Notes on Chapter 1, “Prologue”

Minsky wants to explain how minds work -- how intelligence can emerge from nonintelligence. His perspective is "the Society of Mind": the mind is comprised of agents, which are processes that do very simple (thoughtless) things. "Yet when we join these agents in societies -- in certain very special ways -- this leads to true intelligence."

Minsky's aim is to perform a functional analysis of the mind, sketching out its basic structure first, and filling in the specifics later (possibly modifying the structure too). The structure of the book is geared to reflect (and perhaps to emphasize) this general theme.

1.1 The Agents of the Mind

"To explain the mind, we have to show how minds are built from mindless stuff, from parts that are much smaller and simpler than anything we'd consider smart." Otherwise, we wind up stuck in Ryle's Regress, or the homunculus problem.

Minsky provides a list of 14 different questions, all of which are of great interest to psychologists and cognitive scientists, and all of which appear to be extremely difficult to answer. Minsky's point, though, is that this difficulty is due in part to viewing these questions in isolation. When we see all of these questions as being interrelated -- which presumably occurs naturally from the Society of Mind perspective -- answers to one question will illuminate the others.

One thing to note here: the underlying theme is similar to Simon's (1969) book *The sciences of the artificial* and to Braitenberg's (1983??) book *Vehicles*. Both of these books stress the point that simple systems, when placed in a complex environment (which might be complex because of the presence of other simple systems), produce complex, interesting, and surprising phenomena.

The second thing to note is this -- at the end of the book, we should come back to this set of questions, and see the extent to which the Society of Mind perspective has illuminated their answers!

1.2 The Mind And The Brain

"How could solid-seeming brains support such ghostly things as thoughts?" Mind-body problem is presented here as being fundamental, and mysterious.

However, other problems once seemed intractable, but are no longer so. For example, science has explained away the mystery of "what is life", without appealing to magic principles of animism.

(NB: Or has it? I'm reminded of the point made in Crichton's (1969, p. 201) novel *The Andromeda Strain*, where a group of scientists decide that energy conversion is the hallmark of life. Leavitt, a central character in the novel, wants to argue against this point. He brings in a swatch of black cloth, a watch, and a piece of granite. `Gentlemen, I give you three living things.' He then proceeds to show how their notion of "life" doesn't rule out this claim. As a result, his peers had to change their view. Is the notion of Life more mysterious than Minsky would have us believe?)

Minsky argues that Thought was extremely mysterious a century ago, but the history of ideas that have culminated in the modern field of artificial intelligence "inspired a flood of new ideas about how machines could do what only minds had done previously." Still, the mystery has not been completely removed, because we are at a point where better theories are required. "This book will show how the tiny machines that we'll call `agents of the mind' could be the long sought `particles' that those theories need."

1.3 The Society Of Mind

Even acts that seem simple, like "pick up a cup of tea", can be described as cooperation (and competition) among a large number of different agents. For example, Minsky posits that this simple act involves (at least) GRASPING, BALANCING, THIRST, and MOVING agents.

"Yet none of these consume your mind as you roam about the room talking to your friends...Why not? Because they can depend upon one another. If each does its own little job, the really big job will get done by all of them together: drinking tea."

Huge numbers of processes must be involved in such simple acts, but none appear to require much thought. Even talk itself doesn't require much conscious effort. But underneath there is a lot of deep complexity. "These processes actually involve more machinery than anyone can understand all at once."

So, Minsky's move is to perform a functional analysis (a la Cummins, 1983) on a very simple, ordinary activity -- making things with children's building blocks. "Our study of how to build with blocks will be like focusing a micro-

scope on the simplest objects we can find, to open up a great and unexpected universe."

(NB: We can ask ourselves whether it is really the case that, for the mind, "one can learn the most by studying what seems the least." Nevertheless, this is indeed the practice followed by most cognitivists and by most psychologists from other schools of thought.)

1.4 The World Of Blocks

Minsky starts by positing an agent, BUILDER, which takes a set of children's blocks, and builds a tower from them. While this task sounds simple, it is too complicated for one agent to do by itself.

BUILDER needs help. It must call BEGIN to choose a place to start the tower, ADD to add a new block to the tower, and END to decide whether the tower is high enough.

Each of these agents will also rely on others, too! For instance, ADD must FIND a new block, and GET that block, and PUT it on the tower top. "Before we're done, we'll need more agents than would fit in any diagram."

"Why break things into such small parts? Because minds, like towers, are made that way -- except that they're composed of processes instead of blocks. ... Though all grown-up persons know how to do such things, *no one understands how we learn to do them!* And *that* is what will concern us here."

(Two themes begin to emerge quite markedly in this essay. The first is the notion of decomposing complex processes into (organized) subsets of simpler processes. The second is the notion that this organization is hierarchical, and is an organization imposed on the control of the system. Both of these ideas have a long and distinguished history in cognitive science, and can be found in Miller, Galantner & Pribram's book *Plans And The Structure Of Behavior*, not to mention Simon's classic *The Sciences Of The Artificial*.)

1.5 Common Sense

SEE will be a very complicated agent, as will MOVE. In addition, the system will have to keep track, too, so not to SEE/MOVE/GRASP blocks already on the tower that are being constructed. "When we look closely at these requirements, we find a bewildering world of complicated questions."

I.E., underlying "simple" acts are astonishingly complex information processing problems. (NB: This was one of the lessons provided by the natural computation approach to vision.)

"Common sense is not a simple thing. Instead, it is an immense society of hard-earned practical ideas -- of multitudes of life-learned rules and exceptions, dispositions and tendencies, balances and checks."

So what makes "common sense" so obvious and natural? As skills mature, we (iteratively) build layers of more complex skills on them. The mature skills therefore become more remote. This brings to my mind the transition from a controlled process to an automatic process, as described by Shiffrin and Schneider.

1.6 Agents And Agencies

None of the agents in an explanation can be intelligent. "Accordingly, whenever we find that an agent has to do anything complicated, we'll replace it with a subsociety of agents that do simpler things."

Critical point related to this is that organization -- the structured relationship among agents -- is critical. (NB: As in syntax -- "The butcher is a surgeon" vs. "The surgeon is a butcher".) "It is not enough to explain only what each separate agent does. We must also understand how those parts are interrelated -- that is, how *groups* of agents can accomplish things.

Bottom line: Agencies seem intelligent, but the agents from which they are constructed are not. "If you were to watch Builder work, from the outside, with no idea of how it works inside, you'd have the impression that it knows how to build towers. But if you could see Builder from the inside, you'd surely find no knowledge there. You would see northing more than a few switches, arranged in various ways to turn each other on and off." (NB: This is very reminiscent of the vehicles that Braitenberg describes in his book!).

Notes on Chapter 2, “Wholes And Parts”

2.1 Components and Connections

Agents must be linked to one another by a suitable network of interconnections. The control structure is critical. This is exactly analogous to a computer program. This has implications for how we are to understand large and complex things. "First, we must know how each separate part works. Second, we must know how each part interacts with those to which it is connected. And third, we have to understand how all these

local interactions combine to accomplish what that system *does* -- as seen from the outside."

Minsky sees a link between agents, agencies and societies and nerve cells. Where will he go with this??

2.2 Novelists and Reductionists

For Minsky, reductionists build on old ideas, and novelists create new ones. "Reductionists are usually right -- at least at science's cautious core, where novelties rarely survive for long."

From looking at chemistry or physics, "it really is amazing how certain sciences depend upon so few kinds of explanations." Should psychology be like chemistry or physics? Minsky argues no!! "The `laws of thought' depend not only upon the properties of those brain cells, but also on how they are connected. And these connections are established not by the basic, `general' laws of physics, but by the particular arrangements of the millions of bits of information in our inherited genes. To be sure, `general' laws apply to everything. But, for that very reason, they can rarely explain anything in particular.

But this isn't a call to reject the laws of physics. It is a challenge to find additional laws. This is exactly the view that Pylyshyn proposes in his book, when ZWP argued that cognitive laws capture generalizations that cannot be captured by physical descriptions.

2.3 Parts and Wholes

Terms like "Gestalt" mask our ignorance. "We say `gestalt' when things combine to act in ways we can't explain.

Minsky provides a set of subjective questions, and a set of objective questions. (NB: They are supposed to contrast nicely with one another, but I'm a little too thick to capture the subtleties that M. is trying to communicate here.) Minsky argues that the objective questions are less mysterious. "Many people assume that those `subjective' kinds of questions are impossible to answer because they involve our minds. But that doesn't mean they can't be answered. It only means that we must first know more about our minds."

Minsky argues that if we hide behind pseudo explanations like "holistic", we will never understand aspects of psychology that are of interest to humanists. Instead, for M. we have to find out lots more about mental agents. Question: Will Minsky describe what agents *really* are, or will he just name them -- as he has only been doing early in the book!

2.4 Holes and Parts

"Are life and mind so much more than the `sum of their parts' that it is useless to search for them?" How does a box contain a mouse, when none of its boards show "containment"? Containment is a property that emerges from properly arranging the boards!

"Like *boxing-in*, words like *living* and *thinking* are useful for describing phenomena that result form certain combinations of relationships. But *mind* still holds its mystery -- because we still know so little about how mental agents interact to accomplish all the things they do."

Again, the theme is that mysteries surrounding though (remember the list of questions in Chapter 1.1) will lose their mystique, because they will be seen as emergent properties of agencies (or societies of agencies!). This brings to mind Searle' nice discussion of emergence, when he talks about wetness and water molecules.

2.5 Easy Things Are Hard

Early attempts to program BUILDER revealed a wealth of underlying complexity. A huge number of programs had to be included in the simulation. "In attempting to make our robot work, we found that many everyday problems were much more complicated than the sorts of problems, puzzles, and games adults consider hard. At every point, in that world of blocks, when we were forced to look more carefully than usual, we found an unexpected universe of complications." For example, how do you index a used block, so that you won't attempt to use it again? "Thousands and, perhaps, millions of little processes must be involved in how we anticipate, imagine, plan, predict and prevent -- and yet all this proceeds so automatically that we regard it as `ordinary common sense.'

"In general, we're least aware of what our minds do best." (NB: In my view, this is the key insight provided by AI research. Minsky's basic point is this: we can't trust our judgment about what is simple!)

2.6 Are People Machines?

Many people are offended by the computer metaphor. People are uncomfortable viewing themselves as machines.

Minsky's take on this situation is to view the word machine as being out of date. It carries

with it connotations of trivial or simple mechanisms. "The term `machine' no longer takes us far enough." Modern machines are far more complex than those upon which the connotations of "machine" have been built.

Minsky wants to look at the brain, to see a *really* complex machine, and to get some self-respect in being likened to a machine. (NB: This can be related to the notion of Churchland and others that the brain represents the most complex device in the universe.)

Notes on Chapter 3, "Conflict and Compromise"

3.1 Conflict

At some point, one agent's goal will conflict with another. (e.g., WRECKER vs. BUILDER -- which might have been called by WRECKER earlier!) (NB: This is a basic issue with production system architectures, too.)

Minsky assumes "that conflicts between agents tend to migrate upward to higher levels. For example, any prolonged conflict between Builder and Wrecker will tend to weaken their mutual superior, Play-with-Blocks."

(NB: My sense here is that agents are "active", where activity = the ability to seize control. Conflict weakens this activity, and as a result weakens the ability of the superior agent to repress its competitors. Hmmm..this makes me wonder whether such competitive processes could be modeled in a hierarchical winner-take-all network, because obviously the superior agent is competing with other superiors, just as its subordinate agents are competing against each other.)

3.2 Noncompromise

Agents will get into conflicts when they compete for resources. This results in what Minsky calls the principle of noncompromise:

"**The Principle of Noncompromise:** *The longer an internal conflict persists among an agent's subordinates, the weaker becomes that agent's status among its own competitors. If such internal problems aren't settled soon, other agents will take control and the agents formerly involved will be `dismissed'."* For example, "any conflict inside Play will weaken it and make it easier for Eat or Sleep to take over. Of course, Eat or Sleep must conquer in the end, since the longer they wait, the stronger they get."

This raises the question, in my mind, of how this principle might relate to theories of motivation.

3.3 Hierarchies

Important job of high-level agents is control. "When any enterprise becomes too complex and large for one person to do, we construct organizations in which certain agents are concerned, not with the final result, but only with what some other agents do." As a result of this, there is a set of fundamental control problems to be solved -- who chooses agents to do jobs? Who decides when jobs are done? Who decides how much effort different jobs should take? Who settles conflicts, and how?

A control hierarchy would solve many of these problems. But, as is often the case in complex businesses, the relations between mental agents are not strictly hierarchical. As a result, solving control problems is not only crucial, but difficult.

3.4 Heterarchies

Hierarchies do not always work. "When two agents need to use each other's skills, then neither one can be `on top'." For example, SEE and MOVE are two agents that need each other -- "MOVE to SEE, SEE to MOVE".

"Most of the diagrams in the early parts of this book depict simple hierarchies. Later, we'll see more cross-connected rings and loops -- when we are forced to consider the need for memory, which will become a constant subject of concern in this book." With this passage, a new theme for the text is introduced -- the need for memory, and how it affects control structures.

"If we have enough memory, we can arrange our agents into circular loops and thus use the same agents over and over again to do parts of several different jobs at the same time." This introduces a second critical theme, the recursive nature of mental operations. Minsky's description reminds me a lot about how recursion is implemented in LOGO and LISP.

