Massachusetts Institute of Technology
Artificial Intelligence Laboratory

Working Paper 315

September 1988

Writing and Representation

Philip E. Agre

Abstract

This paper collects several notes I've written over the last year in an attempt to work through my dissatisfactions with the ideas about representation I was taught in school. Among these ideas are the notion of a 'world model'; the notion of representations having 'content' independent of the identity, location, attitudes, or activities of any agent; and the notion that a representation is the sort of thing you might implement with datastructures and pointers. Here I begin developing an alternative view of representation whose prototype is a set of instructions written in English on a sheet of paper you're holding in your hand while pursuing some ordinarily complicated concrete project in the everyday world. Figuring out what the markings on this paper are talking about is a fresh problem in every next setting, and solving this problem takes work. Several detailed stories about representation use in everyday activities—such as assembling a sofa from a kit, being taught to fold origami cranes, following stories across pages of a newspaper, filling a photocopier with toner, and keeping count when running laps—illustrate this view. Finally, I address the seeming tension between necessity of interpreting one's representations in every next setting and the idea that everyday life is fundamentally routine.

Contents

Acknowledgements

Introduction

This paper collects some short essays I've written informally over the last year as I've tried to work through my dissatisfaction with the theories of representation I learned in school. Although the ideas and exposition are only moderately polished and despite their lack of worked-out technical proposals, I am circulating them in this form since they seem to me to articulate some concerns and intuitions that have become widespread in the field over the last couple of years. In reading them it might help to have read the papers of mine and Chapman's listed in the bibliography.

The notion of representation obviously labors under a long philosophical history. These days, though, it also labors under an appreciable technical history, one handed down in the form of an increasingly routinized technical practice of building computer systems that in various ways construct, maintain, and manipulate 'representations.' What I find most interesting about the subject of representation, indeed about artificial intelligence research as a whole, are the interactions between the philosophical and computational issues. I have regularly found that philosophical analyses suggest useful interpretations of difficulties that arise in my technical practice. And I want to believe that our technical practice can help with philosophical inquiries as well. Although I've tried to write in plain English, my principal motivation has been to explore some places where technical questions appear to align with philosophical answers. I don't yet understand how to convert these answers back into technical practice, but I want to.

These notes make both negative and positive suggestions. On the negative side, I am against two related and widespread (though certainly not universal) ideas about representation, namely 'semantics' and 'world models.' These words get used in various ways, so here are some rough definitions.

The notion of semantics holds that a representation has a 'meaning' or a 'content' independent of the identity, location, attitudes, or activities of any particular agent. (For an ambitious exception see Barwise and Perry.) At the risk of collapsing some distinctions, this meaning or content is often understood in terms of a systematic, objective relationship between the representation itself and states of affairs obtaining in the world.

A world model is a component of some physically realized computational system, an object whose internal structure stands in a systematic, objective, analogical relationship to states of affairs presumed to obtain in the world. Some process maintains the model as the world changes; reasoning about the world is a matter of inspecting and manipulating the model. The building and maintenance of world models regularly forces implementers into a seemingly unresolvable trade-off between overly restrictive assumptions and intractable computational complexity.

On the positive side, I would like to make two suggestions about how people use symbolic representations. The first is that *people use symbolic representations to help them make sense of particular situations.* Given a symbolic a representation there isn't much you can do to 'understand' it until you're faced with a particular situation in which it might be relevant.

(Whatever the form of the representation, whether written or spoken or displayed on a video monitor, I shall speak of it, or more precisely its configuration of symbols, as a 'text.' I'll use the word 'representation' when I mean to emphasize the actual material object: the sheet of paper, the speech signal, or the video image. I'm afraid that this use of the word 'text', although perhaps not standard, has become natural to me through exposure to contemporary French philosophy. I hope I haven't been too corrupted by French perversity to express myself clearly to a technical audience.)

My second suggestion is that *what a given text is talking about is a fresh problem in every next setting.* The 'work' of relating a text to a concrete setting—looking around, poking into things, trying out alternative interpretations, watching someone else, getting help—can be either 'mental' or 'physical,' though it is best not to distinguish. Relating a text to a concrete setting takes work because the text might be relevant to the situation in a great variety of ways. The text has a certain amount of, so to speak, 'play.' The more play a text has, the less one can conclude in advance about how it might be relevant to actual situations. The intuition is that texts have a great deal of play, so much that one must simply postpone the majority of one's interpretative effort until the time comes. This is just the opposite of extracting a 'meaning' from a text as soon as it arrives. The only way to explain the point is through examples.

(1) You're on a city street trying to get to a party. You've got a sheet of paper in your hand with directions written on it. The directions say "bear left." Now you've got work to do. What does "bear left" mean with me standing just here, at this intersection, now? It may not be obvious. (If you don't believe me, go to Boston, get on Commonwealth Avenue in Back Bay, and head toward Kenmore Square.) No law of nature connects the phrase on your sheet of paper to the aspects of your physical surroundings that should motivate the details of your actions. The person who wrote the instructions, lacking detailed knowledge of precisely where you'd be standing or traffic conditions or your having the sun in your eyes, had to count on you to fill in for their underspecification of your next move. If all streets were laid out in standardized, discrete patterns then one could rely on simple, deterministic rules to relate instructions to the streets one comes across. In reality, though, street layouts defeat any attempts at simple models. Boston's streets are famous for their inexpressible tangles, but some pretty odd things happen within the supposedly simple geometry of midtown Manhattan as well.

(2) I'm trying to find a friend's house from a set of directions. I received the directions by computer mail and had printed them out so I could carry them with me in the car. I knew the main road well enough but I knew nothing about the residential streets leaving

it. The last two paragraphs read:

> About a mile up from the intersection, look on the left for Elk Tree Road—it's a dirt road with a little bus stop at the end. Follow Elk Tree *past* the first left (Big Tree WAY) to the mailboxes and take the middle of the three-way fork on the right, Upper Big Tree ROAD.
>
> My place is the first one on the left, #27. Park on the left shoulder near my white Honda, and come down the steps and up the stairs to my front deck. (If you went to the main door, you'd get my landlords, not me.)

I had the sense to check my odometer at the intersection so I'd know when "about a mile up" was coming. Even so I somehow missed Elk Tree Road the first time. The bus stop is obvious enough if you're looking for it, but it's a little way up the dirt road and obscured by foliage. The most difficult part, though, concerned the "three-way fork on the right." When I got to the mailboxes (dozens of them, in fact) I had to find some way to interpret the scenery as "the three-way fork on the right." Unfortunately, I could only count two roads on the right. After much backing up and looking around, I decided that a large driveway roughly straight ahead was the third road. Setting off along the "middle" road, I looked for the first place on the left. The road snaked down a hillside with many houses on the right. Finally I came to a house on the left. I couldn't find any street numbers (no surprise out here) but there was definitely a white Honda parked outside. I parked and got out and looked for the place that "come down the steps and up the stairs to my front deck" was talking about. I found such a place but it led me to the side door of a cluttered garage. A great deal of searching and asking around followed. Even though something was clearly wrong, that white Honda made me reticent to abandon any of my previous interpretations. To make a long story short, the middle branch of the three-way fork immediately branched into Lower Big Tree Road steeply down to the left and Upper Big Tree Road steeply up on the right. A large hand-painted sign for Upper Big Tree Road clearly marks the split, but I was already looking for the first place on the left. Writing this now, I realize that it makes perfect sense for *Lower* Big Tree Road to branch steeply down and *Upper* Big Tree Road to branch steeply up, but at the time, I didn't give the right branch any thought at all since I simply did not see the branch *as* a branch. If you're there looking at the branch *as* a branch it's hard to imagine how it looked to me, but it never occurred to me that I had another choice of roads to make. The sign must have occupied a reasonable portion of my visual field, but I was already looking for a house, not a sign, and to the left, not the right. (It takes great skill to make a sign that people will see despite themselves. I'd like to understand precisely how the standardized colors and placements of official road signs, such as the excellent green US Interstate Highway System signs, help in noticing them.) In short, I saw what the instructions told me to see. Told about a "three-way fork to the right"

and unable immediately to see such a thing in that place, I *worked*, successfully, to see it. Told about "the first [place] on the left," I unquestioningly saw what the instructions implicitly told me was there, namely the road on which my friend's house was located.

(3) A friend recently taught me to fold origami paper cranes. In walking me (and a couple other people) through the various steps, she often had to explain by some combination of words, pointing, and demonstrations where and how to fold next. Not having grown up around origami, we didn't have a ready vocabulary for the intermediate forms of a folded crane. The intermediate forms don't look much like cranes. Indeed, much of the magic of crane-folding is that the paper keeps taking on unexpected new identities as you fold it, even long after you've gotten reasonably proficient. These forms can be hard to talk about because they're importantly asymmetrical in nonobvious ways. Even demonstrations are only of limited use if you can't see these asymmetries for yourself. In the course of her explanations my friend said things like "put your finger in the pocket," "fold it back to make a boat," and "make the legs skinnier." Making each of these metaphors refer to parts and aspects of the folded paper always took considerable effort, even though it was always wholly evident in retrospect.

> In the "pocket" case, you can't see the paper as having a pocket (two of them, actually) unless you hold it in a particular orientation and visually, and then physically, separate the flushed edges into peculiarly asymmetrical groups.

> In the "boat" case, the paper becomes boatlike out of nowhere once you manage to reverse the sense of two nearly invisible internal creases and then carefully induce a third.

> In the "legs" case, you're holding a flat diamond-shaped hunk of folded paper that has a slit from one corner of the diamond to its center, but that slit isn't obvious if you hold the paper in the flattened-out form the diamond shape invites. The cranes themselves don't have legs. The "legs," most improbably, become the neck-and-head and tail of the crane shortly afterward.

