more important theoretically than it at first sight appears. For it makes it impossible . . . to formulate the mental facts in an atomistic sort of way, and to treat the higher states of consciousness as if they were all built out of unchanging simple ideas . . . 298

The initial stage of that developing experience called thinking is experience. This remark may sound like a silly truism. It ought to be one; but unfortunately it is not. 299

So far as the significant symbols which the individual uses are stimuli to his own responses, these processes lie in the individual. So far as things, characters, and imagery are indicated, the processes extend beyond the individual. The locus of the mind is not in the individual. Mental processes are fragments of the complex conduct of the individual in and on his environment. 300

The theory of knowing which is advanced . . . may be termed pragmatic . . . It holds that knowledge consists of . . . all the habits that render our action intelligent. 301

Accordingly, just as we say that a body is in motion, and not that motion is in a body we ought to say that we are in thought and not that thoughts are in us. 302

In many ways the pragmatists advanced a framework which seems surprisingly current when viewed against the background of connectionist learning and situated knowledge. The series of remarks above describes a basic


298 William James, Psychology: The Briefer Course, 2.


300 Mead, The Philosophy of the Act, 372.

301 Dewey, Democracy and Education, 344.

arc of this dissertation. Beginning with a frank confrontation with the full import of a nondualist, materialist conception of mental processes, implies that there can be no simple reductionist conception of "ideas." Ideas of this sort are grounded not in logical simples but in experience. But this is not the experience of the sovereign knower whose gaze possesses a scene, but of a person-acting, a person engaged in the world whose mental processes are both about and of the world in a literal way. Mind and thinking extend beyond the brain. Habits, what I have called practices from a pointedly social perspective, are the basis of a pragmatic conception of knowledge and therein lies the origin of intelligent, competent action. We are profoundly enmeshed in a world of thought, a universe of signifying practices; we are literally "in" thought. The world thinks us no less than we think the world.

The valuable tradition that the pragmatists leave us can be usefully extended. While the pragmatists, being determined monists, affirmed a role for both the biological and the social, they did not have the resources to apply to their broader project that we have today. In particular we have developed useful tools in the complex, emergent properties of connectionist learning and in the practices that support situated knowledge. With these we can build an account that constrains our theory and practice, our praxis, more effectively than was possible for the pragmatists.

Connectionist learning, at the biological end of the scale, makes it possible to explain how experience results in associative learning, and it helps inform us how the repetitive, socially-organized, encounters with the full range of category members enable the act of categorization. Understanding the processes that support learning allows us to understand the qualities of the learning experience that are salient to the act of learning. For example, given a particular socially framed category, we can say more about what sort of
examples (central or noncentral) should be presented to form a category with boundaries similar to the socially constructed original. We can also see how the sequence in which they are presented can affect how rapidly the category is achieved and think productively about trade-offs between rapidity of learning and achieving a finely tuned boundary distinction. Certain practices might allow a more rapid entry into the field of activity in which the category is arrayed while others would tend to delay entry but would result in a more accurate category and hence a fuller participation once entry is achieved. Choosing between such practices is part of the craft of teaching under the conception of learning presented here.

The modern conception of practices extends the pragmatist conception of "habits" by stressing the intersubjective nature of their construction and by detailing the potential that such intersubjectively established activities have for becoming sedimented in particular material objects. It is possible to explain in much more detail the practices that compose a cultural object such as a map and its competent use. Being able to see such differences allows the teacher to recognize where things have gone awry for a particular student and to focus that student's attention on those points.

Practices and connectionist learning are much more productive considered in relation to each other than they are when considered alone. Together they show us how the puzzling qualities of knowledge (context dependence, shifting boundaries, etc.) are established and how practices are learned. They fill in the perennial question of how "internalization" occurs while questioning the division that makes internalization a sensible concept. Practices help us understand what is learned and which patterns, out of the plethora of potential patterns the world offers, are signified as carrying knowledge.
Taken together, connectionist learning and situated knowledge form a powerful way to think about learning, a perspective I have labeled situated connectionism.

The Child

In chapter 1, I introduced a child who recognized her mother's face and suggested that a theory of learning adequate to the purposes of education would be able to explain how such a wonder could occur. This first instance of a child's learning is manifestly meaningful learning. Almost every child learns to recognize his or her mother's face easily and fluidly. A theory of learning which leads us in the direction of making such meaningful, readily achieved learning available in the school contexts is what I seek in this dissertation.

I seek a story which can usefully explain this instance of learning:

I sit in a room and watch a young child in a crib beneath an open window. She is beautiful, or I think she is. I sit immobilized, fascinated. A cool gust of wind moves the curtains, annoying the child. Her mother, harried, returns and bends over the crib. The baby perceives motion, turns her head and recognizes her mother. She gurgles, smiles, reaches out and is softly gathered in....