3.5 Destructiveness

What happens when control is captured by other agencies? The release of control can have many, perhaps unpredictable, consequences, as some agents act on their own when released from an agency. When an agency releases control, this does not mean that all of its controlled agents stop!

"Destructive acts can serve constructive goals by leaving fewer problems to be solved. That kick may leave a mess outside, yet tidy up the child's mind."

3.6 Pain and Pleasure Simplified

"Pain simplifies your point of view. When something gives you pleasure, then, too, it's hard to think of other things." In other words, pain's survival value is in how it distracts us from other goals.

While from one perspective pleasure and pain can be viewed as opposites, we also can think of them as being similar. "Why do we find such similarities between antagonistic things? [...] In order to appear opposed, two things must serve related goals -- or otherwise engage the selfsame agencies."

(NB: Again, this makes me think of the winner-take-all view of agency competition. Are we to infer from this chapter that if two capacities are clearly in opposition to one another, then they should be viewed as agents at the same level of a society which are competing for control?)

Notes on Chapter 4, “The Self”

4.1 The Self

We don't have a good definition of self, and "it often does more harm than good to force definitions on things we don't understand."

Minsky goes on to argue that a constructive approach to such definitional issues is to try to learn something about why our old (and wrong) notions about the mind were believed. "When we do this, it shows us that we do not have one such idea [about what the Self is], but many." This is another recurring theme -- multiplicity of ideas. Does Minsky depend on this result, available from introspection, to motivate his notion of multiple agents?

Self-images are defined as beliefs about what we are capable of doing, an about what we may be disposed to do. "We exploit these beliefs whenever we solve problems or make plans. [...] Our ideas about ourselves also include ideas about what we'd like to be and ideas about what we ought to be." So Minsky distinguishes between self-images and self-ideals, the former related to short (and long) range planning, the latter related to long-range goals for self.

4.2 One Self Or Many?

Self is a useful concept, "provided that we think of it not as a centralized and all-powerful entity, but as a society of ideas that include both our images of what the mind is and our ideals about what it ought to be."

"But if there is no single, central, ruling Self inside the mind, what makes us feel so sure that one exists? What gives that myth its force and strength? A paradox: perhaps it's because there are no persons in our heads to make us do the things we want -- nor even ones to make us want to want -- that we construct the myth that we're inside ourselves."

(NB: This reminds me of Dennett's view, in Consciousness explained, that a unitary notion of consciousness (the "Cartesian Theatre") is false.)

4.3 The Soul

"People ask if machines can have souls. And I ask back whether souls can learn. [...]Why try to frame the value of a Self in such a singularly frozen form? [...] The agents, raw, that make our minds are by themselves as valueless as aimless, scattered daubs of paint. What counts is what we make of them."

The merit of the self lies entirely in its own coherency. This makes me think of the amazing complexity of a single cell, embodied in a textbook picture of a walk-through model cell at the Smithsonian. How could such order, such coherence, arise?

"What are those old and fierce beliefs in spirits, souls, and essences? They're all insinuations that we're helpless to improve ourselves."

4.4 The Conservative Self

Why do we need to exploit roundabout techniques to control ourselves? "To understand how something works, one has to know its purposes. [...] To understand what we call the Self, we must see what Selves are for. One function of the Self is to keep us from changing too rapidly."

Long-range planning is likened by Minsky to a stable self, to keeping new agents from undoing the worthwhile work of old -- just like not removing a block form the child's growing tower. This suggests that, in Minsky's view, the coherence of Self comes from striving towards the goals that are defined by self-ideals.

"Selves are practical necessities. They myths that say that Selves embody special kinds of liberty are merely masquerades. Part of their function is to hide from us the nature of our self-

ideals -- the chains we forge to keep ourselves from wrecking all the plans we make."

4.5 Exploitation

Minsky has described a convoluted strategy in which "WORK exploited ANGER to stop SLEEP. But why should WORK use such a devious trick? [...] If WORK could simply turn off SLEEP, we'd quickly wear our bodies out. If WORK could simply switch ANGER on, we'd be fighting all the time. Directness is too dangerous. We'd die."

Devious relations among agents make sense, on reflection. "Deviousness" is a result of checks and balances among agents. Fantasies replace direct links between agents that evolution has removed, but the use of these fantasies requires learning. "Human self-control is no simple skill, but an ever-growing world of expertise that reaches into everything we do."

(NB: I must be a slow learner, picking up on this so late in the game -- but another recurring theme in this book is LEARNING -- learning, at the very least, defines the evolution of an agency. And changes in agencies have been raised at several points before this!)

4.6 Self-Control

Minsky provides a long list of tricks that we could use to force ourselves to work when we are tired: willpower, activity, expression, chemistry, emotion, attachment. "So many schemes for self-control! How do we choose which ones to use? There isn't any easy way. Self-discipline takes years to learn; it grows inside us stage by stage."

(NB: What exactly, though, is growing? Is it the changing of control structure relationships among existing agents? Is building a new agency equivalent to this? Or is it the building of new agents? OR ARE ALL OF THESE THE SAME!--might we come with a universal set of agents, and differences among us evolve as we arrange them into our own individual agencies?)

4.7 Long-Range Plans

"We cannot simply `decide' or `choose' to accomplish an enterprise that makes a large demand for time, because it will inevitably conflict with other interests and ambitions. [...]So how can any long-range plan succeed? [...] We may need some way to make changes that won't let us change ourselves back again. I suspect that, in order to commit ourselves to our largest, most ambitious plans, we learn to exploit agencies that operate on larger spans of time."

Some agencies operate over a short time span, while others operate over a very long one. Does this imply that our character is defined by our slowest-changing agencies?

4.8 Ideals

Ideals "include the standards we maintain -- consciously or otherwise -- for how we ought to think about ordinary matters." Ideals emerge, or are required, in conflict resolution among agencies. "In childhood, our agencies acquire various types of goals. Then we grow in overlapping waves, in which our older agencies affect the making of the new. This way, the older agencies can influence how our later ones will behave." In this sense, then initial agencies are viewed as placing constraints on the form of agencies that will develop later.

"A working society must evolve mechanisms that stabilize ideals -- and many of the social principles that each of us regards as personal are really `long-term memories' in which our cultures store what they have learned across the centuries." In other words, ideals, as emerging properties of a stabilized society, can be viewed as a sort of long-term memory.

Notes On Chapter 5 "Individuality"

5.1 Circular Causality

Causal relations in the real world are rarely simple. Two goals can support one another via circular causality. (NB: Like recurrent processors in a PDP net.) "Then a loop of circular causality ensues, in which each goal gains support from the other until their combined urge becomes irresistible."

Simplifying involves removing such recurrent loops. "There are countless different types of networks that contain loops. But all networks that contain no loops are basically the same: each has the form of a simple chain. Because of this, we can apply the very same types of reasoning to everything we can represent in terms of chains of causes and effects."

But such "straightening out" requires ignoring important interactions, as well as ignoring causal links that move in the opposite direction.

5.2 Unanswerable Questions

Basic questions -- the kinds of questions children ask -- are those that we can see no way at all to answer. These questions are basic *because* they are circular.

"What stops adults from dwelling on such questions endlessly? The answer is that every culture finds special ways to deal with these questions." Institutions assert doctrines of answers -- interestingly, this provides a kind of control structure, because such doctrines (in Minsky's view) prevent us from being stuck on unanswerable questions! Dogma as control! "One might complain that such establishments substitute dogma for reason and truth. But in exchange, they spare whole populations from wasting time in fruitless reason loops."

However, Minsky notes that circular reasoning is a powerful tool if it leads (recursively?) to deeper and more powerful ideas.

5.3 The Remote-Control Self

Standard (lay) view of psychology is what Dennett would call the Cartesian theatre. But such a concept fails because of the homunculus problem. "This concept simply doesn't work. It cannot help for you to think that inside yourself lies someone else who does your work. [...] The idea of a single, central Self doesn't explain anything. This is because a thing with no parts provides nothing that we can use as pieces of explanation."

So why do we so readily accept such flawed answers? "Because so much of what our minds do is hidden from the parts of us that are involved with verbal consciousness."

The theme that is introduced here is that the lay view of Self is doomed, because it leads to circular problems. Indeed, much of Chapter 5 appears to be Minsky's attempt to drive a wedge into our confidence in this lay view, so that he has some chance of an alternative view taking hold.

5.4 Personal Identity

Minsky describes the view of Self as singleton as being paradoxical. On the one hand, it seems to serve us very well in practical life. But in psychology, to explain the mind "the single-agent image has become a grave impediment. To comprehend the human mind is surely one the hardest tasks any mind can face. The legend of the single Self can only diver us from the target of that inquiry."

(NB: Emerging theme -- our everyday introspective view of the mind is useful (and uncomplicated) because it simplifies our lives. But with respect to a scientific understanding of the mind, this single agent view is wrong and misleading and needs to be replaced with a multiple-agent view.)

5.5 Fashion and Style

We like many things that seem to have no practical value. Why? "Fashion and cognition" -- RECOGNIZABILITY, "because familiar styles make it easier for us to recognize and classify the things we see"; UNIFORMITY, "by adopting uniform styles, we protect ourselves from distractions"; PREDICTABILITY, "societies need rules that make no sense for individuals." (NB: This begins to remind me of work in social cognition on the utility of stereotypes, in order to provide some economy of cognitive processing.)

"I do not mean to say that fashion, style, and art are all the same -- only that they often share this strategy of using forms that lie beneath the surface of our thoughts." (Question: Is Minsky trying to establish negative connotations to push a non-stereotyped view of mind?)

5.6 Traits

"What permits a writer to depict such seemingly real personalities?" Agreed upon traits/stereotypes, that's what. Our simplification needs make us rely upon stereotypes, and by exploiting such conventional views of people, writers can easily portray rich personalities in rather few words.

"It's nice to be able to trust our friends, but we need to be able to trust ourselves. How can that be possible when we can't be sure what's in our own heads? One way to accomplish this is by thinking of ourselves in terms of traits -- and then proceeding to train ourselves to behave according to those self images." Theme again: this view of personality requires that complexities be ignored.

5.7 Permanent Identity

"So far as consciousness is concerned, we find it almost impossible to separate the appearances of things from what they've come to mean to us. But if we cannot recollect how things appeared to us before we learned to link new meanings to those things, what makes us think we can recollect how we ourselves appeared to us in previous times?"

Bottom line -- single agent view is engraved upon us, and provides a strong bias concerning our view of mind and self. Minsky is trying to work against this bias.

Notes On Chapter 6, "Insight and Introspection"

6.1 Consciousness

We use our minds, not knowing how they work. Conscious thoughts tell us little about what gives rise to them. Minsky portrays many conscious thoughts as signals that start complicated processing of which we are not aware: "Our conscious thoughts use signal-signs to steer the engines in our minds, controlling countless processes of which we're never much aware. Not understanding how it's done, we learn to gain our ends by sending signals to those greater machines, much as the sorcerers of older times used rituals to cast their spells."

(NB: This reminds me of a key point in a Pylyshyn (1981) Psych. Review paper -- we are aware of the content of thoughts, but not of the machinery that represents this content.)

6.2 Signals and Signs

Analogy is the source of all understanding. (NB: Shades of Robert Sternberg!!). "We make each novelty seem similar to some more ordinary thing. It really is a great discovery, the use of signals, symbols, words, and names. They let our minds transform the strange into the commonplace."

"There are no doors inside our minds, only connections among our signs. To overstate the case a bit, what we call `consciousness' consists of little more than menu lists that flash, from time to time, on mental screen displays that other systems use. It is very much like the way the players of computer games use symbols to invoke the processes inside their complicated game machines without the slightest understanding of how they work."

(NB: This is beginning to draw another analogy to my mind -- Minsky's description of an agency from the inside, looking merely like a set of switches -- he did this in Chapter 1.6. Is consciousness, then, merely a sense of the internal workings of an agency, showing the control structure, but not how individual agents work?)

6.3 Thought-Experiments

Looking and seeing seem simple -- but they are not. "It only seems simple because you're unaware of what is happening." From this, Minsky appears to me to take a very strong stance against introspection as a way of discovering things about the mind: "Thinking affects our thoughts." Later, "consciousness is connected with our most immediate memories. This means that there are limits on what consciousness can tell us about itself -- because it can't do perfect self-experiments."

(NB1: So, this is a continuation of Minsky trying to change self-view -- self-view isn't powerful enough for self-understanding, so we need an alternative (scientific) approach.)

(NB2: Hmmm...when Minsky started selling us the view of multiple selves in his essays in Chapter 4, didn't introspection play an important role in his development of the multiple-agent view?)

6.4 B-Brains

How can a mind watch itself? "Divide the brain into two parts, A and B. Connect the A-brain's inputs and outputs to the real world -- so it can sense what happens there. But don't connect the B-brain tot he outer world at all; instead, connect it so that the A-brain is the B-brain's world."

Now B can monitor and influence A. Why? to control, moderate, or adapt A to improve its performance. This arrangement is "reflective", and is partly "self-aware". (NB: It reminds me a little bit of Braitenberg's requirement in his later vehicles for monitoring overall activity in the vehicle's brain, so that thresholds could be continually adjusted. Perhaps for Braitenberg, the caudate nucleus is a kind of B-brain!)