In each case, the instruction was clearly metaphorical. The speaker hoped to identify certain parts and aspects of the folded paper structure by likening them to pockets, boats, and legs. It was not obvious at first to any of the apprentice paper-folders what structures the speaker was referring to. In fact, much of the speaker's job was to get us to look at our partly-folded origami cranes in just the right way, so that certain parts and aspects would stand out as units for us. (The ones who first 'got' the legs showed the others by grabbing one 'leg' in each hand and making them 'walk.') She was teaching us the skill of seeing our paper *as* having a pocket, a boat, or legs. Although we got better at this skill, it never stopped taking work. The work only became more routine.

(4) The final example comes from my experience teaching people to program computers. If you've long been comfortable sitting in front of a computer terminal, it's easy

to be mistaken about what you're actually teaching. You've got all kinds of neat theory in your head about keyword arguments and normal order evaluation, but none of that is much help to someone who hasn't yet even got the idea of being 'in' the editor as opposed to being 'in' the interpreter. So I sit the student at the keyboard and tell them what to look at and what to type. As they become fluent with the mechanics, my instructions grow more abstract.

> At first it'll be "Type open paren then DEFUN space ...." Often I'll have to explain that "paren" means "parenthesis" or that I'm not asking them to type the words "open" and "paren." Sometimes I'll even have to point out where the parenthesis keys are. Later, though, I'll be able to say "Let's define a function called FACTORIAL."

> When they're learning to read code, you have to point out that there are conventions about indentation that result in common types of code having characteristic shapes.

> Later on, when teaching about using function or class definitions to implement abstractions, you can talk abstractly about hierarchies and common abstractions all day, but it won't do much good until you explain that two hunks of code that look alike are often good candidates for a common abstraction.

In each case, explaining the formalities doesn't do much good until you've managed to connect this knowledge to the student's eyes and fingers.

As these examples illustrate, my prototypical type of representation is natural language, whether as spoken utterances or written texts or so-called 'internal speech.' In each case, figuring out what in the situation the text was talking about always took work. This work often involved creative improvisation and sometimes it required laborious poking at paper and keyboards, looking around, trying over and over, asking for help, or making use of the proddings of a teacher. The work was the work of relating natural language to concrete situations: identifying the things the words were mentioning, seeing materials under metaphorical descriptions, and heuristically associating visual patterns with verbalized technical abstractions. The work required to make sense of "turn left" or "put your finger in the pocket" or "define a function" might differ greatly in different settings or under different conditions.

Does this mean that I disbelieve in AI representation schemes that resemble various extensions or generalizations of first-order logic? Yes it does. I don't resent their formality. Formality is a tool that can be used or abused. Instead, I resent their lack of any account of the work involved in relating a representation to the outside world. Simply positing a correspondence between "John" and John is no help at all.

I want to concentrate on representations written on paper. The image of standing in a kitchen or on a street with a written text in your hand is a good reminder that relating

texts to circumstances requires work and that this work requires understanding in some measure what you're doing. You have to understand what you're doing since the text certainly doesn't. Obviously representations often influence your actions, but you don't understand what you're doing in virtue of owning them. This point is supposed to apply equally to all forms of representation, not just writing. That understanding does not reside in representations is a difficult and consequential idea. The paper's later stories will explore this idea in the context of 'internal language.'

(For readers who know my previously published work with David Chapman, I should point out that the idea of representation I'm discussing in these notes is different from the idea of "indexical-functional representation." From a philosophical point of view, I'm not sure that the entities and aspects of indexical-functional representation should be called representations at all.)

Many words in these essays are in scare quotes. Since scare quotes are not very popular, I should explain briefly why I do this. Double quotes are always simple quotations. Single quotes, though, indicate that I mean to call a word into question. Even though I usually don't want to use the single-quoted word myself, I also don't mean to insult the people who do. Simply quoting a word does not demonstrate its uselessness, nor is it intended to. When I quote a word, I'm asking you to momentarily suspend your routine technical reception of it and listen to its connotations as an ordinary English word, sodden with history and metaphor. For example, in quoting 'content' I want to call up images of containers and contents. In quoting 'structure' I want to recover associations to buildings, to construction, and to other cognate words and ideas. In quoting 'world' I want to draw attention to such manifestly peculiar technical uses of this word as the phrase 'world model.' In each case, attention to words reopens the question of what metaphors are appropriate and useful in formulating technical vocabulary for explaining the lives of agents, whether natural or synthetic.

Nine fairly independent notes follow.

The homunculus and the orbiculus is a fairly glib assault on the notion of a world model, placing it squarely in a long philosophical tradition of attempting to explain the human ability to act competently in the world by pretending that the relationship between person and world is reproduced inside the person's head. This sort of explanation is curiously seductive because it plays to the principal strength of current computational technology, namely building abstractions inside of computers that are almost entirely cut off from the outside world. But it's also fairly easy to find reasons to doubt such views. Not much purpose is served in dwelling on the point, though, since I suspect these doubts will suddenly seem completely obvious once viable alternatives start coming on the market.

Writing as bad and good metaphor for representation makes a first attempt at defining my thesis. It contrasts two very different ways in which one might take written texts as prototypes of representation. The first, "bad," way focuses on the physical properties of

written texts, a long list of which are shared by the sorts of representations one implements with datastructures inside of computers. I suggest that this view of representation derives from a long line of misguided philosophical biases. The second, "good," way of taking writing as a metaphor for representation concentrates on less immediate aspects of written texts. These concern the fact that both the text and your surroundings are outside of you and the consequent necessity of doing work to see the world as being what the text is talking about.

Why we watch videotapes dissects an example of this point. It concerns a videotape of two people assembling a sofa from a kit using a set of instructions printed on a sheet of paper. These people continually had to work together to make the instructions refer to the physical objects they found in the box and they didn't always get it right. The ways in which they made mistakes led to useful hypotheses about the nature of plans. The instructions were a representation of action, but in using them the people had to supply their own interpretive effort and overall comprehension of their task. The process of using the instructions was thus heavily influenced by the nature of interpretation.

A story about photocopier supplies also concerns written instructions. In this case, a secretary is justifiably annoyed because somebody has put laser-printer dry toner in the photocopier. What happened and why? At issue is the selective use people must make of the representational materials that surround them. I put the case for the culprit's defense, arguing that the phrase "dry toner" on a bottle is not 'ambiguous' in a way that anybody could be expected to notice, even if evidence serving to identify and resolve the ambiguity is readily available elsewhere on the bottle's label. One would like to demand that photocopier users be 'careful,' but it's hard to formulate the demand very usefully given that such a problem could hide behind any of the vast number of unarticulated assumptions that form the background of any such activity.

A story about a road sign concerns two words, "Connecticut Avenue", on a road sign. It is a story about a mistake I made because I had not yet learned some fine points of using road signs when I first learned to drive. It's a trivial point that the connections between symbols and things differ depending on the occasion; this story emphasizes that the *methods* by which we connect symbols and things themselves differ depending on the occasion. The methods at issue in this case are bound up with broader legal ideas that assign juridical identities and official, designed conventions to roads, lanes, intersections, signs, and so forth.

A story about some instructions at a performance in an art gallery is another story about instructions, this time a single fairly long sentence spoken (actually hollered) by one person to a group of others. These instructions did not function very well because they were delivered in a context that did not permit their recipients to see, or even imagine, what in the setting the instructions could have been talking about. Some sort of schematic conjuring up of a context often seems necessary to make much sense of an utterance, but here no such conjuring was possible since the structures to which the

instructions referred had been built specially for this performance and thus had never been seen by the people to whom the instructions were addressed.

Two stories about keeping count recounts two stories about trying to keep a running count in my head during a repetitive activity. In the first story I was assigning successive letters to sections of a thesis chapter; in the second story I was counting laps I was running on a track. In each case an odd effect occurs, which both I and others have noticed elsewhere: when it came time to increment the count, I couldn't remember whether the count in my head was the previous index or the next one. I don't know exactly what accounts for this effect, but it is relevant to this paper's themes in that it reflects in a very simple way the plain ambiguity of representations. A remembered letter or number doesn't have its meaning written on it; relating it to one's activity requires interpretive work that can, like any interpretation, suffer from the possibility of multiple interpretations.

A story about my routines for reading the Sunday Globe is a story about an instruction that I hypothesize I issued to myself in the course of reading the newspaper one Sunday morning. The pattern of activity in which the instruction participated had long since become perfectly routine. Nonetheless, the way in which the instruction went wrong, I argue, reveals that it was something very much like an English imperative and not, for example, a precise specification of my intended action stated in anything like a programming language. In this case, a change in the environment led me to make a different sense of an ambiguous phrase than I used to.

A story about the file cabinet in my office is a similar story with a different outcome. Once again there's an instruction I hypothesize I issue to myself in the course of a certain routine pattern of activity, namely fetching stuff from the file cabinet in my office. Here a change in the environment led to a recurring mistake since, I suspect, this phrase continued to pick out the same drawer in the file cabinet even though it should now have picked out a different one. The juxtaposition of these two stories compels some tough questions about how routine use of internal speech works, or even what it means. All I hope to establish so far is the idea that it exists. The work of making sense of your representations in every next situation never goes away, no matter how routine this work becomes.

The purpose of all these stories is not to prove any general propositions. Instead, the stories invite you (perhaps even induce you) to be aware of similar phenomena in your own experience of everyday representation-use. There is immense power in the reciprocal influence between computational investigations and awareness of the manifestations of technical issues in everyday experience. Parallel pursuit of these two kinds of inquiry will, I believe, lead to deeper understandings of why our everyday life is the way it is, why people are the way they are, and why the robots we hope to build can take certain forms and not others.

The homunculus and the orbiculus

In the old days, philosophers accused one another of associating with a sneaky individual called a *homunculus*. From Latin, roughly "little person." For example, one philosopher's account of perception might involve the mental construction of an entity that 'resembled' the thing-perceived. Another philosopher would object that this entity did nothing to explain perception since it required another person, the homunculus, to look at it and determine its identity and properties. Philosophers have been arguing about this issue for centuries. Computational ideas have a natural appeal to philosophers who care about such things because they let one envision ways of 'discharging' the homunculus by, for example, decomposing it into a hierarchy of ever-dumber subsystems.