Can we at least begin to tell the how of that story? I believe we can.

The first move in explaining learning lies in clearing away the hegemony of logic in explaining human thought. The human mind does not, and if we play the game by strictly logical standards, cannot work through discrete objects ordered by the seriality of deductive logic. This conclusion depends on accepting the materiality of thought and locating thought in the human brain and material social practices. In doing so we reject the Descartian idealism which buttresses the present discursive structure. If we settle on this material pattern of explanation it is a short path to the conclusion, based on the architecture of the brain and its observed speed of operation, that human
thought is composed of parallel processes and interactive representations that cannot be mapped onto simple, deductive logic in a credible way.

This denaturalization of logic as the basis of human thought is easier to achieve if we recall that the idea of logic as the basis of the best human thought arose in response to a particular seventeenth-century crisis of representation in Western culture. In solving that problem the West postulated that deductive logic, the grammar of language, and causality in the world all shared the same perfect, timeless form. That postulate is no longer intellectually tenable but our culture has yet to adjust its fundamental discursive structure to accommodate the change. Cognitive science and particularly the research cited herein marshals a large body of evidence against Descartes' extension of that discursive structure which claims that thought is itself logical.

In an alternate story the child is not a "natural" logic machine as Descartes presumed, nor does she "develop" logic by reading it off the world as Piaget believed. Instead the child operates by different principles than those which organize geometry. She operates, if you will, on the contrasting principles of poetry. Meaning-making is a context-dependent, holistic, relational operation which operates in accord with the individual history of the interpreter—such as we habitually consider it to be in the restricted realm of poetry. At the same time much of the context for that meaning-making is established socially, and a much larger amount of knowledge than a Western individualist ideology would have us believe is based in practices that operate and change in the world and which are only later, if ever, internalized through connectionist processes of appropriation.

In the story of the child we start with a situation:

I sit in a room and watch a young child in a crib beneath an open window . . . . A cool gust of wind moves the curtains, annoying the child. Her mother, harried, returns and bends over the crib.
This is an old image, one we all recognize in some way. "Mother and Child"—it might be a painting by a latter-day Rembrandt, chronicling the domestic pleasures of our day. But it is important to recognize that it is also an old situation for this mother and this child. This has all happened before in their lives and their activities are well coordinated. The mother, as harried as she is, does not need the child’s cry or even her arms reaching out to know that she is wanted. It is cool, the wind has picked up—she feels her daughter's need as surely as she feels the wind. So she returns to the room to tend her daughter. The daughter, for her part, may or may not yet know her discomfort, but by the time she sees her mother bending over in the context of a cool breeze and the fluttering drapes she does know her role and is eager to play it.

The baby perceives motion, turns her head and recognizes her mother. She gurgles, smiles, reaches out and is softly gathered in . . . .

What we need to be able to say is how she recognizes her mother and how that recognition is tied to the competent response of reaching out and the reciprocal response of being softly gathered in.

She recognizes her mother through the emergent primitives associated with connectionist architectures. There is a literal quality to the word "re-cognizes" used in a connectionist explanation. The child re-creates her mother; she does not, cannot, retrieve a memory which, in the conventional sense, does not exist. The raw material of this recreation is the impression of her history upon the malleable organization of her brain. That history contains multiple instances in which there has been a cool breeze, an open window, a crib, a fluttering curtain . . . and a mother bending over. The overall pattern is quite reliable and not every element must be repeated in just the same way for the competent response of reaching out to become active. The daughter completes the pattern; she does not analyze the scene and then deduce an appropriate response. There
is no separation, except in our analysis, between the perception of her mother and the act of reaching out. It is all poetry, all perception, all action.

But the act of reaching out is not to be understood as a "natural" human response any more than we believe that the image of Mother is somehow naturally lodged in each child's mind. Both are learned in the same fashion, by repeated presentation of the pattern in a context where the child's emergent perception of category is tied to a socially appropriable response.

Some slowing down and unpacking is called for here. What I am proposing is a particular hermeneutic of learning based on connectionist principles and social practices. In this story the mother is the socially competent actor leading the child into competent activity by presenting experiences from which the child can learn. Like many teachers she acts chiefly out of affection and is largely unconscious of the full effects of her own activity. The mother knows, from her own experience and the lore of motherhood, that her child is uncomfortable in a sudden, cool breeze. At the same time she values fresh air and the image of her child's crib beneath the open window. She knows that reaching out and holding close is warming and comforting—a lesson first learned in her own mother's arms. Holding your arms out as the first portion of the act of holding is a well-established social practice in the world in which she lives, a practice that is habitually interpreted to call for a reciprocal response from the person being summoned by the gesture. And, reciprocally, all understand that the summons may flow in both directions; the child may summon her mother by the same gesture.