"A B-brain could learn to play a role somewhat like that of a counselor, psychologist, or management consultant, who can assess a client's strategy without having to understand all the details of that client's profession. Without having any idea of what A's goals are, B might be able to learn to tell when A is not accomplishing them but only going around in circles or wandering, confused because certain A-agents are repeating the same things over and over again."

In this design, some patterns of connectivity are unstable (e.g., reciprocal connections between A and B), e.g. because of feedback driving the system out of control. Also in this design, many other layers of processors (C-Brains, D-Brains, etc. could be added -- e.g., to monitor the monitors.

(NB: From this, I get a sense of a multiple hierarchy, and it is difficult not to think of layers of hidden units in multilayer perceptrons. In PDP models, such additional layers are sources of computational power. Is this so for this idea of Minsky? Is the relationship between A- and B-brains nonlinear?)

6.5 Frozen Reflection

"No supervisor can know everything that all its agents do." This is why our internal mental workings are not conscious. One approach to simplifying is to focus on specific aspects ("frozen phenomena"). Can the mind do this too? yes -- using (incomplete) memory to reinstate mental states.

"We've now seen several reasons why we cannot simply watch our minds by sitting still and waiting till our vision clears. The only course left for us is to study the mind the way scientists do when something is too large or to small to see -- by building theories based on evidence." In other words, introspection only works when a well-designed (self) experiment has been performed.

6.6 Momentary Mental Time

"It takes some time for changes in one part of a mind to affect the other parts. There's always some delay." (NB: This is similar to Dennett's view in Consciousness explained, and provides more ammunition against the "Cartesian theatre" or the single-agent view.)

It is simply impossible, in general, for any agent P to know for certain what another agent Q is doing at precisely the same time. The best that P can do is send a query straight to Q and hope that Q can get a truthful message back before other agents change Q's state -- or change its message along the way." Because of this, each agency lives in a slightly (or more than slightly) different world of time.

6.7 The Causal Now

"Our everyday ideas about the progression of mental time are wrong: they leave no room for the fact that every agent has a different causal history." Different agencies have different "senses" of time, some work fast and others work slow. Time now becomes operationalized for an agent as the time between state changes. "The slower an agency operates -- that is, the longer the intervals between each change of state -- the more external signals can arrive inside those intervals."

(NB: Now we have some clues as to the technical nature of an agent. Agents have states; over time (discrete "chunks") external signals arrive as input to the agent; with each time "click" an agent can change its state; slow vs. fast agents differ in terms of how long a time "click" lasts.)

6.8 Thinking Without Thinking

We are not introspectively aware of how we think. "If we could really sense the workings of our minds [...] we wouldn't have such varied and conflicting theories for psychology."

"Many people seem absolutely certain that no computer could ever be sentient, conscious, self-willed, or in any other way `aware' of itself. But what makes everyone so sure that they themselves possess those admirable qualities?" In other words, people don't really have much insight into why they differ from machines -- the evidence for self awareness is really weak.

(NB: Again, I sense an argument against the lay view of mentality that is based on casual introspection. It doesn't reveal the way that the mind actually works (which, of course, Minsky views as a society of agents). Question: Will Minsky provide better, or at least alternative, evidence for this position?)

6.9 Heads In The Clouds

"Nothing can have meaning by itself, but only in relation to whatever other meanings we already know." What keeps this view of meaning from collapsing. First, at least it is plausible. Second, rich networks of meanings provide many routes for solving problems -- as long as the meaning system isn't "mush" because of too *many* interrelationships!

"The secret of what anything means to us depends on how we've connected it to all the other things we know. That's why it's almost always wrong to seek the `real meaning' of anything. A think with just one meaning has scarcely any meaning at all."

(NB: That Minsky has this view of meaning doesn't surprise me, seeing as it parallels his notion of an agency. Can individual agents have meaning -- in the context of their casual role in an agency?)

6.10 Worlds Out Of Mind

Each mind evolves, uniquely, a different network of meanings. "The worlds of thought that we appear to like the best are those where goals and actions seem to mesh in regions large enough to spend our lives in. [...] Minds also make up pleasant worlds of practical affairs--which work because we make them work, by putting things in order there."

The danger of simplifying meaning is that the illusion of solving problems can arise. "Perhaps no problem was actually solved at all;

instead, the mind has merely found some secondary pathway in the brain, through which one can mechanically dislodge each doubt and difference from its rightful place!"

(NB: What's the point of this???!!! Maybe, at least, a dig against self-satisfying views of how our minds work.)

6.11 In-Sight

If we could watch brain signals, would they make sense? Not really. After all, observation is mitigated by theory. "One cannot use data without having at least the beginnings of some theory or hypothesis."

Where *do* we get our ideas? From communities, including the society of agents in our head. "Brains don't manufacture thoughts in the direct ways that muscles exert forces or ovaries make estrogens; instead, to get a good idea, one must engage huge organizations of submachines that do a vast variety of jobs." Problem with getting ideas from societies of agents -- most agents can't communicate with one another.

6.12 Internal Communication

If agents can't communicate with one another, then how can people? "The answer is that we overestimate how much we actually communicate. Instead, despite those seemingly important differences, much of what we do is based on common knowledge and experience." In other words, we communicate by exploiting a common heritage of experience.

"The smaller two languages are, the harder it will be to translate between them. This is not because there are too many meanings, but because there are too few. The fewer things an agent does, the less likely that what another agent does will correspond to any of those things. And if two agents have nothing in common, no translation is conceivable."

In other words, agents that can't communicate with one another can't do so because of a lack of common experience. This is why higher-order agents responsible for conscious thoughts are blind, too -- they cannot communicate with, in any deep sense, lower-order agents.

6.13 Self-Knowledge Is Dangerous

"If we could deliberately seize control of our pleasure systems, we could reproduce the pleasure of success without the need for any actual accomplishment. And that would be the end of everything."

For protection, our minds are subject to self-constraints. For example, it is hard to determine what is happening in the mind. For another example, self-ideal agencies are very difficult to change.

6.14 Confusion

"It's mainly when our systems fail that consciousness becomes engaged. [...] A person with an injured leg may, for the first time, begin to formulate theories about how walking works."

Similarly, it is smart to realize when thoughts are confused. For example, if a B-brain could recognize this, then it could use this recognition to drive changes that would alleviate the situation.

Notes On Chapter 7, "Problems And Goals"

7.1 Intelligence

Many insist on a definition of intelligence. But Minsky doesn't want to take this bait, apart from capturing the lay notion of intelligence: "Our minds contain processes that enable us to solve problems we consider difficult. `Intelligence' is our name for whichever of those processes we don't yet understand."

(NB: In many respects, I like this approach -- we waste far too much time arguing about what intelligence is, when we could be discovering new properties of the mind. If anyone disagrees with this view, just spend a little time following any discussion of intelligence or consciousness that appears in sci.cognitive, sci.psychology, or comp.ai.philosophy!)

7.2 Uncommon Sense

Computers are often viewed as stupid. But some of the earliest programs were truly expert systems. Interestingly, common sense is far more complicated than expertise. "What people vaguely call common sense is actually more intricate than most of the technical expertise we admire." Why? Because commonsense requires a greater variety of representations:

"To be considered an `expert', one needs a large amount of knowledge of only a relatively few varieties. In contrast, an ordinary person's `common sense' involves a much larger variety of different types of knowledge -- and this requires more complicated management systems."

7.3 The Puzzle Principle

People assume that machines can't be original or creative, but it is surprisingly easy for originality to be programmed. How? By using what Minsky calls the puzzle principle: "We can program a computer to solve any problem by trial and error, without knowing how to solve it in advance, provided only that we have a way to recognize when the problem is solved." This is typically known as the random generate and test algorithm.

"This possibility makes us reexamine all our old ideas about intelligence and creativity, since it means that, in principle, at least, we can make machines solve any problems whose solutions we can recognize." But, the problem with this is that the generate and test algorithm is extremely slow and inefficient because it is not goal-directed, and thus embodies mindless change.

(NB1: For a really intriguing account of a computer program that draws novel and interesting sciences, see McCorduck, P. (1988) Artificial intelligence: An apercu. In S. Graubard (Ed.) The artificial intelligence debate. Cambridge, MA: MIT Press.)

(NB2: The role of generate and test in psychological theories of creativity is well-entrenched, primarily after Poincare published his introspections on mathematical insight.)

7.4 Problem Solving

Generate and test is too inefficient because of its trial and error nature. Minsky moves to improve on this with his Progress Principle: "Any process of exhaustive search can be greatly reduced if we possess some what to detect when `progress' has been made. Then we can trace a path toward a solution, just as a person can climb an unfamiliar hill in the dark -- by feeling around, at every step, to find the direction of steepest ascent."

But, progress may be hard to recognize for difficult problems. Solution? Decompose the problem! "The most powerful way we know for discovering how so solve a hard problem is to find a method that splits it into several simpler ones, each of which can be solved separately."

Another approach is to embody knowledge in machines, because "the most efficient way to solve a problem is to already know how to solve it. Then one can avoid search entirely." (NB: This point becomes crucial for the discussion of K-lines in Chapter 8.)

But, embodying knowledge is problematic too. "We must discover how to acquire the knowledge we need, we must learn how to represent it, and, finally, we must develop processes that can exploit our knowledge effectively. To accomplish all that, our memories must represent, in preference to vast amounts of small details, only those relationships that may help us reach our goals."

The irony of all of this is the paradox of expert systems, i.e., it is easier to program an expert system than it is to program common sense. That is why Minsky has focused on "easy" problems to this point in his book.

(NB: Key note here -- decomposition is a key strategy to solving a very tricky problem: how the mind works. What would Braitenberg say about this strategy?)

7.5 Learning And Memory

For complex problems, it is impossible to claim that we only learn what we are rewarded for. (NB: Best example of this -- language.) "Those twin ideas -- reward/success and punish/failure -- do not explain enough about how people learn to produce the new ideas that enable them to solve difficult problems that could not otherwise be solved without many lifetimes of ineffectual trial and error. The answer must lie in learning better ways to learn."

Ultimately, learning better ways to learn will require an account of memory.

(NB1: Nice problem-solving twist to standard cognitivist attack on behaviorism -- generate and test has already been slammed for its inefficiency; here, Minsky essentially equates behaviorism with generate and test.)

(NB2: The theme of this section is important to keep in mind when reflecting back on Braitenberg's book -- Braitenberg relies pretty much exclusively on "those twin ideas" that Minsky is presenting here as being inadequate for the job of explaining cognition.)

7.6 Reinforcement And Reward

What happens in our minds to make things easier if we've done them before? One approach to this question is analogous to Hebb learning: "So let's take `reward' to mean that if agent A has been involved in arousing agent B, the effect of reward is, somehow, to make it easier for A to arouse B in the future and also, perhaps, to make it harder for A to arouse other agents."

But Snarc demonstrated to Minsky that such learning would not work for complicated problems -- another attack on behaviorism (not to mention connectionism). "We cannot learn to solve hard problems by indiscriminately reinforcing agents or their connections, Why is it that among all the animals, only the great-brained relatives of man can learn to solve problems that require many steps or involve using the same agencies for different purposes? We'll seek the answer in the policies our agencies use for accomplishing goals."

(NB: Theme here, as in preceding sections, is that associationism is not enough. Bever, Fodor and Garrett (1968 -- bug me for the exact reference) present a nice proof of this, for my taste. But what Minsky makes interesting here is the hint of what can be done to fix the problem -- looking at how goals are accomplished.

7.7 Local Responsibility

Whether a behavior is rewarded or not can depend on local success (e.g., did an agent do its job) or global success (e.g., did the agent doing its job help the system). Reinforcement on the basis of local success is easy to build into a machine. "It is harder to implement a global learning scheme because this requires machinery to find out which agents are connected all the way to the original goal by unbroken chains of subgoals."

When global success is used as a criterion for reinforcement, agents may *not* learn from their experiences! (NB: I think this idea is *very* cool -- it is worth considering for a while, and thinking about the implications for learning!!)

Both approaches to reward have advantages -- maybe a system must be flexible enough to choose which to use, and when to use it.

"The global scheme requires some way to distinguish not only which agents' activities have helped to solve a problem, but also which agents helped with which subproblems." (Question: Is this a version of the credit assignment problem, or does it require a much more refined notion of goals and subgoals. Similarly, when credit assignment is solved by a technique like error backpropagation in a connectionist network, is this focusing on local success, global success, or both?)

7.8 Difference-Engines

"Goal" has many complicated meanings. "A `goal-driven' system does not seem to react directly to the stimuli or situations it encounters. Instead, it treats the things it finds as objects to exploit, avoid, or ignore, as though it were concerned with something else that doesn't yet exist. When any disturbance or obstacle diverts a goal-directed system from its course, that system seems to try to remove the interference, go around it, or turn it to some advantage."

Difference engines -- general problem solvers a la Newell, Simon, and Shaw -- can give the impression of being goal-driven. In a difference engine, agents work to reduce the difference between current and desired states of affairs. Is this approach to "goals" too simple or too complex? Minsky will flesh this out.

(NB: Minsky provides a very nice picture of a difference engine here. Compare this picture to the Predictor vehicle that Braitenberg described. Are their similarities? Was Braitenberg also concerned with goal-driven behavior when he played with Predictor?)

7.9 Intentions

"The impression of intention is only in the watcher's mind." (NB: Compare this statement to Braitenberg!) So, when does an agent *really* have a goal? Minsky identifies two characteristics -- persistence, and the internal representation of the goal. "The idea of a difference-engine embodies both elements: a representation of some outcome and a mechanism to make it persist until that outcome is achieved."