I think, though, that the tangled ruts of argument about homunculi distract attention from a more subtle and telling issue. If the homunculus repeats in miniature certain acts of its host, where does it conduct these acts? The little person lives in a little world, the host's surroundings reconstructed in his or her head. This little world, I decided, deserves a Latin word of its own. So I talked to my medievalist friend, who suggested *orbiculus*. One way to say "world" is *orbis terrarum*, roughly "earthly sphere." But *orbis*, I am told, extends metaphorically in the same ways as "world" in English: one might speak of the world of a peasant or a movie director, meaning roughly their existential world, "the world they live in." So the orbiculus is one's world copied into one's head.

Where can we find orbiculi in AI? All over. A 'world model' is precisely an orbiculus; it's a model of the world inside your head. Or consider the slogan of vision as 'inverse optics': visual processing takes a retinal image and reconstructs the world that produced it. Of course that's a metaphor. What's constructed is a *representation* of that world. But the slogan would have us judge that representation by its completeness, by the extent to which it is a thorough re-presentation of the corresponding hunk of world.

You'll also find an orbiculus almost anywhere you see an AI person talk about 'reasoning about X.' This X might be solid objects, time-extended processes, problem-solving situations, communicative interactions, or any of a hundred other things. The phrase 'reasoning' about X suggests a purely internal cognitive process, as opposed to more active phrases like 'using' or 'acting upon' or 'working with' or 'participating in' or 'manipulating' X. Research into 'reasoning about X' almost invariably involves all-out representations of X. These representations will be judged according to whether they encode all the salient details of X in such a way that they can be efficiently recovered and manipulated by computational processes. In fact, discussions of these 'reasoning' processes often employ metaphors of actual, literal manipulations of X or its components. In practice, the algorithms performing these abstract manipulations tend to require a

choice between extremely restrictive assumptions and gross computational intractability (see Brachman and Levesque 1984, Chapman 1987, Hopcroft and Krafft 1987).

If you favor slogans like 'using X' over slogans like 'reasoning about X,' someone will ask you,

> "But we can reason about things that aren't right in front of us, can't we?"

Mentalism offers seductively simple answers to many questions, and this is one of them. According to mentalism and its orbicular slogans, reasoning about a derailleur proceeds in the same way regardless of whether the derailleur is in front of you or in the next room. Either way, you build and consult an orbicular derailleur-model. If having the derailleur there helps you, it is only by helping you build your model.

This is, of course, contrary to common experience. As we all know, the first several times you try to reason about a derailleur (1) it has to be sitting right in front of you and (2) you have to be able to look around it, poke at it, and take it apart. I've disassembled and reassembled several derailleurs. Yet I just tried and failed to figure out how a derailleur changes gears, or even to list the parts involved. I've got the rough idea, but that's all. I could do it right away it if you gave me a derailleur, but sitting here with my notebook there's no way. The question arises, why aren't several disassemblies and reassemblies of derailleurs enough to build a mental model of them? Why not just one? One receives several answers to this sort of question, but my favorite (and also the most common) is what I call the *gratuitous deficit* response. For example,

> "Maybe you build a model but it decays."

> "Maybe there isn't enough capacity."

Look what we have here. We have a theory that makes everything easy for itself by arraying before itself encodings of every salient fact in the world. And then we have this sort of lame excuse that pops up to disable the theory when anyone notices its blatant empirical falsehood, without at the same time admitting that anything is fundamentally wrong with it. Maybe the computational complexity of reasoning with realistic world models is trying to tell us something about why it's harder to think about a derailleur in the next room.

Maybe what we store when we gain experience with a derailleur or a city or a recipe is more specific. Perhaps it is more biased to the specific things you've had to remember in the course of the activity. Perhaps it is more closely tied to your goals at particular moments of the activity. Perhaps it is more organized around the experience of the individual situations that arise in the course of the activity. These are difficult ideas. The question is complicated and messy and poorly worked out. It's hard. But that's to be expected. Expecting it to be easy is a sign of addiction to the easy answers of the orbiculi.

Writing as bad and good metaphor for representation

The notion of symbolic representation is near to the hearts of most everyone in artificial intelligence. Indeed, within the technologically informed human sciences, cognition is almost universally understood to involve, in some fashion or another, the manipulation of assemblages of symbols called representations. The vast majority of this research assumes symbolic representations to have certain properties: they are

objects (neither events nor processes),

passive (not possessing any sort of agency themselves, though *cf* Birnbaum),

static (not apt to undergo any reconfiguration, decay, or effacement, except through an outside process or a destructive act of some agent),

structured (composed of discrete, indivisible elements whose arrangement is significant in some fashion),

visible (can be inspected without thereby being modified), and

portable (capable of being transported to anyone or anything that might use them without thereby being altered or degraded).

(Latour makes a similar list.) Although the cognitivist understands symbolic representations as abstract mental entities, observe that all of these properties are shared by texts written on pieces of paper. Indeed, words like 'structured,' 'inspected,' 'modified,' 'transported,' and 'altered' must be interpreted metaphorically, by extension from similar operations performed on physical materials such as paper, in order to apply to abstractions inside of computers. Observe also that most of these properties are either deficient or absent for spoken utterances, which evaporate as quickly as they are issued and which can only be decomposed into discrete elements with a certain amount of wishful abstraction. Thus we can speak, within limits at least, of a writing metaphor for representation.

This conception of representation-as-writing is topical for several reasons, some within artificial intelligence and some in other fields. Within the field, the growth of the connectionist movement has lent urgency to the seeming conflict between symbolic manipulation and the sorts of relatively simple, uniform, statically and locally connected, highly parallel hardware generally imagined to compose the human brain (see for example Elman, Fodor and Pylyshyn, Hutchins, Rumelhart *et al*). Within anthropology (broadly

speaking), the work of Goody, Ong, Harris, Latour, and others has challenged views of cognition that, they claim, tend to make universal principles out of certain psychological and social phenomena found only in literate cultures. (For formal citations see the end of the paper.)

This section has three purposes. First I want to present evidence that symbolic representation in artificial intelligence is, historically, modeled on written texts, as opposed to (say) photographs or spoken utterances. Then I want to describe some ways in which writing is a bad metaphor for symbolic representation. These arguments implicate some prevalent technical methods. Finally I want to describe some ways in which writing is a *good* metaphor for symbolic representation. These arguments suggest some new technical directions, a few of which are being taken up in various projects.

Representation as writing

Much of the relevance of philosophy and linguistics to AI derives from the many habits of thought that AI has inherited, largely unchanged, from these fields. Roy Harris, among others, has argued that ideas about representation in philosophy and linguistics have been heavily biased by writing. He observes that these fields have given a central position to those aspects of human utterances that survive a conventional written representation. One might read in a textbook,

> Suppose that John says to Mary,
>> Please close the window.

and this will be taken to be a sufficient specification of some (usually hypothetical) event. We do not normally wonder, and only very exceptionally are we told, about several aspects of John's action:

- his tone of voice,
- his articulation of the various phonemes,
- the shape of his intonation,
- the timing of the various elements within the utterance,
- the timing of his utterance relative to other actions and events,
- his position relative to Mary and the window,
- his posture,
- his gestures,
- his facial expression,
- the direction of his gaze,

12

- whether and when he has caught Mary's gaze.

(For the horrors of trying to make written notations of these things, see Heritage or Levinson for an introduction to Jefferson's notation system used in conversation analysis.)

Given that these aspects of normal speech regularly affect the import of utterances, a written sentence must be considered a terribly poor representation—at best a coarse idealization—of a spoken utterance. But philosophical and linguistic analyses have almost always proceeded solely on the basis of this idealization, a tradition that AI has carried on. The point is not that any of these fields talk about writing; only that they concentrate on the aspects of representation that writing normally captures. See Harris for a detailed critical history of the role of writing in the philosophy of language.

It is understandable that tone, articulation, and the rest should have been left out of account by philosophers and linguists whose principal medium of communication is written scholarly texts. This tendency has been reinforced by the practice of reasoning about hypothetical scenarios such as John's request to Mary. When one describes a hypothetical situation in writing, the aspects of that situation that one's writing practices don't capture don't ever get specified. As a result, theories will naturally tend to lean on distinctions that writing captures, not on the many distinctions it doesn't.

That's philosophy and linguistics; how about AI? AI's technical practice has been influenced in a hundred mutually reinforcing ways by philosophical and linguistic habits of thought. But among the many routes by which the writing metaphor into AI practice, one particular moment stands out, namely when Newell and Simon invented symbolic programming.

Here's what happened. Almost all the domains that Newell and Simon have studied, especially in their earlier work, have been domains like cryptarithmetic in which the 'world' with which an agent interacts is made up entirely of a sheet of scratch paper. Newell's production system models do not contain separate mechanisms for the scratch paper and for the agent's 'short-term memory.' No separate process parses the ordered markings on an image of the paper and extracts the information needed to decide which productions to fire; instead that information—all the information available on the scratch paper and all its salient relationships—is always automatically available to the mechanism. As a result, Newell's notion of short-term memory is, to put it mildly, influenced by the imagery of scratch paper. Newell and Simon invented symbolic programming in order to implement the sorts of structures and operations that their models specified. List structures, like scratch paper, and like the symbolic structures of all subsequent AI programming languages, have all the aforementioned properties of writing and none of the aforementioned properties of speech.

Writing in the head

13

People invented writing for several good reasons. One of those reasons is that you can do things with paper that you can't do in your head. And there are good reasons why you can't do those things in your head. One is that there's nothing in your head that's anything like paper.