Enmeshed in this world, the mother enacts her role consistently. She goes to her child when she anticipates the child's discomfort and holds out her arms. She may hesitate at that moment, expecting this hesitation will somehow, as it does so often in her life, trigger an appropriate, reciprocal response of reaching.
The child, for her part, explores the use of her own body. She thrashes, and waves, and grasps at things. Sometimes the child holds both her arms out from her body. In this environment an ecology of behavior takes shape over the history of their interaction. When the mother infers that a child is uncomfortable she holds out her arms and picks her daughter up. She is often right about this and her daughter learns to associate the eminent cessation of discomfort with this part of her mother's repertoire of activity. Arms held out from the child's body are taken to mean a desire to be picked up, especially when the child is under conditions where she may be uncomfortable. In this dance of action and response it may be difficult for the outsider to locate the initiator of an activity; indeed it may be difficult to separate these as an actor. But as the pattern recurs and strings itself out over time, the action of both cohere into well-coordinated activity. The mother reaches out and the child is held close. Or, finally, the child reaches out and is softly gathered in. They have learned to hold each other and have learned what such an attachment means.

To tell this story in the current discursive context without referring to a set of concrete material processes is to play into a hegemonic structure which, pleasantly and with heart-felt agreement that there is something profoundly correct in the story, easily consigns the recounting to the personal and to the irrational—to the particular. It is only "a" story. I want to claim that it is more than that. It is an analogy for all our stories about learning. It applies to the biologist as well as to the baby, and it applies to biology students be they in grade four or fourteen.

Learning is grounded in the central reciprocal relation between the malleable material organization of our body and the patterns which recur in the world. Humans are pattern completing animals. We seek pattern and insist on finding it; we have little choice—the basic architecture of our brain prepares us
to categorize in particular ways. Many, on some accounts all, of these patterns are socially constructed in the activity between individuals.

To move toward reconceptualizing learning and knowledge in this way is to move away from general, eternal forms and toward particular, social patterns. It is to move away from logic and toward poetry. We move from what Freire called a banking model of education to an image of education as the co-construction of appropriable patterns.

This dissertation attempts to establish a basis within the current discursive structure for challenging our self-understanding as logical beings. By using the hegemonic tools of the discourse—proof, disproof, empirical evidence, and deductive logic itself—I attempt to establish a space where another image of our capacities may flourish. In using these tools and by arraying the privileged signifiers of science, logic, statistics, and empirical evidence, I hope to establish a claim that cannot be dismissed as simply romantic, particular, teleological, or "affective:" the traditional, hegemonic ways to dismiss the cognitive understandings that I have tried to establish. Admitting that such claims have a truth of their own but ghettoizing them in different, pointedly noncognitive, realms is a potent way to circumscribe their influence. In the story I wish to tell a mother's love in not "ineffable," "natural," or "intuitive;" it is effective action in the world continuous with and learned in the same fashion as the practices that establish a biologist's expertise in cellular identification.

Taking this tack allows us to see schools differently. We are used to thinking of the sedimented practices of schooling as a problem in teachers lives and for student learning. But here we have an opportunity to see practices as effective constituents of both identity and learning. And we are used to thinking of learning as a problem, separate from and more difficult to achieve than other constituents of effective activity. Here we have an opportunity to see learning
as a social activity continuous with action and perception. In this context we can question the current assumptions of schooling and the practices which are produced by and which produce those assumptions.

This approach takes us to a place which may appear strange, a place where individuals are not discrete and where the change we call learning takes place as often between people as within. We move into a space constructed of networks of relations rather than causal chains and where constraints are more interesting than causes. In some ways, though, this is not a completely unfamiliar space for the educator. Teachers practice in a messy, contingent world where causal chains are seldom informative for practice. Teachers tend to see the classroom as a particular, unique setting, each one of which places its own constraints on the activity of the teacher. It is perhaps appropriate to move into a way of viewing the teachers’ workspace through an analytic lens which does not do violence to their perception. Moving to the view of learning and knowledge advocated here can have large consequences in this regard.