But is this really a goal? Minsky (correctly) points out -- and I suspect that Braitenberg would too -- that it doesn't matter, because it is a very practical description. "The difference-engine scheme remains the most useful conception of goal, purpose, or intention yet discovered."

(NB: This makes me think a lot about the theme developed in Braitenberg's Vehicles. Is "goal" much more complicated from the uphill road of analysis? Minsky seems to be echoing that view here.)

7.10 Genius

Why separate genius from common sense, when the latter remains so mysterious to us? "We shouldn't let our envy of distinguished masters of the arts distract us from the wonder of how each of us gets new ideas."

Genius is built from everyday components, but also includes unusually effective ways to learn. Minsky suspects that giftedness might emerge from learning powerful new ways to learn.

(NB: This makes me think a little bit about a general strategy in cognitive psychology, to make very general claims about human cognition. Individual difference work -- and identifying genius could be regarded as looking for individual differences -- does not make up very much of the cognitivist view of the mind.)

Notes on Chapter 8, "A Theory of Memory"

8.1 K-Lines: A Theory of Memory

The notion of memory-as-a-place raises many "how" questions about representation. "The next few sections explain a theory of memory that tries to answer all these questions at once by suggesting that we keep each thing we learn close to the agents that learn it in the first place."

Minsky proposes knowledge-lines or K-lines. A K-line associates itself with agents; when the K-line is activated, this re-activates these agents. "In other words, we `memorize' what we're thinking about by making a list of the agents involved in that activity." More than one K-line can be attached to an agent.

"The basic idea is simple: For each familiar kind of mental job, your K-lines can refill your mind with fragments of ideas you've used before on similar jobs. In such a moment, you become in those respects more like an earlier version of yourself." In other words, memory is viewed as the re-activation of knowledge that was useful for solving similar problems previously. This related directly to a point Minsky made in Section 7.4, that the most efficient way to solve a problem is to already know how to solve it.

(NB: This gives a nice functional account of a K-line. But what is it with respect to other accounts (e.g., implementational, or at least technological). K-lines can be activated, and can activate agents. Should a K-line thus be viewed as a special (or even *not* special) kind of agent?)

8.2 Re-Membering

"Your mind is in a new state, with agents Q aroused. Something in your mind suspects that Q is similar to P -- and activates P." So, a new situation might lead to an old k-line being activated. This leads to two sets of active agents (current and old). Maybe all will cooperate. But what if conflicts arise?

There are many ways to resolve conflicts -- e.g., maybe give priority to K-line's agents, or maybe let the old agents win, or maybe use the principle of noncompromise (which Minsky described in Section 3.2). (NB: I don't get a clear sense of how these are to be resolved, and I suspect that Minsky isn't concerned about committing to one approach or another.) "The ideal scheme would activate exactly those P's that would be most helpful in solving the present problem. But that would be too much to ask of any simple strategy."

8.3 Mental States and Dispositions

Scientists avoid "mental state talk", and prefer to stick to "scientific" ideas that come from information processing theory. "This has produced many good theories about problem solving, pattern recognition, and other important facets of psychology, but on the whole it hasn't led to useful ways to describe the workings of our dispositions, attitudes, and feelings."

Minsky argues that the K-line mechanism could still work for these more subjective states, and in fact can explain some memory paradoxes: e.g., "Why do we find it easier to recollect our attitudes and feelings than to describe what actually took place?" Minsky's answer -- K-lines can reactivate a complex state with many components, a state that can be experienced, but which (because of its complexity) is hard to describe.

But, K-lines raise their own problems, namely, comprehension or recall of very specific facts. "Once we think in terms of K-line memories, it becomes easy to imagine, at least in principle, how a person could recall a general impression of a complex previous experience -- but it becomes hard to understand how a person can so easily comprehend a specific statement like `John has more candy than Mary'."

(NB: Minsky is presenting the K-line model as an inversion of typical memory theories. The typical theories deal with specific facts well, and with general impressions poorly. K-line theory reverses this. Question: Why does Minsky want to take this inverted route?)

8.4 Partial Mental States

Making new ideas means keeping more than one idea in mind at once. Lots of ideas, if individual features of concepts are explicitly represented (NB: This form of representation reminds me of Treisman's feature maps.) Therefore, we can talk of total or partial mental states.

To do this, Minsky introduces a key simplification -- the notion that agents are binary (on or off). From this, "A `total state' of mind is a

list that specifies which agents are active and which are quiet at a certain moment. A `partial state' of mind merely specifies that certain agents are active but does not say which other agents are quiet." Partial states are therefore incomplete descriptions, and because of this we can validly talk about a person being in more than one partial mental state at the same time. Interestingly, conflicts will arise here when agents within divisions (i.e., as if within same kind of feature map) are activated simultaneously (compare "small white ball" to conflict in "round square").

This really does bring to my mind parallels with Treisman's notion of visual attention

8.5 Level-Bands

How does the mind arouse so many -- but not too many -- memories so quickly? Minsky's approach is the level-band theory. "The basic idea is simple: we learn by attaching agents to K-lines, but we don't attach them all with equal firmness. Instead, we make strong connections at a certain level of detail, but we make weaker connections at higher and lower levels."

From this perspective, K-lines don't activate all agents equally well. Furthermore, the weakly activated agents serve as default memories that are activated (to fill a slot), but which can be easily changed. "Default assumptions embody some of our most valuable kinds of commonsense knowledge: knowing what is usual or typical."

(NB1: Level-band theory makes me think of Rosch's work on conceptual levels.)

(NB2: If agents are binary, how does level band theory work? All in connections, or is agent activity stochastic (like in a Boltzman machine or Hopfield net)?)

(NB3: Schema theory -- or, given the author, more properly frames -- is alluded to here.)

8.6 Levels

All that memory can really do is recall our minds to previous states. Level-band theory can help explain how old memories can be adapted to solve current problems. e.g., how to move BUILDER for towers into BUILDER for houses. "Most of the skills embodied in BUILDER's middle level-bands will still apply. These seem to embody the sort of knowledge that is most broadly and generally useful, whereas uppermost and lowest level-bands are more likely to be based on aspects of the problem that are specific to an older goal or to the particular details of the original problem." So, the advantage of level-band theory is the ease of change of less-relevant agents.

"We started out by using level-bands for describing things -- but we ended up using them for doing things! In the next few sections we'll see that it is no accident that level-related ideas play many different roles in how we think."

(NB: This is really, really nice stuff. I wonder -- if we took a look at Rosch's work on basic level categories, could this same sort of story be maintained? Is one advantage of the basic level ease of learning new kinds of basic level categories?)

8.7 Fringes

Memories should weaken their attachments to fine details, and to global details. "Lower Band: Beyond a certain level of detail, increasingly complete memories of previous situations are increasingly difficult to match to new situations. [...] Upper Band: Memories that arouse agents at too high a level would tend to provide us with goals that are not appropriate to the present situation." Minsky calls these two bands "fringing effects".

"We can think of the lower fringe as concerned with the structures of things, and we can think of the upper fringe as involved with the functions of things. The lower levels represent `objective' details of reality; the upper levels represent our `subjective' concerns with goals and intentions." Minsky is arguing for a tight connection between structure and function -- things and goals -- where this connection is mediated by K-lines.

(NB: Two thoughts. First, I am again struck by Minsky's slant on Rosch's work, where Minsky's innovation is to apply basic level to goal-directed behavior or change. Second, the continuum between function and structure makes me think of another continuum -- from computation through algorithm to implementation.)

8.8 Societies of Memories

There are two ways to make new memories. First, you attach a new K-line to all the agents that were recently active. Second, you attach a new K-line to older K-lines that were active recently. (NB: Important point -- now we see that K-lines can be attached to other K-lines too -- perhaps it is not bad now to think of K-line as a special kind of agent.) This builds K-line memory trees.

"The kinds of mental states that this `hierarchical' type of memory produces will be

based more on stereotypes and default assumptions than on actual perceptions. Specifically, you will tend to remember only what you recognized at the time. So something is lost -- but there's a gain in exchange. These `K-line memory trees' lose certain kinds of details, but they retain more traces of the origins of our ideas."

Another really neat idea -- Minsky is going beyond association between ideas to higher-order associations between associations. What precedent is there for this idea?)

8.9 Knowledge-Trees

K-lines can form societies. How can we keep them orderly (e.g., in nice hierarchies)? Minsky's solution: level-band theory again. "When making a new K-line memory, do not connect it to all the K-lines active at the time but only to those that are active within a certain level-band."

Can this be done without specifying what "level" means for K-lines? Minsky says yes: "Something like this will happen automatically, simply because the new K-line societies will tend to inherit whatever hierarchy already existed among the original agents that become connected to those K-lines."

New problems emerge, though. "This policy of connecting new K-lines to old ones must be used in moderation. Otherwise, no new agents would ever be included in our memories. Furthermore, it should not always be required to produce simple, orderly hierarchy-trees." But the level-band idea still needs to be adopted.

(NB: There must be some representation of time in Minsky's system, so that it can identify K-lines that were "recently active".)

8.10 Levels and Classifications

"Level" is a complex term (NB: just like "goal"). Basic idea for any notion of "level", though, is order and hierarchy. "Usually, we tend to think that each of those hierarchies illustrates some kind of order that exists in the world. But frequently those orderings come from the mind and merely appear to belong to the world." In other words, no single hierarchy is correct because "it depends on what you want to use it for."

Our classifications can resemble level-schemes and hierarchies. But the hierarchies always end up getting tangled and disorderly because there are also exceptions and interactions to each classification scheme. When attempting a new task, we never like to start anew: we try to use what has worked previously. So we search around inside our minds for old ideas to use. Then, when part of any hierarchy seems to work, we drag the rest along with it." This dragging along of the rest of the hierarchy is what can mess up the nice ordering.

(NB: This makes me think of Pylyshyn's tri-level approach, and criticisms of it. Many of the criticisms come from the notion of dragging the perspective around.)

8.11 Layers of Societies

The K-lines of each agency grow into a new society. Here, Minsky introduces some terminology. S-agents are the original agents, and S-society is their society. K-lines can get added to these agents as memories are recorded, this is the K-society.

"But this will lead to a different problem of efficiency: the connections to the original S-agents will become increasingly remote and indirect. Then everything will begin to slow down -- unless the K-society continues to make at least some new connections to the original S-society. That would be easy to arrange, if the K-society grows in the form of a `layer' close to its S-society."

"If arranged this way, the layer pairs could form a curious sort of computer. As S-agents excite K-agents and vice versa, a sort of spiraling activity would ensue." This feedback leads to problems -- chaotic explosions, etc. How is this to be controlled? "By specifying which level-band should remain active and suppressing all the rest. Indeed, that is precisely the sort of coarse control that a B-brain might exercise, since it could do all this without needing to understand the fine details of what is happening inside the A-brain." So, a third controlling agency is required. Now, too, we see an A-brain being decomposed into two layers, the S-society and the K-society.

Finally, Minsky makes the claim that the creation of new layers is the process by which mental capacities develop.

Notes On Chapter 9 "Summaries"

9.1 Wanting and Liking

Again, Minsky attacks "simplifying": "What makes us want to compress so much into such inexpressive summaries as `like', `prefer', and `enjoy'? Why try to reduce such complex things to simple values or amounts of pleasurable quality?" Minsky's position is that we cannot accept such summaries at face value. "Neither the state of the world nor that of the mind is ever so

simple that it can be expressed in a single, one-dimensional judgment."

The simplest summaries are demanded by the highest levels of the mind when decisions are made. (NB: This gives a very strong sense of increased abstraction as one moves to high levels of agents). Such summaries hide lots of internal machinery used to support this decision. "The relation between wanting and liking is not simple at all, because our preferences are the end products of so many negotiations among our agencies. To accomplish any substantial goal, we must renounce the other possibilities and engage machinery to keep ourselves from succumbing to nostalgia or remorse. Then we use words like `liking' to express the operation of the mechanisms that hold us to our choice."

(NB: This is interesting to me. A common view in cognitive psychology is that we do not have direct access (e.g., via introspection) to the machinery underlying our thoughts. Here, Minsky is making a similar point, but for different reasons -- summaries that neglect details make it easier to stick to accomplishing goals!)

9.2 Gerrymandering

"In a complex human brain, a great many layers of agencies are interposed between the ones that deal with body needs and those that represent or recognize 0our intellectual accomplishments. Then what is the significance, in these more complicated systems, of those pleasant feelings of accomplishment and disagreeable sensations of defeat? They must be involved with how our higher-level agencies make summaries." (Q: Is a major goal of higher-level agents to make summaries? Or do their behaviors merely permit us to interpret them as summarizers?)

"The only way to solve hard problems is by breaking them into smaller ones and then, when those are too difficult, dividing them in turn. So hard problems always lead to branching trees of subgoals and subproblems. To decide where resources should be applied, our problem-solving agents need simple summaries of how things are going." (NB: Relate this to the practice of functional analysis!) (NB: Minsky is arguing that summaries oversimplify, hiding internal complexity -- but are still necessary. This makes me think of ideas like cognitive economy and stereotypes, and their role in cognitive theories.)