AI research is very often caught in a pattern whereby mechanisms that seem extremely 'expressive' and 'powerful' and 'flexible' and 'general' turn out not to scale up. Let's return to the properties that AI has ascribed to symbolic representations—objects, passive, static, structured, visible, and portable—and consider how they lead to difficulties of scaling and implementation.

Symbolic programming languages such as Lisp clearly endow their datastructures with all six properties. It implements these properties using pointers. Pointers make objects visible to processes. Pointers connect the components of structures. One can effectively transport a structure by 'passing' a pointer to it. Structures only change when processes change them. Pointers are typically realized as binary encodings of memory locations, but what's important is the metaphor of pointing. Pointers do not obey any locality beyond that of their own connectivity. Thus they are eminently reconfigurable. Every variable in every process and every component of every structure is capable of pointing at any object in the machine at any time. (For related discussions of these points, see Chapman's *Connections, encodings, and descriptions.*)

At least two sorts of difficulties arise in computing with pointers. The first is that they require their implementation medium to be infinitely reconfigurable. Thus they pull hard both against the locality of the physical space in which that implementation medium must be realized and against the locality of connection of any particular machinery. Everything can't be physically connected to everything else. This difficulty manifests itself in various ways. On serial machines we observe it in the complexities of dynamic storage management and its interactions with storage technologies (especially the tendency to virtual-memory thrashing and the resulting need for compacting and localizing garbage collection). On parallel machines we observe it in the complexities of shared-memory management and message routing (especially its probabilistic nature and the resulting need for complex and hard-to-understand synchronization and arbitration protocols).

The second characteristic difficulty of computing with pointers is at the level of algorithmic complexity. Just as you can write any symbol you like on the next bit of blank paper, a pointer places very little constraint on what might be found at the other end of it. Even when pointers are typed the types typically have large ranges. As a result, algorithms suffer from the combinatorial arbitrariness of the structured objects with which they must reason. Thus the exponential terms that stubbornly crop up in analyses of algorithms sufficiently general to account for the seeming range of human reasoning abilities.

If writing is a good metaphor for symbolic representation, then, it is not because we've got things in our heads that're objects, passive, static, structured, visible, and portable.

14

These properties of writing don't help us to understand human use of symbolic representation *in general* because they're precisely the properties of written texts that are *specific to written texts*. Far from picking out the essence of symbolic representation, they dwell on the physical activity of using a written text: inscribing, passing about, inspecting, storing, and destroying. Writing was an important innovation precisely because it was the first technology of symbolic representation that permitted these extraordinarily useful forms of activity. Activity of these forms has to go on outside our heads because it's not practical for them to go on inside our heads.

Writing as representation

In what ways, then, can writing serve as a model of the essence of symbolic representation? Consider a situation in which you're using a sheet of paper with some writing on it to help you organize a concrete activity. Perhaps you're using a recipe to help you cook dinner. Or perhaps you're using some directions to help you get to a party. The sheet of paper has a funny paradoxical position in this scene.

> Even though it's a physical object with a definite size, mass, and location relative to the other materials in your surroundings, it plays its role—at least qua representation—entirely through your interpretation.

> Even though it seems to speak perfectly clearly about the materials of this particular situation, it only manages to do so because of the work you do to figure out what in your surroundings it's talking about.

The sheet of paper in your hand is both part of the material situation and doubly removed from it, separated from it by you yourself.

> It underdetermines the sense you make of it because it is separate from you—after all, someone else faced with the same materials under the same conditions would probably do something different.

> And it underdetermines what in the situation it picks out because you are separate from your surroundings—after all, you could probably use the paper equally well with different materials under different conditions.

As with written texts, I want to claim, so it is with all symbolic representations. Where a written text is outside you in a physical sense, all symbolic representations of whatever form are outside you in a particular metaphorical sense. In your head or in your hand, all texts are fundamentally indeterminate.

15

This view of symbolic representation conflicts with the widespread proposal—manifested by, for example, the model theory of first order logic—that a text has a 'meaning' independent of the activity of any individual. The world does not come innately parcelled out into the categories we find mentioned in written texts. Instead, *people use representations to help them make sense of particular situations. What a given text is talking about is a fresh problem in every next setting.*

(For those who care about such things, this is what Jacques Derrida means by the word "writing." His assertion that philosophers have suppressed writing has caused much confusion when interpreted more closely to the sense of representation-as-writing that I denounced earlier. For introductions to Derrida's philosophy see Culler and Norris. I have also been influenced by Garfinkel's ethnomethodological ideas about representation use. For an introduction see Heritage.)

Likewise, 'understanding' a text is not a matter of possessing something called its 'meaning.' Instead it is the ability to use the text in the situations one encounters—or perhaps the situations one *might* encounter. It does not matter whether an individual's efforts to make sense of representations in successive situations are infallible, systematic, consistent, or readily articulable. All that matters to the average representation-user is that the job get done.

This idea has many consequences for computational theory. To describe cognition as a matter of processes operating upon symbolic representations is much less help than we had once hoped. Indeed, the relationship between internal processes and their internal symbolic representations is qualitatively the same as the relationship between a whole rational agent and a sheet of paper. Even if we put symbolic representations inside the heads of our robots, regardless of their properties and regardless of whether we intend our robots as models of human beings, we have not qualitatively reduced the problem of specifying the essence of rational action.

These consequences are negative. What can we say that's positive? For the moment, we can speculate about alternate technical forms that theories of symbolic representation might take. I believe that our ability to use symbolic representation is formed in interaction with the practices of representation use we encounter in our everyday activities. These abilities have as many forms as there are forms of representation use around us: speaking and hearing give rise to internal speech, creating and inspecting physical symbols give rise to visualization, and so forth. The 'internal' processes—I quote the word because I doubt if they're best considered as wholly internal, rather than as elements of complex forms of interaction—aren't precisely like the 'external' processes because the insides of our heads aren't like the outsides. Instead, they are strongly shaped by the machinery we have in our heads and more generally by the practicalities of physical realization of complex forms of activity. For example, I strongly suspect that internalized speech is far more consequential than internalized writing because our machinery is far better suited for reproducing speech than writing. (For a great deal of insight about the

nature of internal speech see Wertsch's survey of the Vygotskian school.) If this is true, then many properties of speech that philosophy and linguistics have treated as marginal phenomena—tone of voice and intonation come to mind—will have to be readmitted to our theoretical center stage.

Why we watch videotapes

One evening recently I learned something valuable about plans as Lucy Suchman, Randy Trigg, John Batali, and I were watching a videotape of two fellows, whom we'll call John and Paul, assembling a sofa from a kit. The point of this story isn't so much what I learned as how I learned it. It's a story about the value of a cultivated discipline of videotape watching. In particular, it's a story about learning to locate where you end and others begin.

In this tape, John and Paul make a mistake. Fairly early on in the process, they put some washers where they don't belong. They discover their mistake much later on when they find they don't have enough washers to put the last few parts together. The question is, why did they make their mistake?

Their mistake certainly wasn't caused by inattention to the instructions. They spend the first 12 minutes of their 40-minute-long project just reading the instructions, relating each line of instructions to the materials they've removed from the box. They work carefully and try to form articulated reasons for their actions. This makes their mistake a compelling study in what people do with representations.

The kit provides them with a collection of dowels, bolts, nuts, and washers for putting this sofa together. These parts first come into play when they form the "base" of the sofa by affixing two "rails" and a "base spreader" to two pre-assembled rectangular "base leg assemblies." Step A of the instructions reads:

> Place a 7/16″ × 2″ dowel (1) in each end of the front and rear rails (2) and attach to the leg assemblies (3) as shown. Secure each end with 3 1/4″ bolts (4), washers (5), and nuts (6). Place the base (short) spreader (7) between the leg assemblies and secure each end with a lag screw (8) and washer. Be sure all the bolts are tight.

The parenthesized numbers correspond to the list of parts in the upper left corner of the page of instructions. At issue is the second sentence, which (dropping the numbers) reads:

> Secure each end with 3 1/4″ bolts, washers, and nuts.

The issue is where the washers go. This sentence is not clear. Is there a washer under the head of the bolt, under the nut, or under both? They end up putting a washer at each end of each bolt.

After a great deal of stepping frame by frame through several segments of the tape we finally found the moment where John and Paul made their mistake. After much effort, they are finishing up with the first sentence of the instructions:

> Place a 7/16″ × 2″ dowel in each end of the front and rear rails and attach to the leg assemblies as shown.

This takes them about five minutes. Before they're quite done with this, though, Paul turns his attention to the bolts, washers, and nuts. After sorting them out under himself, he lifts one corner of the structure and inserts a bolt in its pre-drilled hole. John, however, would clearly like to finish the first step. With Paul occupied with the hardware, he seems at a loss for something to do. He repeatedly tries to account for what Paul is doing. He asks Paul whether he's experimenting and if that's the next thing for them to be doing, but he doesn't get much acknowledgement from Paul.

At this point something fascinating happens. John picks up the instruction sheet, which has been lying by his side. While John is inspecting the instructions, Paul removes the bolt he has inserted into the side of the base assembly, brings it down to the floor, puts a washer on it, lifts it back up, and re-inserts it in its hole. John then looks up from the instructions and back toward Paul.

This sequence was entirely obvious in retrospect even though we had already been through this section of tape several times without any of us noticing it. Perhaps, I figured, we had here an explanation of the mistake. Maybe John, occupied with the instructions, had simply not noticed Paul placing the fateful washer on the bolt.

Just now, though, Lucy said something important, to roughly this effect:

> Where in the tape is the warrant for your assumption that the matter of whether a washer belonged under the head of the bolt was an issue for them? It's clearly an issue for us, but is it an issue for them?

We went back through the tape and, indeed, we found no evidence that they had ever thought about it. They were quite concerned with whether there should be washers under the nuts, but they never appear to entertain doubts about the heads.