Conclusion

In abandoning formal, axiomatic logic as the foundation of its perspective and adopting a perspective which depends on emergent structure in the face of experience to account for knowledge, a situated connectionist approach asks a fundamental question about the very nature of being human. Should its insights prove compelling its success will force an intellectual reconsideration of tremendous proportions on our society. Because the first and most important areas of success for this viewpoint appear to be in models of perception and learning, education will be one of the first areas to come under question. New primitive functions will replace the older store, search, and match operators
upon which so much of our educational practice is founded. (Drill and practice is the most direct analog of this preunderstanding.) Operators such as associative recall, content-addressable memory, learning via exposure to example, context specificity, and the power curve of learning will all become first order primitives. Such a perspective implies that attempting to teacher-proof a classroom by mandating the teaching of only “objective” facts in a lockstep order is doomed not only to failure but also to actually damage a student’s ability to learn by separating the “fact” from its context, a context which is absolutely essential to its recall and productive use at even the lowest “level.” This perspective implicitly argues that facts cannot be discrete and immutable because the distributed representations interact, the pattern in which they are embedded affects their recreation and use in actual practice, and they change, to some extent, each time an associated memory is stored. The implications for teaching of such a conception are profound: teaching the right answer would be recognized as futile enterprise, whereas symbolic logic implies that there is always a universal best response. Much of what is implicitly endorsed here is already a part of the repertoire of good teachers but it exists as an isolated result of their own experience, and one which is difficult to justify in a fully articulated way. This perspective would provide a consistent

---

303 As relaxation models are more extensively explored, even more fascinating primitive functions are coming to the fore. Rumelhart and McClelland (p. 126), for instance, discuss primitives for which we do not have simple names: “relax into a state that represents an optimal global interpretation of the current input” and “retrieve the representation in memory best matching the current input, blending it into plausible reconstructions of details missing from the original memory trace.” See: David E. Rumelhart and James L. McClelland, “PDP Models and General Issues in Cognitive Science,” in Parallel Distributed Processing, Volume 1: Foundations, ed. David E. Rumelhart, James L. McClelland, and the PDP Research Group (MIT Press: Cambridge, MA, 1986), 110-149.
understanding of such practice which would reach as far down to the roots of how we believe people think as has the rationalistic explanation traditionally.

To my mind the opportunities for shaping our common vision of education loom large. The potential for reshaping an interpretation of how (and thereby what) it means to be human presents itself. Such an opportunity is not likely to be offered twice in anyone’s lifetime. As educators we are obligated to examine our own experience and understanding of what education consists of and to honor that understanding. As one of the primary “consumers” of academic psychology we are an important part of the universe of discourse which will decide which perspective on knowing will prevail. We will vote with what we teach and, more importantly, with what we practice and how we justify that practice.

It is possible to understand how the child knows her mother. In coming to that understanding we trade a presumed stability for a dynamic and unstable conception of ourselves. Any deep change carries risks and educators bear a special responsibility to our society for educating the young. It is my hope that a new common sense understanding of what it means to think may arise in a space which this dissertation explores, and that the newer understanding will prove more valuable and less misleading in the practice of education than the understanding which it supplants.
WORKS CITED


Boole, George. The Laws of Thought. Chicago: Open Court, 1854/1940.


APPENDIX

Category Learning Program

Instructions and System Requirements

This program is implemented as a HyperCard Stack. It illustrates a situated connectionist approach to learning bifurcated categories.

The program is on a Macintosh diskette in a pocket on the rear cover.

Upon launch you will be presented with an opening screen which briefly explains the purposes of the program. It is designed to track multiple users over time and requests your name in order to begin a record of your interaction with the program. After you enter your name you may click the button on the bottom of the screen and it will open to the first learning screen. Enjoy!

You will need:

An Apple Macintosh computer
A 13 inch monitor or larger
The ability to display 8 or more levels of gray or color
System 6.05 or higher
HyperCard 2.0 or higher
Approximately 800K of available RAM
VITA

John St. Julien was born and raised in Baton Rouge, Louisiana, son of a Cajun father and Kentucky hill country mother. He was a good student and an aggravating adolescent—being considered an activist as a result of participation in the local civil rights struggle. He graduated from Robert E. Lee High School in 1970, from Louisiana State University in sociology in 1973 and, following a long-established tradition of that field, sought employment as a carpenter. He married Layne Darby and after helping to raise two children he returned to graduate school in education, gaining state certification and, ultimately, pursuing a masters and, in 1994, a doctorate in that field.

His intellectual interests grow out of interlocking commitments to place, children, and learning. During his initial teacher training and student teaching experience he became convinced that there was a curious void at the heart of the educational enterprise: no one seemed to be able to say just what learning was. He was drawn into an expanding study of conceptions of learning involving cognitive science, computational theory, practice theories, artificial intelligence, neurology, intellectual and discursive history, and educational practice, research, and theory. The present work is the fruit of that search.

John has participated in designing and implementing the present Holmes teacher education program at Louisiana State University and is particularly interested in pursuing cognition and design issues in a teacher education program focused on praxis.
DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: John A. St. Julien

Major Field: Education

Title of Dissertation: Cognition and Learning: The Implications of a Situated Connectionist Perspective for Theory and Practice in Education

Approved:

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

Date of Examination:

April 4, 1994