9.3 Learning from Failure

It may be more important to learn from failure than from success, in order to minimize chances of damaging procedures that work quite well. One method to minimize this damage is to use censors and suppressors. "A safer way to deal with this would be to modify M by adding special memory devices called `censors' and `suppressors'...which remember particular circumstances in which M fails and later proceed to suppress M when similar conditions recur. Such censors would not tell you what to do, only what you shouldn't do; still, they prevent your wasting time by repeating old mistakes."

The real world is not mathematically precise. So, "we can't so willfully make up the rules for objects that already exist, so our only course is to begin with imperfect guesses -- collections of rough and ready rules -- and then proceed to find out where they're wrong."

(NB: All of this brings to mind Gold's theories of learning (which indicate that powerful learning techniques do involve learning from positive and negative information) and how these theories are at odds with empirical studies of language learning. Is Minsky following a similar path here? Specifically, how much of (say) language learning is learning from failure?)

9.4 Enjoying Discomfort

There is more to motivation than immediate reward. When we succeed, a lot goes on in the mind. "These kinds of complications make it impossible to invent good definitions for ordinary words like `pleasure' and `happiness'. No small set of terms could suffice to express the many sorts of goals and wants that, in our minds, compete in different agencies and on different scales of time."

(NB: This passage brings to mind Miller, Pribram and Galatner's treatment of behaviorist motivation theories in their book *Plans and the structure of behavior*. They pointed out that there is a lot cognitive going on influencing our notion of reward, which is why (for instance) we do not develop compulsive urges to mail letters (for more details on this, read their stuff!!).

Notes On Chapter 10, "Papert's Principle"

10.1 Piaget's Experiments

Watching children might be a way to see how mind-societies grow. Minsky takes as example Piaget's "conservation of quantity" experiments. "In the next few sections we'll examine the idea of `more' and show that it conceals the workings of a large, complex Society-of-More -- which takes many years to learn."

(Q: So, now is Minsky going to show us how his approach can account for an interesting empirical phenomenon?)

10.2 Reasoning about Amounts

There are many possible accounts of why young children fail to conserve amount -- don't understand "quantity", are unduly influenced by "extent", focus on what changes, not what remains the same, lack of logic. "Every one of the explanations has some truth in it, but none reach the heart of the issue."

Minsky's key move is this: "The younger children possess the ideas they need but they don't know when to apply them! One might say that they lack adequate knowledge about their knowledge, or that they have not acquired the checks and balances required to select or override their hordes of agents with different perceptions and priorities."

(NB: On one hand, people might not be very interested in this point, saying "Oh yes, we've known for a long time that this is a problem of metacognition." However, don't dismiss it so! The really neat idea here, one that isn't well developed in psychology, is that failure to conserve reflects an inadequate control structure!)

10.3 Priorities

Minsky's approach to conservation is to propose some example agents, TALL, THIN, CONFINED. The existence of such agents in a child is justified on the basis of the kinds of judgments that children make. Importantly, Minsky defines these agents in such a way that they always conflict. So, failure to conserve reflects one approach (a rudimentary and ultimately flawed approach) to resolving such conflicts.

(NB: This reminds me of the fact that conflict resolution problems are basic fare for anyone programming a production system. Minsky's point seems to be that, for the child, new conflict resolution schemes will emerge or will be learned.)

10.4 Papert's Principle

Another approach to conflict resolution is to use the principle of noncompromise. In order for this to work, a new intermediate level of agents must appear in the Society-of-More hierarchy. "The new APPEARANCE administrator is designed to say `more' when the agent TALL is active, to say `less' when the agent THIN is active, and to say nothing at all when something appears both taller and thinner. Then the other new administrator, HISTORY, makes the decision on the basis of what CONFINED says."

This puts the emphasis on control structure, which is highlighted in Minsky's statement of "Papert's Principle: Some of the most crucial steps in mental growth are based no simply on acquiring new skills, but on acquiring new administrative ways to use what one already knows."

(NB: Control structure is a major theme of Newell and Simon's work on cognitive simulation via production systems. But this stuff here makes me think of hidden units in connectionist models -- adding the new layer of agents is similar to adding a new layer of hidden units to a network. This makes me wonder, a bit, whether hidden units can be viewed as serving a control structure function in a PDP network.)

10.5 The Society-Of-More

"More" has many different meanings, and each involves different agencies. Minsky provides alternative agencies for other kinds of conservation problems. "You might complain that even if we needed these hordes of lower-level agencies to make comparisons, this system has too many middle-level managers. But those mountains of bureaucracy are more than worth their cost. Each higher-level agent embodies a form of `higher-order' knowledge that helps us organize ourselves by telling us when and how to use the things we know. Without a many layered management, we couldn't use the knowledge in our low-level agencies; they'd all keep getting in one another's way."

Q: Does every distinction in meaning reflect a distinction in agencies?

10.6 About Piaget's Experiments

Piaget's results were once treated skeptically, because "they contradict the traditional assumption that children are much like adults, except more ignorant." But these results and their implications have survived many subtle challenges.

"What is the significance, then, of evidence that the young children do possess methods that could give correct answers -- and yet they do not use those abilities? As far as I can see, such evidence would only further support the need for explanations like those of Papert and Piaget."

(NB: Personally, I'm not so sure that Papert and Piaget are generating similar, or even complementary, explanations.)

10.7 The Concept of Concept

Are concepts learned? Or more precisely, "why do we feel we have to think of what we learn as things or concepts? Why must we `thingify' everything?" "Thingifying" isn't always bad, but Minsky feels that in general it is disastrous for a science of mind.

"Instead of assuming that our children come to crystallize a single `concept of quantity', we must try to discover how our children accumulate and classify their many methods for comparing things." Minsky's rule for learning is to always try to combine related agents first.

(NB: Two points. First, societies-of-more are learned. Second, what does it mean for agents to be related? Minsky hints that it has to do with shared brain resources, and therefore possibly similar brain locations.)

10.8 Education and Development

Why can't we speed up the acquisition of conservation? Because simulations of mature agencies require too many special rules and children are unable to manage this complexity. (NB: Again, here is the role of a maturing control structure -- to efficiently manage such complexity).

Minsky notes that conflict resolution is much more difficult than it has appeared in his book to this point, because he has been concentrating on simple examples. He hints that later in the book he will complement his notion of competition among agents with the notion of cooperative processing.

(NB: One of his key attacks on connectionism has been problems with scaling up (see, for instance, the epilogue in the 1988 edition of Perceptrons. One wonders whether conflict resolution methods that have been proposed to this point will scale up to deal with more complex situations!)

10.9 Learning a Hierarchy

What is the best way to change the brain? Build a detour around the old, but don't remove the old until you are sure that the new route works. This suggests to Minsky an approach to adding new intermediate layers to hierarchies. New agents (and connections) form, and only later do existing structural links fade away.

(NB: First, do old agent interconnections simply wither away from disuse? could some variant of Hebb learning be used to produce this kind of modification of an agency? Second, this approach to learning intrigues me. It makes me wonder what would happen to, say, cascade correlation if the original connections in the network were not fixed after new processing units were added.)

Notes on Chapter 11, "The Shape of Space"

11.1 Seeing Red

What possible kind of brain-event could be correlated to the meaning of an ordinary word? Lots of different brain processes must be involved. "But this should not be taken to mean that no machine could ever have the range of sensibilities that people have. It merely means that we aren't simple machines; indeed, we should understand that in learning to comprehend the qualities of vast machines, we are still in the dark ages."

(NB: Does this suggest that consciousness etc. emerges from sufficient complexity? NB2: The theme of this chapter appears to be the relationship between the world and the brain.)

11.2 The Shape of Space

How can the brain learn about the external world? Sense is a complicated illusion. "We never actually make any direct contact with the outside world. Instead, we work with models of the world that we build inside our brains." So, how do neural signals lead to the experience of sensations?

Key point raised by Minsky is that our sensations carry their "meaning" via their relationships. "There is little that one could say about any `single touch' -- or about what any single sense-detecting agent does. However, there is much more to be said about the relations between two or more skin touches, because the closer together two skin spots are, the more frequently they'll both be touched at the same time."

(NB: This makes me think of the need for hidden units to produce sensitivity to higher-order relationships. Minsky is really arguing that sensations are not unary properties, but at the very least must be binary properties (i.e., involve relationships between two different things).)

11.3 Nearnesses

The spatial relations among touch sensors on the skin are mirrored in the brain. Because of this, "in general, the greater the extent to which two stimuli arouse the same sensors,

the more nearly alike will be the partial mental states those stimuli produce -- and the more similar those stimuli will `seem' simply because they'll tend to lead to similar mental consequences."

"Other things being equal, the apparent similarity of two stimuli will depend on the extent to which they lead to similar activities in other agencies." (NB: This implies a very strong relationship between brain states and mental states.)

Provided that there is spatial regularity in sensors, agents can learn topographic maps. The spatial relationships begin imposing constraints on the kind of inferences (or representations) that can be drawn by higher-order agents.

(NB: First, doesn't this remind you a lot about Braitenberg's vehicles? Second, it makes me think quite a bit about the natural computation approach to vision, where certain natural constraints concerning the physical world (e.g., 2D projected topography) are mirrored in the early stages of transducers.)

11.4 Innate Geography

Minsky begins by presenting a topographic learning principle: "The nerve pathways that preserve the physical nearness relations of our skin-sensors can make it easy for inner agencies to discover corresponding nearnesses about the outer world of space." Minsky defines "space" as a society of nearness relations among places. But no one knows how the brain makes such maps. "It is a wonderful fact that, in principle, one can deduce the global geography of a space from nothing more than hints about which pairs of points lie near one another! ... If brains do something of this sort, it might illuminate a problem that has troubled some philosophers: `Why do we all agree on what the outer world of space is like?'".

(NB: This section makes me think about learning associations between processors that are simultaneously active, which in turn makes me think of Hebb learning, which in turn makes me think of NMDA receptor sites (talk about associations!). Could Hebb learning account for the development of "innate geography"? After all, I believe that there is lots of evidence for NMDA receptor sites in the retina. Do such sites exist for other sensory modalities as well?)

11.5 Sensing Similarities

"What we learn depends on how we classify." As a result, early thought depends more on wiring than on experience -- in particular, wired-in similarity relationships. The key thing to keep in mind about similarity is that it is relational. "For just as there is nothing to say about a single point, there's nothing to be said about an isolated sensory signal. When our REDNESS, TOUCH, or TOOTHACHE agents send their signals to our brains, each by itself can only say `I'm here." The rest of what such signals `mean' to us depends on how they're linked to all our other agencies."

(NB: First, relationships as far as Minsky is concerned appear to be relations among agents. Second, Minsky recounts the problems that Johnson faced with his dictionary, because relationships needed to be taken into account. We face a similar problem with our dictionary, but can deal with it (more effectively?) by taking advantage of relational links built into the HTML.)

11.6 The Centered Self

How do we learn about the world beyond the skin (i.e., the third dimension)? "Every motion of your body, head, or eye makes drastic changes to the image in your eye. How can we extract any useful information when everything changes so rapidly? ... It appears that our brains have evolved with special mechanisms that help us compensate for motions of the body, head, and eye." This makes it easier for agencies to learn from visual information. But we still don't understand how this is done:

"Perhaps we start by doing many small experiments that lead to our first, crude maps of the skin. Next we might start to correlate these with the motions of our eyes and limbs; two different actions that lead to similar sensations are likely to have passed through the same locations in space. A critical step would be developing some agents that `represent' a few `places' outside the skin. Once those places are established (the first ones might be near the infant's face), one could proceed to another stage: the assembly of an agency that represents a network of relationships, trajectories, and directions between those places".

(NB: The basic idea here is that the notion of 3D space is bootstrapped. Is there any evidence for this view in the developmental literature? Also, here we get explicit notion that agents `represent'. But why does Minsky put scare quotes around this? Do agents represent, or not?)

11.7 Predestined Learning

There is not clear-cut boundary between heredity and environment, which leads to Minsky's notion of "predestined learning". "Why make the brain use a tedious learning process when the final outcome seems so clear? Why not build

in the answer genetically? One reason could be that learning is more economical." Why not make learning about space predestined -- this will give economy of learning.

Constraints on learning make it predestined. "We acquire our conceptions of space by using agencies that learn in accord with processes determined by inheritance. These agencies proceed to learn from experience -- but the outcomes of their learning processes are virtually predestined by the spatial geometry of our body parts."

(NB: The hypothesis emerging here is that mental universals -- e.g., conservation of quantity -- may result from learning in an environment that is structure or constrained in an important way. Two things to keep in mind though -- the constraints could be in the structure of the environment being learned about, or could also be in the structure of the system that is learning! One example of this is Dawson, Kremer and Gannon's paper on finding simple cell receptive fields in connectionist networks, but only when the network is initially constrained.)

11.8 Half-Brains

Brain is build of many pairs of agencies that are cross connected. This has led to the old of idea of the brain as a "meeting of anti-principles". "On one side stands the Logical, across from Analogical. The left-side brain is Rational; the right side is Emotional. No wonder so many seized upon this pseudoscientific scheme: it gave new life to nearly every dead idea of how to cleave the mental world into two halves as nicely as a peach."

The problem is that the brain has many parts, not just two. "My own theory of what happens when the cross-connections between those brain halves are destroyed is that, in early life, we start with mostly similar agencies on either side. Later, as we grow more complex, a combination of genetic and circumstantial effects leads one of each pair to take control of both. Otherwise, we might become paralyzed by conflicts, because many agents would have to serve two masters." In other words, Minsky's theory of lateralization is deeply rooted in the context of maturing control structures.