This is interesting. They have read the instructions carefully with every appearance of understanding them. And when they eventually realize they are in trouble they are able to read the exploded diagram to discover that it does not portray washers under those bolt heads.

The truth, it turns out, is more interesting than my detective story about flukes of synchronization. It has to do with the kind of use that John and Paul were making of the instruction sheet. Certainly they made frequent reference to the instructions; the instructions influenced everything they did. Nonetheless, it is not really accurate to say that they were 'following' the instructions.

The issue isn't whether the instructions precisely specify their users' every move. John and Paul's enactment of the first sentence, "Place ...," takes about five minutes, so they're filling in some details, to put it mildly. Even given this, though, one might hold that the fine structure of their activity was, say, an expansion of standard patterns of activity mentioned implicitly in the instructions, perhaps on the model of the computer-programming notion of a subroutine call.

But, no, it wasn't like this. Rather, John and Paul used the instructions as a sort of oracle. When an issue arose, they would address it to the instructions, working to achieve their coherence, their reference to the materials at hand, and their character as specifying the activity. Naturally a large number of issues arose, as to "what now," choices among similar parts and seemingly symmetrical possibilities for placing them, and so forth.

But plenty of other issues did not arise. One issue that did not arise is whether the bolts attaching the rails to the side assemblies had washers under their heads. We might imagine that Paul, in sorting out the hardware, was working from first principles, deciding to place a washer under the bolt head because bolt heads should not have to turn against wood. But we have—at least so far—no more evidence that he made a positive decision to this effect, a positive choice of this move as against other possible moves, than we have that he made a positive decision to maintain his balance or to keep breathing.

Once I thought about this, I decided that John and Paul couldn't help but use the instructions in this way. When they first read through the instructions, before assembling the sofa, John read the second sentence,

Secure each end with 3 1/4" bolts, washers, and nuts.

out loud, without evincing any sense that there was anything problematic about it. It made sense. Securing wooden parts with bolts, washers, and nuts is something both of them know about. We have no evidence that their understanding of this sentence was sufficiently detailed as to specify numbers of washers-per-bolt, and plenty of evidence that it was not.

John and Paul had no more reason to question how many washers the sentence specified than they did to question how many ends the rails had, whether "each end" really meant each end of each rail, whether one should double the nuts, or any of countless other matters. Nor do we know if they were clear about, say, which direction the bolts should pass through the parts. No doubt if we stopped them and asked they would've come up with the right answer pretty quickly, but that's a different matter. In short, the sentence is no more—and no less—'ambiguous' in the problematic regard than it is in any of these other, more peculiar ones. We could go on all day listing and clarifying the ambiguous aspects of the instructions. No sentence, or any other hunk of text, has a clear, definite, unique meaning that anyone could finish specifying, whether by playing

twenty questions or in any other way. A sentence is a mutely indeterminate string of symbols until it is brought to bear on some actual situation. Plans are no exception.

If John and Paul were using the sofa-assembly instructions as oracles rather than as direct specifications of action, a new and interesting problem opens up: in virtue of what were John and Paul able to make this use of the instructions? It's not exactly that they already knew what they were doing. But they did possess some fundamental competence which the instructions served only to supplement. The instructions played this supplementary role by, among other things, directing John and Paul's attention to materials they were already capable of comprehending and by answering specific queries John and Paul were already capable of formulating.

Critical to my having learning what I have just reported was Lucy's advice to try separating issues-for-us from issues-for-them. Lucy picked up the habit of making such distinctions from the ethnomethodologists. These people are utterly obsessive empiricists who have developed and articulated a practice of distinguishing the interpretations for which you have warrant from the ones for which you don't.

Central to the ethnomethodologists' practice of interpreting videotapes (or anything else) is the idea that the members of a social setting engage in a continual, cooperative construction of the situation's reality—a fiction, so to speak. If we, as observers trying to understand the methods by which people get along, join the participants in taking these fictions for granted, we won't even be able to ask our scientific questions, much less answer them. Likewise, if we read our own concerns and questions back into the activity we are studying, as if its participants' reality transparently subsumed ours, we will certainly ask only wrong questions. In short, we must try to locate the boundaries between ourselves and the people on the videotape.

That we can never finish demarcating these boundaries is a consequential phenomenon in itself. It has important negative implications for any methodology that would present a finished, objective set of rules for studying human activity. It also has important theoretical implications as a positive fact about human relationships. In the course of our cooperative activity, we rely in endless ways on the imperfection, indeed the superficiality, of our separation. Before we have our own private thoughts, before we can fall prey to insidious and tragic failures of communication, before we even feel the need to engage in positive acts of imparting information to one another at all, we dwell together in a radically inarticulable common world whose roots reach bottomlessly into our commonalities of environment and biology and cultural heritage. Insofar as this common experiential world is the site and basis of so much of our learning, its properties must be central to any scientific theory of human cognition.

A story about photocopier supplies

When I used to have jobs where I worked behind a counter, I was always amazed how oblivious people are to signs. If the machine was going to be back up in an hour, you could put a big red sign with stars and arrows

THE MACHINE WILL BE BACK UP AT MIDNIGHT

anywhere you liked and people would still walk up to the counter and ask "when will the machine be back up?" This can get on your nerves.

Today's text is a manifestation of this effect.

> Date: Tue, 17 May 1988 17:05 EDT
> From: J
> To: All-AI
> Subject: Xerox copiers and Lazer Printers
>
> Due to someone's IGNORANCE, CARELESSNESS, or LACK OF PATIENCE, SOMEONE PUT DRY IMAGER FOR THE LAZER MACHINE INTO THE XEROX MACHINE on the 8th floor. These supplies, although both dry imager, ARE NOT INTERCHANGEABLE!!! It says on the box it comes in (and on the bottle itself) which machine it is for.
>
> We were warned by the Xerox people before that if this happened again, they may discontinue servicing our machines–not to mention the cost of having it corrected (or maybe having to get a replacement).
>
> If there is a problem with the xerox or lazer machine on the 7th floor and you do not know how to correct it or have a question, see D.... She is in charge of the overall care of those machines as I am the machines on the 8th floor (I can be located at ...).
>
> Because of one person's lack of resourcefulness (he/she could of went down to the 7th floor or seen me or D), we are all suffering! If it is after hours and you are not sure what to do, it is better to do nothing than ruin a machine.
>
> –J

In the old days, when the photocopier needed dry toner, one would look around the room, see the bottle marked "dry toner," open the machine, find the reservoir marked "dry toner," and put the contents of the bottle into the reservoir. Sometimes they would

put the toner in someplace else, or they'd put something else into the toner reservoir, and these are very bad things to do to a photocopier, but they didn't happen that often.

Nowadays, though, many rooms with photocopiers in them also have laser printers in them. And laser printers also take dry toner. Except in the copier/printer room on the 8th floor, the two types of dry toner are incompatible. So what happens? Exactly the same thing as before. Except that the bottle marked "dry toner" one happens to come across first might or might not be the correct dry toner.

Does this happen because people are too lazy to check whether it's laser or copier toner? After all, as J points out, "It says on the box it comes in (and on the bottle itself) which machine it is for." Do they just go irresponsibly ahead putting laser toner in the copier figuring it's probably OK? Did they plough ahead despite a nonspecific but conscious uncertainty about how to replace the toner?

Although we must appreciate J's situation, I think all these hypotheses are unnecessary. Put yourself in the place of someone to whom the photocopier is asking for dry toner. This person was making dozens of different moves in the course of their actions. So far as this person could tell, any of these moves could be mistaken in dozens of different ways. The actual problem, namely that the bottle marked "dry toner" was not actually the correct substance, is pretty obscure, as if somebody parked a car nearly identical to yours a couple spaces down. In the case of the wrong car, some discrepancy would no doubt force itself upon your attention before you got too far along. Sure there was evidence of the mistake readily available, but what was the poor victim supposed to do, make a complete list of all the things they might be doing wrong and go looking for evidence to rule out each one?

The problem, in short, is only obvious in retrospect. Types of dry toner are obviously an issue for the person who had to deal with the consequences of somebody's having gotten it wrong, and for us as theorists now, but it doesn't follow that we can expect it to be an issue for anyone who tries to replace dry toner in the copier. It's not necessarily even that the person "didn't know what they were doing." Someone could become perfectly proficient at replacing toner in copiers and still run afoul of this difficulty, simply because whether they had the right type of toner had never become an issue for them. The arrival of the laser printer would invalidate one of the implicit assumptions of their toner-changing routine, but they wouldn't necessarily have any way of knowing this.

I interpret this story in terms of a larger theme about what people do with representations. Recall the story about the two guys trying to assemble a sofa from a kit. Given an instruction that was ambiguous about the placement of bolts, they did what seemed right without it ever being an issue for them where exactly the bolts belonged. This case is similar but a little more complicated. I suspect that the person in question, seeing that dry toner was needed, looked around the room for containers marked "dry toner." Finding such a container, they proceeded as if nothing was wrong. As far as they could

tell, nothing *was* wrong. The phrase "dry toner," to us, is 'ambiguous' between "laser dry toner" and "copier dry toner." But that ambiguity is only an ambiguity for us. For all we know, "dry toner" could also be ambiguous in an unlimited variety of other ways.

One might say, the person in question should have read the rest of the label. But the world is full of representations; how do you decide when to stop reading them all and start doing something? We might hear this should-have as an instance of the Ur-heuristic: look around. Looking around has a way of making things become issues for you. But the person wasn't in doubt and had no reason to be ...

The dispute can go on and on. In the end, all we can do is stop moralizing and fix the equipment. Make all the toner bottles so different that it's physically impossible to install the wrong stuff. Or try to reduce the risk by training everyone. Or resign yourself to locking up the toner bottles.

Here, by the way, is a second recent manifestation of this phenomenon:

Date: Thu, 19 May 88 11:21:30 EDT
From: D
To: 7ai, 8ai
Subject: [something else]

. . .