(NB: Now would be a good time to refresh one's mind about what Braitenberg said about cross-connections!)

11.9 Dumbbell Theories

We often dichotomize the world. "Such divisions all have flaws but often give us useful ways to think." But, dichotomies can lead us into making false analogies. So, "dividing things in two is a good way to start, but one should always try to find at least a third alternative. If one cannot, one should suspect that there may not be two ideas at all, but only one, together with some form of opposite."

"Before we're drawn into dumbbell schemes, we owe it to ourselves to try to understand their strange attractiveness."

(NB: Explaining this attractiveness might be at the root of the next chapter, which focuses a great deal upon categorization, but uses as a foundation the idea of drawing structured analogies between two domains: structural and functional.)

Notes on Chapter 12, "Learning Meaning"

12.1 A Block-Arch Scenario

Minsky describes a situation where two new block structures with identical descriptions produce different affordances. "So the child must find some way to change the mental description of Block-Arch." This is akin to the Piagetian notion of accommodation.

(NB: The analogy between structure and function begins with this section. On the one hand, we have a structural description -- Block-Arch. On the other hand, there is a functional affordance that is (or may be) offered by such a structure -- Hand-Change. Finding some Block-Arch's with the affordance, and some without, drives the change in the structural description. In other words, a breakdown in the structure/function relationship is the impetus for conceptual change.)

12.2 Learning Meaning

Learning is a hard word to define, because most definitions are too broad. "The problem is that we use the single word `learning' to cover too diverse a society of ideas."

In the preceding Block-Arch example, there were at least 4 different ways to learn: Uniframing (combining several descriptions into one), Accumulating (collecting incompatible descriptions), Reformulating (modifying a description's character), and Trans-framing (bridging between structures and functions). (NB: These are all in the cast of this chapter!)

Minsky want to introduce a new vocabulary concerning learning, because with respect to

psychology, ideas from AI are new and deserve new names. "Our Block-Arch scenario is based on a computer program developed by Patrick Winston in 1970. Winston's program required an external teacher to provide the examples and to say which of them were arches and which were not. In my unprogrammed version of this, the teacher has been replaced by the concern of some agency inside the child to account for the emergence of that mysterious Hand-Change phenomenon." In other words, motivation to learn is important.

(NB: Again, here, mysteries mount -- instead of talking about societies of agents, we are talking about societies of ideas. Is this the terminology that is used when agents are viewed as representing mental content? Or are societies of ideas qualitatively distinct from societies of agents? And on a different not, Minsky here is making a strong move towards unsupervised learning.)

12.3 Uniframes

A uniframe is defined as a description to apply to several different things at once, and Minsky is concerned with how uniframes are created. He points out that facts can be dissected, enforced, prevented, tolerated in the process of building a uniframe. But what facts should such operations be applied to? Ideally, only the essential ones. "But how can we judge which facts are useful? On what basis can we decide which features are essential and which are merely accidents? Such question can't be answered as they stand. They make no sense apart from how we want to use their answers. There is no single secret, magic trick to learning; we simply have to learn a large society of different ways to learn!"

(NB: In other words, identifying the essential facts is largely a matter of active discovery. Along different lines, what exactly is a uniframe? Is it an agent? Is this a new primitive for Minsky? Again, I'm unclear as to the representational nature of agents -- and a uniframe is obviously a representational medium!)

12.4 Structure and Function

Learning is viewed as connecting to preexisting meanings (and reminds me a bit of the Piagetian notion of assimilation). "To learn new words or ideas, one must make connections to other structures in the mind."

"But one can't learn what something means merely by tying things to names. Each word-idea must also be invested with some causes, actions, purposes, and explanations." The basic move here is for the use of analogy to learn and enrich ideas. Furthermore, the analogy is being drawn between structural descriptions (physical properties) and functional descriptions (affordances).

(NB: To my mind, this isn't a particularly novel idea as presented here. Will an agency approach give a more interesting spin on the role of analogy/metaphor in thinking? Are such operations more natural for agencies than for other proposals about the nature of mind?)

12.5 The Functions of Structures

"Many things that we regard as physical are actually psychological." In other words, many categories are defined functionally, not physically. (NB: This was a basic and early move by Miller and Johnson-Laird in their book on language and perception.)

Because of this, "we need to combine at least two different kinds of descriptions. On one side, we need structural descriptions for recognizing chairs when we see them. On the other side we need functional descriptions in order to know what we can do with chairs." But this is not simple association. Instead, there are presumed to be highly specific links between structural and functional descriptions. "With that knowledge we can do amazing things, like applying the concept of a chair to see how we could sit on a box, even though it has no legs or back!" This is why uniframes need to incorporate this kind of knowledge.

(NB: The purpose of detailed structure/function mappings is to provide generalization of knowledge. It is still not clear how these mappings are instantiated -- e.g. by agents.)

12.6 Accumulation

Uniframing doesn't always work. Result: collections of examples. Why does this difficulty emerge? Because of the many-to-one relationship between structures and function. Why is this a problem? Cognitive economy -- examples are hard to reason with. "When we try to reason about things, accumulations can be nuisances -- because then we'll be forced to find a different argument or explanation to justify each separate example."

"A simpler theory of when we start new uniframes would be that in the brain, there is an architectural constraint on how many K-lines are directly accessible to various types of agents." (NB: Does this then mean that uniframes are K-lines between agents?)

(NB: Note that accumulation is the second learning strategy, after uniframing, so a second member of the cast has been introduced. Q: Given the many-to-one problem, are uniframes even possible?)

12.7 Accumulation Strategies

"Most people find some reasonable compromise [between learning styles], though a few of us lean more in one direction than the other. I'm sure we all use mixtures of different learning strategies -- accumulations of descriptions, K-lines, uniframes, or whatever." Unwieldy collections motivate uniframing.

(NB: Emerging theme (which I am again slow to appreciate): single concepts are useless, because no single concept can meet all our requirements of it. Multiple (competing?) descriptions will be the order of the day. This view parallels the agents vs. unitary self theme developed earlier in the book, and provides an indication that agents are doing the representing. But confusion arises -- in this section, uniframes and K-lines are given separate mention. Does this mean that uniframes are not K-lines, or that K-lines are a part of uniframes?)

12.8 Problems of Disunity

Your purposes determine when you should accumulate and when you should uniframe. Grouping by similarity can be done on the basis of structure or perceived function. "At one moment you may wish to emphasize a similarity; at the next moment, you may want to emphasize a distinction. Often, we have to use both uniframes and accumulations in combination."

Why is it so hard to define the essence of an accumulation? "Many good ideas are really two ideas in one -- which form a bridge between two realms of thought or different points of view." We find accumulations on the structure side of the structure/function bridge because of many-to-one relationships. "Our different worlds of ends and means don't usually match up very well. So when we find a useful, compact uniframe in one such world, it often corresponds to an accumulation in our other worlds."

The bottom line: "we know only a very few -- and, therefore, very precious -- schemes whose unifying powers cross many realms."

(NB: This and preceding chapters remind me to re-visit prototype vs. exemplar debate in the concept literature -- a good source is Smith and Medin's book.)

12.9 The Exception Principle

An important principle about rule modification is the Exception Principle: It rarely pays to tamper with a rule that nearly always works. It's better just to complement it with an accumulation of specific exceptions." (NB: What does Minsky mean by rule?) Exceptions are always a problem when attempting to define (or create) concepts. But "concept perfection" by means of using more and more accumulation will not work. "The power of ordinary words like `fish' comes from how we make them span so many meaning-worlds at once. However, in order to do this, we have to be able to tolerate many exceptions." Important warning comes with this -- don't confuse conceptual inventions for natural phenomena!

12.10 How Towers Work

It is bad psychology to assume that "obvious" things are simple. "Many such tings are done for us by huge, silent systems in our mind, built over long forgotten years of childhood."

Key theme: "obvious" is the result of long development of giant cognitive machinery. This point was established in early chapters too.

12.11 How Causes Work

Cause is a mental invention, but not completely. "Causes are indeed made up by minds -- but only work in certain parts of certain worlds." What is a cause? First and foremost, a compact description. Such compactness is due to the nature of the world; effects are localized. "Why does a block retain its size and shape when it is moved? It is because we're fortunate enough to live within a universe in which effects are localized. ... This can happen only in a universe whose force laws work in close accord with the `nearness' of time and space -- in other words, a universe in which entities that are far apart have much less effect on each other than ones that are close together. In worlds without constraints like that, there could be no things or causes for us to know."

(NB: This is bringing to mine Rosch's view of categories reflecting the structure of the world -- the idea that mental concepts (e.g., the key components -- the types -- of Jackendoff's conceptual structure) are shaped by natural constraints.)

12.12 Meaning and Definition

Meanings aren't definitions, though definitions do appear to work better when meanings

are treated not as things, but as processes in the mind. "Of course, it is no great surprise to find that `game' has a more psychological character than does `brick', which we can define in physical terms without referring to our goals. But most ideas lie in between. We saw this in the case of `chair', which we cannot describe without referring both to a physical structure and to a psychological function."

(NB: I don't find much of this surprising, as Minsky even borrows Wittegnstein's most famous example. But is there something new in this notion of requiring concepts to relate both physical and psychological attributes?)

12.13 Bridge-Definitions

A single definition rarely works, because structural descriptions by themselves are too specific, and functional descriptions by themselves are too general. "But we can often capture an idea by squeezing in from several sides at once, to get exactly what we need by using two or more different kinds of descriptions at the same time. Our best ideas are often those that bridge between two different worlds!" Minsky is particularly keen on bridging structure and function; such bridges must be actively built during learning. "It helps to be given a good definition, but still you must mold and shape each new idea to suit your own exiting skills -- hoping to make it work for you the way it seems to work for those from whom you learn."

(NB: Minsky is quite explicit here, definitions being described as "networks of connections and constraints among our agencies." Agencies are playing a representational role, and connections among agents seem to be the source of uniframes. Furthermore, the bridging alluded to here strikes me much like the Piagetian notion of assimilation.)

Notes on Chapter 13, "Seeing and Believing"

13.1 Reformulation

How can uniframes classify disparate-seeming things, such as many different kinds of arches? Minsky says that this requires reformulating the description of the members of the category, which appears to require making a move to a more abstract (and therefore less structural?) description. One part of this involves dividing objects into essential and auxiliary portions. The other is to abandon old classification methods that no longer work. This is reformulation. "Reformulation is clearly very powerful -- but how does one do it?"

(NB: Free associating, reformulation brings to my mind the notion of a paradigm shift. Will Minsky now proceed to describe how reformulation works?)

13.2 Boundaries

Some of the secret of creativity lies in finding new ways to look at things. Minsky suggests that this can involve imagining boundaries that do not really exist, and treating real boundaries as though they do not exist. "We're always changing boundaries! Where does an elbow start or end? When does a youth become an adult? Where does an ocean change into a sea? Why must our minds keep drawing lines to structure our reality? The answer is that unless we made those mind-constructed boundaries, we'd never see any `thing' at all! This is because we rarely see anything twice as exactly the same."

"Making sense" requires finding permanence, which in turn requires us to abandon much of what we see.

(NB: Two points: First, much of this reminds me of Gordon's Synectics program for creative thinking, which was very "metaphoric" in nature, and asked practitioners to make the unfamiliar familiar, and the familiar unfamiliar. Second, the importance of imaginary boundaries, and unimportance of real boundaries, can't help but make one think of Grossberg's THEORY of everything -- even though Grossberg and Minsky are talking about very different kinds of things!)

13.3 Seeing and Believing

Minsky spends much of the rest of the chapter considering children's drawing of people, particularly drawings of people that are essentially heads with attached sticks for arms and legs.

"We normally assume that children see the same as we do and only lack our tricky muscle skills. But that doesn't explain why so many children produce this particular kind of drawing, nor why they seem so satisfied with them." Minsky suggests one view might be that the child does not have an image in mind when they draw. Instead, perhaps children have a set of features (propositions!!) in mind that serve as a description of what is to be drawn, and then employ a drawing program that is guided by this description. A simple or unsophisticated drawing program will produce unrealistic results, but can still satisfy the internal description -- which might be why children prefer these kinds of drawings.

(NB: Wow! Shades of the imagery debate here; would Kosslyn's model produce child-like images? Are internal descriptions like sets of propositions? If so, are agents to be viewed as propositions? This would make sense, if they were also viewed as being similar in nature to perceptrons...)

13.4 Children's Drawing-Frames

"It does not make much sense to speak of what a person `really' sees, because we have so many different agencies." How do children's drawings of people evolve? Minsky describes a scheme which involves relatively minor changes to the drawing program.

"I suspect that after children learn to make recognizable figures, they usually move on to face the problems of representing much more complicated scenes. As they do this, we should continue to appreciate how well children deal with the problems they set for themselves. They may not meet our own grown-up expectations, but they often solve their own versions of the problems we pose."

(NB: This illustrates a kind of reformulation (of the drawing routine), but not WHY it occurs, which is a question that keeps coming to mind for me.)

13.5 Learning a Script

For a child, the next drawing task involves depicting two or more people. "This involves wonderful problems about how to depict social interactions and relationships -- and these more ambitious projects lead the child away from being concerned with making the pictures of the individual more elaborate and realistic. But drawing learning does not stop.

Minsky now explores the notion of "practice making perfect", which strikes him as being odd -- "You might expect, instead, that the more you learned, the slower you would get -- from having more knowledge from which to choose! How does practice speed things up?"

Minsky proceeds to suggest that what goes on is a special type of learning by bridging, in which links are created from an existing program to a set of new and simpler processes (which he calls a script). The script reflects the essence of the program, provides a more compact notation for it, and presumably allows the program's consequences to be produced by more efficient means. "The people we call `experts' seem to exercise their special skills with scarcely any thought at all -- as though they were simply reading preassembled scripts. Perhaps when we `practice' to improve our skills, we're mainly building simpler scripts that don't engage so many agencies."

Q: Is building a script an instance of reformulation? Another q: What is Minsky really talking about when he describes a program or a script? How are these ideas, central to conventional AI, to be converted into his SOM perspective? For example, is a society of agents a program?

13.6 The Frontier Effect

Minsky starts this chapter with a nice example of a developmental trait revealed by a drawing task -- see p. 138 for this rendition of the "frontier effect" -- the tendency to place new features at locations that have easily described relationships to other, already represented features. Copying is hard to do for children -- Minsky suggests that instead they use abstract descriptions of scenes as they "copy", and that the use of such descriptions produces this effect.

(NB: The frontier effect makes a nice case for the importance of relationships among features, which was a key theme of Chapter 11 ("The shape of space").

13.7 Duplications

Sometimes, it makes sense to count a single feature more than once. Minsky provides a nice example in which "we seem to see two complete arches, despite the fact that there aren't enough legs to make two separate arches."

"The double-arch problem also offers a choice of description styles. If you plan to build several separate things, you'd better keep count carefully or take the risk of running out of parts. But if you do that all the time, you'll miss your chance to make one object serve two purposes at once." This is easier to do by focusing upon functional descriptions instead of structural descriptions. "This doesn't mean that functional descriptions are necessarily better. They can make it hard to keep track of real constraints; hence they have a certain tendency to lead toward overoptimistic, wishful thought."

(NB: It is amusing to me to relate this last point to the "biological vacuum" of classical cognitive science, or, at the opposite end of the spectrum, to Paivio's "house of cards" attack on the nonempirical nature of computational theories).

Notes on Chapter 14, "Reformulation"

14.1 Using Reformulations

When we can't solve a problem, we try to find a new way to look at it. "Reformulation is the most powerful way to attempt to escape from what seems to be a hopeless situation." Minsky suggests that while an inventor might appear to be full of a vast array of novel ideas, in reality the inventor might simply be using variations of far fewer themes -- reformulation as the mother of invention! "In the long run, the most productive kinds of thought are not the methods with which we solve particular problems, but those that lead us to formulating useful new kinds of descriptions." New ideas often have roots in old ones.

(NB: The inventor example makes me think of creativity in rule-governed systems -- infinite variety coming from finite (but compositional and systematic) resources).

14.2 The Body-Support Concept

Simple cuts can often seem meaningful (e.g., the cut between "body" and "support"). Power of such ideas is that they can be extended to many domains via bridging definitions. "But the point is that it is not enough just to link together descriptions from two different worlds ... we must also know how this is done.

(NB: I didn't get it from this section, but later it is quite evident that Minsky views body-support as a fundamental idea, much like the cultural metaphors that Lakoff and Johnson argue permeate language. At my first reading of this chapter, I focused instead on the distinction between knowing how and knowing that, and wondered whether the requirement for both of these kinds of knowledge analogous to Minsky's notion here of detailed structure/function links.)

14.3 Means and Ends

We have many ways to connect 0our resources to our goals, and each of these ways has its own criteria for identifying essential parts. In fact, we might employ several of these approaches simultaneously when solving problems.

"The quality of our understanding depends upon how well we move between those different realms. In order to translate easily from one of them to another, we must discover systematic cross-realm correspondences. However, finding these is rare. ... What is remarkable about the body-support concept is how often it leads to systematic cross-realm correspondences." Cross-realm translations are the roots of fruitful metaphors.

(NB: For many, many examples of things that might count as being as fruitful as body-support, see Lakoff and Johnson's book *Metaphors We Live By*)

14.4 Seeing Squares

"We often self-impose assumptions that make our problems more difficult, and we can escape from this only by reformulating those problems in ways that give us more room." What makes us see squares in so many different things? "Psychologists have long wondered how we recognize such similarities but often forgot to ask how we recognize the very simplest forms of squares in the first place. Which comes first in recognition, specific features or global shapes? It must depend upon one's state of mind."

(NB: Two points. Local features vs. global form theories of perception are on opposite sides of a very long-standing debate in the literature. One example of a contribution to this debate is Dawson & Harshman, 1986. Second point is that Minsky now appears to be making a very strong appeal to top-down influences on perception. How cognitive does he want to be with this position?)

14.5 Brainstorming

"Your mind is constantly preparing ways to do [reformulation of complex scenes] by building up connections between different kinds of descriptions. Then, when you finally change your view to find another way to look at things, you can apply a lifetime of experience as easily as turning on a switch." Minsky suggests brainstorming as a method to break out of old thinking habits. "However, when you switch to unfamiliar views of things you may get new ideas, but you also put yourself in the situation of a novice; you become less able to judge which new ideas are likely to be compatible with any of your older skills."

(NB: One theme emerging here, which makes sense with emphasis on bridging definitions that link new material with old, is that old ideas can dominate or structure new.)

14.6 The Investment Principle

Unfortunately, some ideas gain their influence undeservedly. "The Investment Principle: Our oldest ideas have unfair advantages over those that come later. The earlier we learn a skill,

the more methods we can acquire for using it. Each new idea must then compete against the large mass of skills the old ideas have accumulated." In other words, it is easier to do new things in old ways. In the short-term, this is OK.

"I don't mean to say there's anything wrong, in principle, with using what you are comfortable with and already know. But it is dangerous to support your old ideas merely by accumulating ways to sidestep their deficiencies."

(NB: For an alternative slant on this, where even short-term effects can be negative, take a look at the functional fixedness literature in problem-solving.)

14.7 Parts and Holes

Minsky reformulates arch as a container. "Why focus so sharply on the concept of a container? Because without that concept, we could scarcely understand the structure of the spatial world." Hint at the end of this section is that the container concept is as powerful and general an idea as was the body-support concept.

14.8 The Power of Negative Thinking

To know how boxes keep things in, geometry is not enough. You also have to know how moving works. "The diagram below depicts an agency that represents the several ways an arm can move inside a rectangle ... If we connect each of these sub-agents to the corresponding side of our four-sided box frame, each agent will be able to test whether the arm can move in the corresponding direction."

(NB: This section is important for a number of reasons. First, Minsky provides an explicit example of agencies representing -- different agents detect motion in different directions (or, more correctly, the ability to move in different directions), and also are connected to properties in the world (which must also be represented by agencies). Second, Minsky introduces other functional primitive -- inhibitory connections between agents. This feels more and more like connectionist-style modeling to me. Third, we get a sense here of bridging of a different sort -- bridging between agencies that represent structural descriptions (i.e., properties of the arch) and agents that represent functional affordances (i.e., ability to move).)

14.9 The Interaction-Square

"What's so special about moving left or right or up or down? At first one might suppose that these ideas work only for motions in a two-dimensional space. But we can also use this square-like frame for many other realms of thought, to represent how pairs of causes interact." An example of such an interaction is how a combination of horizontal and vertical movement produces diagonal movement.

How do we learn to control movements mediated by our joints? Minsky suggests that we use interaction-square arrays. "An interaction-square array provides a convenient way to represent all the possible combinations...If square-arrays can represent how pairs of causes interact, could similar schemes be used with three or more causes?" No, too complicated to use.

(NB: Two points. First, Braitenberg -- with his use of networks to represent hyperspace -- does not shy away from multidimensional representations. Second, is the interaction-square another primitive? Or is it merely a convenient description of how different agents interact, as might be the case for the example given in Chapter 14.8?)

Notes on Chapter 15, "Consciousness and Memory"

15.1 Momentary Mental State

"Consciousness does not concern the present, but the past: it has to do with how we think about the records of our recent thoughts." How can thinking about thoughts be possible at all? Minsky's move is to have agents sensitive to events inside the brain.

Minsky then proceeds to argue for a limited capacity of consciousness, which produces a serial stream of consciousness. Why can't we reflect on the present? Because the present is not stable: "It is virtually impossible to speak of the shapes of things that change into something else each time we try to think of them. And that's what happens when we try to think about our present thoughts -- since each such thought must change our mental state."

(NB: This notion of consciousness, i.e., thinking about thoughts, reminds me directly of Minsky's earlier discussion of B-brains.)

15.2 Self-Examination

What do we mean by "consciousness"? Questions about consciousness are really questions about our recent past. "Why is it so hard to talk about our present state of mind? We've already seen several reasons for this. One is that

the time-delays between the different parts of a mind mean that the concept of a `present state' is not psychologically sound." (NB: See also Dennett's attack on the Cartesian Theatre in his book Consciousness Explained.)

"It is unlikely that any part of the mind can ever obtain complete descriptions of what happens in the other parts, because, it seems, our memory-control systems have too little temporary memory even to represent their own activities in much detail." (NB: This strikes me as a critique of introspection, in the sense that limited capacity will limit the accuracy of introspection. This could be viewed as another move against lay theories of the self/mind, as were developed earlier in the book.)

15.3 Memory

The mind has to "juggle fragments of its mental states." How do agents keep track of fragments during imagining? "They must be stored as memories. But what do we mean by that? Some readers may be surprised to learn that biologists still have no well-established theory of what happens in our brains when memories are formed." Minsky introduces LTM vs. STM distinction, and then suggests that we don't remember very much. "Our various agencies selectively decide, unconsciously, to transfer only certain states into their long-term memories -- perhaps because they have been classified as useful, dangerous, unusual, or significant in other respects."

(NB: First, the limited-capacity-of-memory theme is reflected here too, this time with respect to transferring information into more permanent memories. Second, the notion of "keeping track of fragments" reminds me of the kinds of problems that must be faced in visual cognition, problems which Pylyshyn's FINST mechanisms are supposed to solve.)

15.4 Memories of Memories

Many of our childhood memories are likely reconstructions -- not really memories at all. "I suspect that this `amnesia of infancy' is no mere effect of decay over time but an inevitable result of growing out of infancy." Minsky describes memories as "reconstructions of previous states of mind", which is certainly consistent with the notion of K-lines that he introduced a while back. "Memories are processes that make some of our agents act in much the same ways they did at various times in the past." Question: From this view of memory, how does limited capacity enter the picture? Is there a limit to the number of agents that can be active at any time?

15.5 The Immanence Illusion

Memories are only reproduced mental fragments. "Then what makes some recollections seem so real? The secret is that real-time experience is just as indirect! The closes we can come to apprehending the world, in any case, is through the descriptions our agents make." This leads to Minsky's "Immanence Illusion: Whenever you can answer a question without a noticeable delay, it seems as though tat answer were already active in your mind." (NB: This makes sense to me from the view that consciousness = recent memory.) Not all information is immediately available, though, which again reminds me of the temporal problems with consciousness that Dennett describes in *Consciousness Explained*.

15.6 Many Kinds of Memory

Minsky argues for many different kinds of memory, describing them as different kinds of memory agencies. Why do we have so many? Minsky doesn't answer this question here. (NB: Is splitting memory into different memory agencies going to buy us more than the more traditional view of having different memory systems, which (as far as I can tell) is motivated by exactly the same evidence that Minsky is keying into here?)

15.7 Memory Rearrangements

Here Minsky describes symbol manipulation (if I can really use that phrase here) in terms of swapping the states of two agents, which requires memory buffers to succeed. These buffers must be specialized enough to store only that information to be swapped, and must also require sensitivity to timing, in order to perform the correct swap at the correct time.

(NB: There are some architectural hints here -- is Minsky buying into the structure/rule distinction? Control structure is very important in this section too, which strikes me as a crucial leaning-away from connectionism.)

15.8 Anatomy of Memory

How are memories controls? In this chapter, Minsky proposes several different mechanisms. Micromemory units are proposed as temporary K-lines that can quickly store or restore the states of many agents in an agency. Short-term memory units are used to store the memories of the micromemory units themselves. Information from either of these types of memories can also be stored in more permanent memory systems.

(NB: The thing that strikes me about this stuff is that the memory control structure is crucial, but in order to work it needs memory too. Is Minsky falling into Ryle's regress with respect to his memory systems? This is one chapter that has lots of potential architectural information, and I do wish that it was given a more detailed treatment by him.)

15.9 Interruption and Recovery

For system interrupts to work, memory is required, because after the interrupt call is successfully completed, the system must return to the state that it was in prior to the interrupt. "Why do we so often get confused when we're interrupted? Because then we have to keep our place in several processes at once. To keep things straight, our memory-control machinery needs intricate skills." Importantly, we are largely unaware of the intricate control structure required for this to work.

(NB: "Keeping track of multiple things" argument strikes me as having the same structure as Pylyshyn's arguments for FINSTs.)

15.10 Losing Track

Why do we lose track in language, but not in vision? "One reason is that our visual-systems support more simultaneously operating processes than our language-systems can, and this reduces the need for any process to interrupt the other one." In language, it takes lots of learning to keep track well.