I cleaned up the ninth floor printer area and carted off the two boxes of paper for recylcing. Please note that there are approximately 10 opened reams of 3-holed paper up there. Perhaps the literacy rate is low for ninth floor printer users.

A story about a road sign

I was 17 and hadn't been driving very long. Having grown up in a sterile outer suburb, I was unfamiliar with the conventions and street knowledge, so to speak, involved in city driving. One particular evening I was driving to someplace in the northern suburbs of Washington DC, an area to which I had never been, though having grown up nearby I had heard about it and knew the names, but not the relationships, of the major roads. I was heading west on Adelphi Road (or perhaps it was University Boulevard) and I wanted to take a left on Connecticut Avenue, so as to be heading south toward the city. The road had curved and gone through some difficult shifts and branchings, though, so I was aware of being somewhat disoriented. In particular, I was unsure whether I was still on Adelphi Road. As a result, I was investing great effort in inspecting the territory for evidence.

At one point, I saw in the median a short, wide sign on a pole, white with a black border and black sans serif lettering, reading "Connecticut Avenue" (or perhaps "Connecticut Ave"). "Aha," I said to myself, "I'm already *on* Connecticut Avenue!" A queasy uncertainty accompanied this insight, but then 17 is an age of more or less perpetual queasy uncertainty. The sign was about 50 feet short of a major intersection, but this did not make much of an impression on me since, as far as I was concerned, I had figured out where I was and was in the right place. Incongruities accumulated over the next several miles, though, and I realized I had made a mistake. I then quickly figured out what the sign meant and in general what such signs mean.

It is often said that the relationship between representations and reality is mediated by social conventions. This slogan is often interpreted as meaning that society posts a table of conventional symbol-thing correspondences to which one can refer in decoding a particular symbol. Quite the contrary, the social conventions around representations concern the methods by which one is to discover what things a symbol is talking about in particular situations. In this case, I had to learn a method for relating certain kinds of signs to a certain setting's materials (roads, intersections, turn lanes, etc). I was aided in this by some important ideas I brought to the situation: the idea of a 'road' as an individual, of my being 'on' some particular road, of roads meeting in 'intersections', of 'looking for' the intersection between the road I'm on and some other particular road, and of some as yet unidentified intersection 'coming up.' More implicitly, I was exploiting the fact that road signs are designed to support the projects people commonly have along roads and that my particular project just then, of looking for a particular road, is a common project. Since the issues for me just then were few and clear, interpreting the sign as being there *in order to* support my project made it easy to formulate a new idea

about the sign's significance.

Furthermore, I could formulate my insight as concerning 'such signs', not just 'that sign', since such signs stand in a designedly generic relationships to 'such roads' and 'such intersections' when one is engaged in 'such tasks.' The sign's success in standing in a generic relation to the surrounding materials and to my purposes nearly effaces the sign's identity as a particular object. I relate to the sign, now that I've acquired the necessary skills, the way I relate to 'such signs', as opposed to the way I would relate to an improvised sign hand-painted on scrap lumber and nailed to a tree out in the countryside. Someone who had never looked for a road would, I expect, have a more difficult time making any sense of the "Connecticut Ave" sign's relation to the surrounding territory, or even perhaps formulating the idea that the sign *had* a relation to the surrounding territory.

I would like to understand how explicit an idea one needs of the conventional nature of the relationships of official signs to their settings. The officialness of the sign is part of a larger network of official facts: the existence of a road or intersection is a plain, ordinary fact out in the forest somewhere, but in a city it is also a *legal* fact. In Boston we joke that the roads started out as cow paths, but we have a sense that at some point they changed their status, acquiring names and street numbers and significance vis-a-vis laws such as "right turn on red after stop." The cow-path story, however apocryphal, reflects something important about the nature of these roads. The roads and intersections can only be legal facts because they are, or lend themselves to construal as being, in some broad sense, *already* facts for individuals engaged in the 'merely physical' activity of getting somewhere by following roads.

The black-and-white "Connecticut Avenue" sign signaled its official status, in part, through its neat, standardized look. Somehow we understand that only official agencies (which can be governments, corporations, schools, clubs, or just about anything in various contexts) have access to the technology to give things that neat, standardized look and to place them in such convenient, appropriate, significant places as the medians on major roads. Unscrupulous advertisers sometimes play on this assumption, making their newspaper ads look like real news columns or making their envelopes look like those containing official documents or announcing themselves to tourists as "information centers" and the like. The aim, presumably, is to encourage the same generic, automatic, trusting response which such a look normally occasions in other, situations of routine interaction with 'official' representational materials. People are certainly capable of seeing objects as 'made to look like' something they're not, but only once such an idea has occurred to you.

A story about some instructions at a performance in an art gallery

Last Saturday a friend and I went to the MIT reference gallery to see a weird performance put on by the current pair of resident artists. The performance took place in a windowless room that's about twenty by thirty feet with quite a high ceiling. When we arrived the doors hadn't yet opened, so we and the rest of the audience milled about outside. Finally the time came and the person who worked behind the desk wandered out into the lobby where everyone could see her and yelled something like, "OK the doors are open." Then as people were drifting toward the door she moved into the middle of the crowd and yelled something fairly complicated that went something like this:

> You can sit on the chairs or you can go to the back wall and look through
> the windows or you can go up on the balcony, but don't lean against the side
> wall.

None of us could actually see the inside of the performance space as she was saying all this, so we had no way of knowing what she meant by windows, side wall, back wall, balcony, *etc.* One could feel the crowd being uncomfortable about this, many of them turning to their neighbors in an attempt to get clarification. I found myself trying to visualize the scene, but I had no idea how to place even the "side wall," much less the balcony and the windows. Windows? Both the impossibility of visualizing the scene and the effort spent trying seemed to make the instructions unusually hard to remember, as if they were nonsense syllables or Zippyisms, and several people could be heard repeating parts of them over to themselves or to their neighbors.

I found this very amusing, and adopting a gently ironic imitation of the register and diction of the yeller I said something like, "you can stand between the monsters but don't sit on the toadstools." Not very many people got the joke, I'm afraid, and especially not the yeller, who replied, if you see any of those things in there please let me know. I found this interesting in itself. The person who yelled those instructions was obviously quite familiar with the room, and even though she couldn't see it just then any better than any of the rest of us, she presumably had no problem visualizing what she was talking about and no evident awareness that others *were* having a problem.

Once we got in there, there was an audible rush to make the words attach to parts of the room, which was dark and full of elaborate and peculiar wooden structures. The "side wall" is immediately there on your right, verifiable by the readily visible chairs along it, and the "balcony" could be found along the back wall with a little scanning.

27

The "windows" weren't at all obvious; they were windows in the wall supporting the balcony, behind which one could stand. I suggested we go for the balcony; although no stairs were immediately visible it was pretty obvious where they should be and others were already headed that way ahead of us.

In this story, the peculiar relationship between the instructions and the physical setting disrupted a part of language understanding that normally goes unremarked. The yeller's experience of "You can sit on the chairs or ..." differed from the audience's in that only she had a working familiarity with the network of structures and spatial relationships it referred to. For the audience these relationships, lacking any reference to any familiar or imaginable setting, were difficult to remember in the way that arbitrary formal symbols are hard to remember in laboratory experiments. Heaven knows how these crazy artists might have set up this space!

The situation resembles the mnemonist's method of loci: if someone tells you about something that happens in a familiar space—or a space for which you have a familiar model, like imagining someone else's cooking story taking place in your own kitchen—its objects and actions get themselves 'placed,' in a way that can be either quasi-visual or kinaesthetic or, more commonly, both. And they tend to stay where you put them. If you can't put the various elements of the story in their places you won't be able to hang onto them, just as we had trouble hanging onto the performance space's balcony and windows. I don't understand any of this.

Taking language out of context

This story connects to a larger theme about language. I want to believe that an utterance doesn't really have anything like a 'meaning' outside of the particular concrete context it's involved in, but this makes it a little hard to explain how we can talk in the abstract about people and things that're distant or hypothetical. Consider the sorts of examples that linguists ask you to evaluate out of any context, like those sentences about abstract Johns and Bills where you're supposed to say whether it's OK for the pronoun to refer to (a) John (b) Bill. Often I've had the experience of finding an interpretation not-OK until I do a fair amount of work elaborating hypothetical contexts. And I'm not even talking about my habit of imagining weird 'pathological' contexts that can make any utterance "OK," but just interpretations that the linguists and I alike find OK in retrospect.

I suspect that the only reason that such exercises get consistent results, when they actually do, is because of the consistency across a segment of culture about the foggy default contexts that people will construct to make the sentence make any sense at all. The ability to do this *at all* is quite culture-specific; many anthropologists have had

the experience of administering such tests, either grammar tests or syllogistic reasoning tests, and having their informants insist on knowing who John and Bill (or whatever names they use) are, on bringing in additional 'irrelevant' information from their own experience, and so forth. These people just haven't been trained in 'reasoning' about the poorly-sketched decontextualized quasi-situations that're evoked by grammar and syllogism quizzes. (See Cole and Scribner.)

Heath (1986) discusses how this capacity arises. She has studied middle-class child-raising customs, especially the way that middle-class parents use rituals like bedtime-story-reading to introduce children to decontextualized letters, words, forms of speech, and so forth. Children who don't get this training, at least according to Heath, have a much harder time relating to decontextualized school exercises. (Unfortunately, Heath did not pose the question of whether this mismatch reflects the shortcomings of working-class child-rearing or the artificiality of decontextualized school exercises.) None of these abilities are normal; they are quite complex phenomena overlaid on top of the ability to relate language to present contexts or to evoked but concretely familiar contexts. This is unfortunate in a way, because it has led to two quite different phenomena—one of them more fundamental and universal than the other—being run together and confused, with the later, more articulated phenomenon (the ability to perform certain tricks with decontextualized representations) getting all the credit.

Two stories about keeping count

Just now I was in the editor moving through a long file of text putting in section dividers. Each section divider looked like

\section*{a. Introduction}

I would put in a divider, then I would move down through the text a screenful at a time looking for the right place for a new divider. I had taken the precaution of inserting the just-mentioned first section divider into the kill ring so I could yank it out into each next right place. Having done so, I could edit in the correct section letter (b, c, d, ...) and type in the title.

The problem was in remembering the correct next letter. It has to be the next letter in order from the last section's letter, but I couldn't ever remember that last letter with sufficient assurance. So I kept having to move back and find the previous divider and check. This settled, rather annoyingly, into a routine of its own, aided by the fact that the previous search string, now, was always "\se." After three or four of these, I resolved to remember the letter. After typing a new section divider, I would say the letter to myself and try to keep it in my head until it got time to type the next one.

When I got to the next place for a section divider, then, I would have a letter in my head. Unfortunately, though, I wasn't sure whether it was the last letter or the next letter, so I had to go back and check all over again. I think in each case it was the last letter, but I was never sure. This went on for a few more letters until I noticed the effect and started typing this note instead.

This story is analogous to something that happened the first time I ever ran five miles. It was on a quarter-mile dirt track on the Stanford campus. Five miles on a quarter-mile track is a lot of laps. The problem was keeping track of the laps. I could keep a number in my head, but I could never remember with certainty whether it was the number of laps I had completed or whether it was the number I was about to say when I got to the start/finish line again. Since each lap took about three days, you see, I spent a lot of time thinking about that next number along. At the end I ran a couple extra laps to be sure.

These two stories demonstrate an especially simple case of the indeterminacy of representations. The bit of internal speech I was keeping in my head and repeating to myself originally meant "I have just entered section c" or "I have finished 23 laps," but all I was repeating to myself was "c" or "23." This was clearly the experience of it: I would say "c" or "23" to myself, both with and without moving my mouth. These simple strings

have many possible interpretations in various contexts; in these particular contexts two interpretations came to mind for each of them. Perhaps in a slightly different world I would have been unsure among three interpretations, or more.

Of course the two competing interpretations were not arbitrary. In each case I knew perfectly well that the letter or number was in my head because I was trying to label successive sections or count successive laps. That was all I was doing. But the fact that I couldn't keep the possible interpretations straight says something to me. It won't do to interpret these 'values,' "c" and "23," as being kept in specialized registers that were allocated and assigned fixed meanings by some program. I didn't have a last-section or last-lap variable whose value got manipulated correctly in virtue of some invariant a program maintained with regard to it. Instead, I had only a single 'register,' namely my ability to say some English utterance to myself in my head.

Cognitive psychologists speak of this effect in terms of short term memory's being a matter of echoic memory, which is to say that you can only keep an utterance 'alive' by repeating it to yourself. A consequence of this fact is that the contents of 'short term memory' (whatever that is) suffer from all the difficulties of interpretation of any other utterance. As with any other representation, using an internal utterance requires doing work to figure out what in one's current situation the representation is talking about. Being an 'internal' rather than an 'external' representation does not change this fact.

A story about my routines for reading the Sunday Globe

This is a story about the indeterminacy of interpretation of plans, a theme I first outlined in my story about my experiences watching the videotape of two guys putting together a sofa. Here, though, the plan (if we want to call it that) was something I said to myself as part of an old and settled routine.

What I'm hypothesizing is that a routine that starts out being mediated by an imperative utterance—even a command you subarticulate to yourself—will still be governed by all the 'underdetermination' and 'ambiguity' and 'indexicality' and 'ellipsis' of that utterance, long after you've lost all awareness of any English being involved. Just because the activity has gotten 'compiled' doesn't mean that the connection between plan-text-over-here and concrete-situation-out-there becomes any less problematic. Perhaps you don't ever stop saying the plan-text to yourself (whatever that means); perhaps you only have routine patterns of acting on it when the conditions are sufficiently (in a qualitative sense) similar to the conditions that obtained as the routine settled down.

The Boston Globe recently began an expanded (about twenty pages) and loudly touted arts section in their Sunday edition. Called (foolishly) "Arts *Etc*," it gathers together all the Sunday movie, arts, and book reviews, and various arts schedules and advertising, and throws in some high-brow cultural commentary that occasionally reaches the intellectual level of the average review in the weekly Boston Phoenix. This section doesn't have clearly delineated departments except for the last few pages, which are marked off as the book review pages. The book review department, in fact, is wholly unchanged from the pre-Arts-Etc Sunday Globe. The first of the book review pages has its own banner and distinctive format, and all the longer book reviews begin on that page and continue inside, where there are also shorter reviews and lists of best sellers and so forth.

Now, it often happens when I am reading the newspaper that I'll come across the continuation of an article that looks interesting even though I didn't notice it when I was reading the page on which it began. So I'll have to back up to the earlier page to read the beginning of the article.

Last Sunday, then, I was reading a book review in the Globe. In particular, I was in the interior of the book review section, having followed an article from the book review section's first page (which, let us say, was page C15), when I came upon a headline about an author I was interested in reading about. Focusing on this headline, I found that it was a continuation. Whereupon, oddly, I turned to page C1—*i.e.*, the very front page of the whole arts section—not to page C15—*i.e.*, the first page of the book review section. I

knew that all the book reviews began on C15, not C1, but I turned to C1 anyway. When I got C1 in front of me, it was not at all what I expected; momentarily confused, I then figured out I should turn to C15 instead.

Saying "C1" and "C15" is of course very misleading here. I don't think I knew what the correct section or page numbers were. My mistake, I think, turned on my never having reflected on the somewhat odd relationship between the two pages in question: the book review section was a clear "part" of the arts section, but the arts section didn't have any other clear "parts." Both page C1 and page C15—if that's what they were—were the "front" of something—namely the arts section and the book review section, respectively. I had long been familiar with the Sunday Globe book review section's format, and its design and layout did little to make it look like a part of the superordinate arts section.

What I think happened is this. When I went to turn to the beginning of the review I wanted to read, I turned to "the front," perhaps even "the front of the section." I'm not sure exactly what I mean by double-quoting those two English phrases, but I want to mean something fairly literal. That is, I think I made my mistake because I was saying something to myself in English, either subarticulately or wholly within my head (whatever that means), and the phrase in question was ambiguous and I interpreted it wrongly.

I've been reading newspapers for at least fifteen years and Boston Sunday Globe book reviews for almost ten years. Stumbling upon the continuation of an article and wanting to find its beginning is a routine I've probably been through many hundreds of times. And at least a couple dozen of these episodes must have occurred during my reading of the Sunday Globe's book reviews.

I should stop to wonder why I knew to turn to the "front" of the section, since not all articles in the rest of the paper start at fronts of sections. Perhaps at some point I articulated to myself the fact that the book reviews start at the front of the book review section. I don't think I actually inspected the page number in the phrase "Continued from page C15." But a plausible alternative is that I knew that I was only a single page into the section, if only because I know that the book review section is only a few pages long, which meant that a continued article could only have started on the front page.

So this was not the first time I've said "go to the front of the section" to myself in my head in such a situation. This internal uttering-to-myself and the actions I typically have occasion to take in consequence of it must certainly have worn deep grooves in my brain by now. You might think that it was so thoroughly 'compiled' that it no longer resembled English. Yet still it was capable of this very language-like underspecification of the situation. I still had to figure out what the English phrase was referring to in this specific concrete situation, and even though this figuring occurred perfectly automatically, smoothly, and routinely, it was still problematic—just as problematic, perhaps, as it was the first time, at least in the sense that it was still possible to get it wrong.

So "the front of the section" was ambiguous in this situation. But all of this still

doesn't explain why, on the particular moment in question, it led me to turn to C1 rather than C15. Before the Globe reformatted and ballyhooed its silly "Arts *Etc*," I had never given any particular thought to the idea of the Sunday Globe having an "arts section." In fact I clearly recall the first Sunday of the new section: the front page of the paper—*i.e.*, A1—had an ad for it that caught my eye, and despite the dippy name I decided to give it a fair try. And in fact it contained a fairly perceptive article about the Soviet cimena, whose best recent product (a film called "Repentance") I had recently been very impressed by. The matter of "the Globe's new arts section"—in exactly those words—had thus clearly been on my mind. I don't want to conclude that the arts-section interpretation was 'stronger' than the book-review-section interpretation, but whatever is operating as we constantly use background information to 'fill in the details' of utterances when determining their relevance to particular concrete situations was operating here as well.

A story about the file cabinet in my office

Background: the file cabinet in my office

I've got a file cabinet in my office. It has five drawers, but until recently I only had files in the top three. The bottom two were storage for a bunch of junk, much of which was left over from when I shared the cabinet with Daniel Weise—empty files whose tabs had Daniel's handwriting, a pile of Time magazines from back when I was interested in naive argumentation, a couple bags of mardi gras trinkets, a now-empty box of blank cassette tapes, *etc.*

When I was doing my taxes, I finally cleared up a pile of papers that'd been gathering in the vegetable crate I keep on my desk. The crate contained some files in yellow folders for personal stuff—taxes and MIT bills and bank statements and travel records—but most of the papers were unfiled, just heaped at the end of the slumped-over row of files. So I made files for many of these papers: I sorted them into piles and recruited a new hanging file folder for each new category of papers—taxes, MIT, Oxford, airlines, my car, job-hunting, credit cards, *etc.*

Each drawer of this file cabinet has a distinct category of stuff. I have a hard time remembering what the categories are, but when I need a file I almost invariably head for it without any thought or hesitation.

> Top drawer: sources for my published papers, *e.g.*, drafts and figures
>
> Second drawer: papers filed by subject (*e.g.*, biology) instead of author— papers by author are in the cabinet out in the hall
>
> Third drawer: copies of papers to give away, both my own and friends'

So when I assembled these new files, even though the third drawer had plenty of room for the new files I recruited a new file drawer for the new category—personal papers. I threw out most of the junk that had been living in the fourth drawer, shuffled the rest elsewhere, and inserted the new files in that drawer.

What happens

That's the history, here's the phenomenon. I often have occasion now to find one of these personal files, either to fetch something or to stash something that arrived in the mail. When I do so, however, I very often find myself opening the third drawer instead of the fourth.

Why is this? My guess is as follows. I'll bet that I often thought of the third/papers-to-give-away drawer as the "bottom drawer." When Daniel and I shared the file cabinet, the top drawers were mine and the bottom ones were his. I believe he used the bottom drawers for his personal files because they were easy to reach when he was sitting in his chair at his desk. I took the top couple drawers, leaving a sort of no-man's-land in the middle, which was just as well since it saved us the bother of explicitly apportioning the cabinet. When I recruited the third drawer—I believe it was after Daniel left—it was "the bottom drawer," meaning something like "the bottom one of my drawers." When I went to get a paper to hand out, I went to "the bottom drawer" or "the bottom occupied drawer" or something like that. The drawers with Time Magazines and stuff in them were just a junk heap; the stuff in them was there because the space happened to be vacant, not because the space was specifically assigned to them. These bottom two junk-storage drawers were definitely a grey area in my experience of the room; since I had never deliberately explored them or arranged them or assigned them a purpose, psychically I did not 'own' them. Once Daniel owned them; now they were orphans.

So now I recruited one of these outside drawers. The fifth drawer—the actual physically bottom drawer of the cabinet—is still out in the forest, but now I've cleared the fourth drawer and annexed it to my personal space. So when I go for this drawer to retrieve some personal papers, it's "the bottom drawer." The part of me that relates the quasi-linguistic string—actually I tend to think of it as a routinized internal utterance—"the bottom drawer" to the physical file cabinet, I suspect, hasn't heard the news. It's in the habit of turning "open the bottom drawer" into the actions that open the third drawer.

Why do I think that it's "the bottom drawer" rather than some other description? To say that it's "the personal files drawer" begs the question since that's not especially operational; most likely I know that "the personal files are in the bottom drawer" or something like that. Most likely it's not "the fourth drawer," since I can't think of any reason to believe that this would lead to any action except the correct one.

One slightly different possibility does make sense. Perhaps it's "the next drawer down," that is, the description I used back when I recruited the third drawer. If the part of me that fetches files still thinks of the file drawer as the top drawers plus an extra drawer that I've annexed, then I'm not going to have any ready way to distinguish the third and fourth drawers, both of which have been "the next drawer down" at different times.

What I'm trying to figure out here

Implicit in both this story and the Boston-Globe-arts-section story are some ideas I'm trying to work out about representation and action.

> 1) Writing model of representation. I tried to define this in "Writing as good and bad metaphor for representation." I mean the writing metaphor in a particular sense, the less prevalent, "good" sense, which asks you to think of a representation as something that's not part of you. It is, as we say, "a resource in situated action." A representation doesn't have a once-and-for-all meaning. Instead, you have to make sense of it in each next situation. There are no complete or systematic or guaranteed rules for this making-sense. And when you *do* manage to figure out what in some situation some representation is talking about, there is no way to finish listing what about the situation enabled you to do this.

> 2) Dependency model of routine evolution. All forms of activity are snap-shots in the evolution of routines. The routines themselves are complicatedly intertwined around the patterns of society and the physical layout of partic-ular places. The model is, when you think a new thought (whatever that means) in some situation, you construct a dependency circuit connecting the thought's premises to its conclusion. When you believe those same premises again (whatever that means) in some future situation, the dependency circuit reasserts the conclusion.

I'm trying to connect these two ideas. I want to portray most concrete activity as involving routinized representation-using—where I mean the writing model of 'using' representations. The writing model talks about "making sense of a representation in each next situation by figuring out what in the situation it is talking about." So, to take an example of Chapman's, "turn left" might engender a variety of different actions in different situations.

Now this each-next-situation business sounds like a lot of work. If it really required a fresh, novel effort to make sense of "turn left" or "open the bottom drawer" on every moment of every day then we'd never be able to do anything. But I don't think it's that bad. What you have in practice is a patchwork of routinized methods that have worked in various situations in the past. The methods themselves have accumulated by dependency maintenance. In some past situation someone said "take the next left" and you took certain specific concrete actions did figure out what this meant you should

do: you looked around, you performed visual operations, maybe you walked around and looked at various things, and so forth. You had your own various reasons for doing all these things, and all the actions and all their reasons got recorded by the dependency network, so now it's ready and waiting to happen again in new situations. In subsequent situations some or all of the reasons might not have applied, so instead you took various other actions, and these themselves led to new dependency circuitry.

Perhaps after enough of this you assemble enough routines to apply "turn left" to almost all of the left-turn situations you encounter in the average day. You've developed habits of interpretation. So whenever you tell yourself—or someone else tells you—"turn left" you'll be able to do it right away, 'automatically,' without any hesitation or difficult figuring-out.

These routines, like all routines, will evolve. Very often the evolution of a routine involving a representation will permit you to undertake the activity without the representation actually being present. For example, using a recipe ten times will often let you make the dish without referring to the recipe. One precondition for this effect seems to be that you understand the reasons for the recipe's instructions, but this is a vague idea that needs to be twisted around uncomfortably to fit the actual phenomena. In any event, there's an important sense in which the representation itself never goes away. Even when you're routinely deciding to turn left or add salt, you're still—in some sense I wish I understood—saying the utterance "turn left" or "salt to taste" to yourself, and you're still interpreting it just like any other natural-language utterance you need to relate to a concrete situation.

Now, where do effects like the arts-section page-turning mistake and the bottom-file-drawer mistake come from? Habits of interpretation are just that, habits. When you store a dependency record, you store a list of reasons why the action is a good one. But this list cannot be complete. The success of your actions always depends on aspects of the current situation that you've never articulated. If these aspects change and nothing brings the change—or more accurately, the relevance of the change to your habitual line of reasoning—to your attention, your routine might lead you into a mistake. This, I suspect, is what happened in both cases. The Boston Globe changed its arts section's format so that two different pages might count as "the front page" and my file cabinet changed so that a different drawer counted as "the bottom drawer" (or "the bottom one of my drawers" or whatever it was).

References

Philip E. Agre and David Chapman, Pengi: An implementation of a theory of activity, Proceedings of AAAI-87.

Philip E. Agre and David Chapman, What are plans for?, Submitted to the *AI Magazine*.

Jon Barwise and John Perry, Situations and attitudes, MIT Press, 1983.

Lawrence Birnbaum, Integrated processing in planning and understanding, Yale Computer Science Dept PhD thesis, 1986. See especially Chapter 9, "Goals as active mental agents."

Ronald J. Brachman and Hector J. Levesque, The tractability of subsumption in frame-based description languages, Proceedings of AAAI-84, pp 34-37.

David Chapman, Connections, encodings, and descriptions, unpublished manuscript, 1987.

David Chapman, Planning for conjunctive goals, Artificial Intelligence **32**, 333-377, 1987.

David Chapman, A system illustrating situated instruction use, unpublished manuscript, 1988.

Michael Cole and Sylvia Scribner, *Culture and Thought*, John Wiley and Sons, 1974.

Jonathan Culler, *On Deconstruction*, Cornell University Press, 1982.

Daniel Dennett, *Brainstorms*, Bradford Books, 1978.

Jeff Elman, Finding structure in time, CRL Technical Report 8801, Center for Research in Language, University of California at San Diego, 1988.

Jerry Fodor and Zenon Pylyshyn, Connectionism and cognitive architecture: A critical analysis, *Cognition*, January 1988.

Harold Garfinkel, *Studies in ethnomethodology*, Polity Press, Oxford, 1984. Originally published in 1967.

Jack Goody, *The domestication of the savage mind*, Cambridge University Press, 1977.

Roy Harris, *The Language Machine*, Cornell University Press, 1987.

Shirley Brice Heath, What no bedtime story means: Narrative skills at home and school, Language in Society **11**, pp 49-76, 1982. Reprinted in Bambi B. Schieffelin and Elinor

Ochs, eds, *Language Socialization Across Cultures*, Cambridge, 1986.

John Heritage, *Garfinkel and ethnomethodology*, Polity Press, 1984.

John E. Hopcroft and Dean B. Krafft, The challenge of robotics for computer science, in Jacob T. Schwartz and Chee-Keng Yap, eds, *Algorithmic and Geometric Aspects of Robotics*, Volume 1, Lawrence Erlbaum Associates, 1987.

Edwin Hutchins, Mediation and automatization, ICS Report 8704, Institute for Cognitive Science, University of California at San Diego, 1987. Also in the *The Quarterly Newsletter of the Laboratory of Comparative Human Cognition* 8, 47-58, 1986.

Bruno Latour, *Science in action*, Harvard University Press, 1987.

Stephen C. Levinson, *Pragmatics*, Cambridge University Press, 1983.

Marvin Minsky, *The society of mind*, Simon and Schuster, 1985.

Allen Newell and Herbert Simon, *Human problem solving*, Prentice-Hall, 1972.

Christopher Norris, *Deconstruction: Theory and practice*, Methuen, 1982.

Walter Ong, *Orality and literacy*, Methuen, 1986.

David E. Rumelhart, Paul Smolensky, James L. McLelland, and Geoffrey E. Hinton, Schemata and sequential thought processes in PDP models, in James L. McLelland and David E. Rumelhart, eds, *Parallel Sequential Processing: Exploration in the Microstructure of Cognition*, volume 2, *Psychological and Biological Models*, MIT Press, 1986, pp 7-57.

Jeff Shrager and D. Klahr, Instructionless Learning about a Complex Device, *International Journal of Man-Machine Studies* 25, 153-189, 1986.

James W. Wertsch, *Vygotsky and the social formation of mind*, Harvard University Press, 1985.