(NB: First, FINST stuff suggests that visual "keeping track" is harder, or perhaps more important, than Minsky believes. Second, I think that the distinction between vision and language in terms of the number of required interrupts is really very, very clever.)

15.11 The Recursion Principle

"The best way to solve a hard problem is to break it into several simpler ones, and break those into even simpler ones. Then we face the same issue of mental fragmentation. Happily, there is another way. We can work on the various parts of a problem in serial order, one after another, using the same agency over and over again. Of course, that takes more time. But is has one absolutely fundamental advantage: each agency can apply its full power to every subproblem."

"The Recursion Principle: When a problem splits into smaller parts, then unless one can apply the mind's full power to each subjob, one's intellect will get dispersed and leave less cleverness for each new task." (NB: Logo programs are good demonstrations of the computational power, and the efficiency of described operations, that recursion permits.)

To take advantage of recursion, we must be able to remember an interrupted state, which requires fast but short-term memories.

(NB: Are Elman's context units a good example of this kind of micromemory? Or do they have too little control?)

Notes on Chapter 16, "Emotion"

16.1 Emotion

Minsky views emotions as types of thoughts, types of thoughts that we wrongly credit with lots of stuff that reason does. Emotion is viewed as a product of conflicts among goals. "The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any emotions."

16.2 Mental Growth

How do our minds form? "We'll start by envisioning a simple brain composed of separate `proto-specialists', each concerned with some important requirement, goal, or instinct like food, drink, shelter, comfort, or defense. But there are reasons why those systems must be merged." What reasons for merging? To allow control structure to deal with conflicts, and to get components to exploit the knowledge of others. Furthermore, new goals must be learned. "Consequently, our genes must build some sort of `general-purpose' machinery though which individuals can acquire and transmit goals from one generation to another."

(NB: This is a crucial chapter, because I think here we are getting some proposals for primitives of the system. Specifically, I think that proto-specialists are Minsky's primitives. This is really intriguing to me, because these primitives really remind me a lot of some basic components of learning theory (drives/motivation); components which cognitivists have moved away from!)

16.3 Mental Proto-Specialists

NB: Right from the start of this section, I can see the need to relate Minsky's proto-specialists to things like Maslow's need hierarchy, and basic facts about motivation.) A set of individual proto-specialists is not a practical way to design an organism. "Most animals economize by

having all their proto-specialists share common sets of organs for their interactions with the outer world. ... Another kind of economy comes from allowing the proto-specialists to share what they learn. Whether you seek warmth, safety, nutrition, or companionship -- eventually you'll have to be able to recognize and act in order to acquire the objects you need." (NB: Is this an alternative to modularity?)

16.4 Cross-Exclusion

There must be a mechanism to select one goal from a set of competing goals. Minsky suggests a mechanism called cross-exclusion for this task; cross-exclusion is essentially a winner-take-all network of connections among agents. This kind of network can instantiate the principles of noncompromise. "Cross-exclusion groups can also be used to construct short-term memory units. Whenever we force one agent of such a group into activity, even for a moment, it will remain active (and the others will remain suppressed) until the situation is changed by some other strong external influence."

(NB: This is a very important architectural principle, with lots of history -- see for example Grossberg's work. Note that in such a system, there must be recurrent processing of signals from one agent to another.)

16.5 Avalanche Effects

The problem with networks is that as activity spreads, more and more things turn on, which can lead to paralysis or interference that stops the network from functioning. Cross-exclusion is one method that could be used to solve this problem; Minsky suggests others (conservation, negative feedback, censors and suppressors). "These methods are simple enough to be applied inside small societies, but they are not versatile enough to solve all the management difficulties that can arise in the more complex societies we need for learning to solve harder problems."

(NB: Theme here is focusing upon a particular control problem, self-regulation. Was this theme also a key issue that motivated Grossberg's work?)

16.6 Motivation

"Focusing" of learning is desirable, and could be implemented as a separate memory system for every goal. Cross-exclusion mechanisms might provide this separation, even if a common memory system is used. But this will lead to control problems too. Why? "Because each separate specialist is much too small and specialized to understand how the others work, the best each can do is learn to exploit what the others can do, without understanding how they do it.

(Q: What does this material have to do with the section title?)

16.7 Exploitation

"How could any specialist cooperate when it doesn't understand how the others work?" Minsky's answer is that you don't need such understanding: "Each part of the mind exploits the rest, not knowing how the other parts work but only what they seem to do." (NB: This strikes me as an interesting variant of functionalism, because it is a "functionalist attitude" being incorporated into the functioning of the components!)

16.8 Stimulus vs. Simulus

One agency can activate another by merely imagining a stimulus. Such fantasies work because they are merely summaries of the type that high-level agencies typically process. (NB: One example of this sort of thing, perhaps, is Finke's work on producing visual aftereffects by having subjects imagine part of the adapting display.)

16.9 Infant Emotions

Minsky takes sudden, drastic changes in infant moods as evidence that baby's minds are made up of nearly separate agencies. How does one explain such striking shifts? "One explanation of those striking shifts in attitude is that one agency attains control and forcibly suppresses the rest. Another view is that many processes continue at once -- but only one at a time can be expressed."

What would be the advantage of the single-expression design? "Perhaps that artificial sharpening promotes the child's welfare by making it easier for the parent to respond to whichever problem has the greatest urgency."

(NB: Again, this chapter reinforces the notion that proto-specialists are tied to very basic needs, and are the developmentally primitive agents.)

16.10 Adult Emotions

What are emotions? Tough question! Little agreement! Why? Because "when we learn

such words, we each attach to them variously different and personal accumulations of conceptions in our minds."

"Our earliest emotions are built-in processes in which inborn proto-specialists control what happens in our brains. Soon we learn to overrule those schemes, as our surroundings teach us what we ought to feel." (NB: This strikes me as interesting, in the sense that emotions appear to be portrayed as cognitive primitives. This makes sense, though, if you view emotions as the expression of conflicts among goals, and if the primitive goals (agents) in the system are very basic and important needs -- strong emotions will result when such goals conflict!)

Notes on Chapter 17, "Development"

17.1 Sequences of Teaching-Selves

How is the adult "sense of unity" produced from the infant proto-specialists? "We'll speculate that this coherency is acquired over many `stages of development'. Each new stage first works under the guidance of previous stages, to acquire some knowledge, values and goals. Then it proceeds to change its role and becomes a teacher to subsequent stages."

How is such teaching to be accomplished? Merely by recognizing and rewarding solutions, or by providing new goals. New stages are suppressed until tested (see Minsky's previous account of adding new layers to a hierarchy for a similar point).

"How could so many steps and stages lead to any sense of unity?" Minsky speculates that old stages still remain, to be used when required. "One's present personality cannot share many of the thoughts of all one's older personalities -- and yet it has some sense that they exist. This is one reason why we feel that we possess an inner Self -- a sort of ever-present person-friend, inside the mind, whom we can always ask for help."

(NB: Keep in mind this notion of development, and consider it in the context of Shultz' use of cascade-correlation. Q: Why do new stages emerge? What drives this kind of development? These are questions that Minsky has not yet answered for me!)

17.2 Attachment-Learning

Minsky uses a difference engine formulation to describe how different types of learning are activated by different contexts (e.g., being scolded by a stranger vs. being scolded by a parent). With ordinary failure contexts, the learner changes methods used to accomplish the goal. With fear contexts, the learner might change the description of the situation. With attachment-learning -- parental provided context! -- the learner modifies their notion of which goals are worthy of pursuit. (NB: All of this strikes me as a very non-connectionist approach to learning, because it is strategic -- the context determines what is to be changed!)

17.3 Attachment Simplifies

What is the function of childhood attachment? 1) Safety. 2) Child uses this relationship to learn goals from older people. "Even though there are many ways a child could learn about ordinary causes and effects, there is no way for a child to construct a coherent system of values -- except by basing it upon some already existing model." Given that goal-directed behavior is critical for Minsky (after all, this is what describes a control structure!), one gets the sense that attachment is viewed as a primitive process that provides a context in which goals themselves are learned.

17.4 Functional Autonomy

How do we invent new goals? By decomposition! Simpler subgoals eventually become detached from higher-order goals that they subserve, and evolve into more ambitious pursuits.

(NB: This is worrisome to me, because I find it hard to fit into the general practice of functional analysis. Perhaps what is missing from the section is the notion that as subgoals become more and more removed, the vocabulary required to meet the subgoal changes, and becomes more challenging. For instance, identifying a "red detector" is a simple functional goal, but explaining the biochemistry of light transductions is very complicated -- once the vocabulary shift is achieved!)

17.5 Developmental Stages

Why does development appear to be stage like? "I'll argue that nothing so complex as a human mind can grow, except in separate steps. One reason is that it is always dangerous to change a system that already works."

"Another conservative strategy is never to let a new stage take control of actual behavior until there is evidence that it can outperform its predecessor. (NB: This reminds me of cascade-correlation model of development, because Min-

sky points out that development is occurring during the plateau!)

(NB: How does this material relate to Brainerd's famous BBS article on developmental stages?

17.6 Prerequisites for Growth

What controls the pace of mental growth? One suggestion is the requirement for prerequisites, which also produces developmental stages. "The reason we know so little about how children's minds grow is that we can't observe the processes that are responsible. (NB: The bottom line here is that theories of development are underdetermined by the data.)

17.7 Genetic Timetables

Timing of adding new "management" layers to an agency is critical, and could be controlled genetically. This could explain universal stages. "Once it becomes too hard to change an old agency, it is time to build another one; further progress may require revolution rather than evolution. This is another reason why a complex system must be grown in a sequence of separate steps."

17.8 Attachment-Images

"Our attachment mechanisms force us to focus on our parent's ways, and this leads us to build crude images of what those parents themselves are like. That way, the values and goals of a culture pass from one generation to the next. They are not learned the way skills are learned." Attachment might be a method of imposing coherence on a bunch of mindless agencies. (NB: Here, Minsky shows a strange mix of Freud and cognitivism. This reminds me of an article that I read in the Annals of Theoretical Psychology that argued that Freud's notion of cognitive functions was pretty mainstream for his day.)

17.9 Different Spans of Memories

Parent/child bonds must be based on some sort of memory. "I suspect that attachment-bonds involve memory-records of a type that can be rapidly formed but them become peculiarly slow to change." Here, Minsky relates this kind of memory to a range of different phenomena. (NB: Different spans of memory can also be related to earlier notion of different time-lines of experience or agents.)

17.10 Intellectual Trauma

Minsky again gives ode to Freud. Social and intellectual failures are argued to have similar consequences; intellectual failure produces a fear about one's state of mind. (NB: This reminds me of the old social perception literature where stooges would call a short line long and as a result so would a subject, who then had to be seriously debriefed because of concerns about their visual faculties.)

17.11 Intellectual Ideals

"Human though is not based on any single and uniform kind of `logic', but upon myriad processes, scripts, stereotypes, critics and censors, analogies and metaphors. ... We can make intellectual attachments, too, and want to think the way certain other persons do. ... I suspect we depend as much on images of how we ought to think as we do on images of how we ought to feel."

(NB: This section makes sense to me, given Chapter 16's view of the similarity between emotion and intellect.)

Notes on Chapter 18, "Reasoning"

18.1 Must Machines Be Logical?

Minsky takes issue with the view that machines must think with perfect logic, noting that logical accounts of the internal workings of a machine are not the same as accounts of the future actions of that machine. (NB: This reminds me a lot of Braitenberg.) Minsky argues that we only use logic to think after the fact, to "clean up" accounts of how problems were solved, and that logic does not explain much about how we think. (NB: This last point reflects the book's strong emphasis on control structure.)

18.2 Chains of Reasoning

Lots of ideas are easily phrased as chains, and we often delete internal links and just speak of the ends. "For generations, scientists and philosophers have tried to explain ordinary reasoning in terms of logical principles -- with virtually no success. I suspect this enterprise failed because it was looking in the wrong direction: common sense works so well not because it is an approximation of logic; logic is only a small part of our great accumulation of different, useful ways to chain things together."

(NB: The anti-logic stance adopted here is consistent with lots of psychological literature,

and is pursued in remaining sections. Note, though, that it is also a popular connectionist move...)

18.3 Chaining

Chaining is important because it can be used in many different domains. Minsky works the analogy here that breaking of chains is equivalent to the failure of reasoning. Question: Is chaining merely associational? This might be all that is required to get the properties that Minsky describes here.

18.4 Logical Chains

What is the difference between chains and logical arguments? "The difference is that in logic there's no middle ground; a logic link is either there or not. Because of this, a logical argument cannot have any `weakest link'." Common sense reasoning looks for overall plausibility, and does not test each and every link in an argument. Minsky suggests that logic has limited application in cognition, providing one method to detect weaknesses in old ideas.

(NB: This does raise the important question of why many assume that thinking is logical, when lots of scientific evidence suggests that it is not.)

18.5 Strong Arguments

What does strength -- as in "strong argument" have to do with reasoning? Strength of the chain, which is not a logical notion! Strength is reflected in parallel, converging lines of evidence, redundant links. (NB: Here is a very neat notion -- parallelism (instantiated by converting chains into bundles, chains of bundles) provides damage resistance to types of reasoning!)

18.6 Magnitude from Multitude

"Why don't we just use cool, clear, faultless reasoning to prove we are right? The answer is that we rarely need to know that anything is absolutely wrong or right; instead, we only want to choose the best of some alternatives."

Minsky suggests two strategies for such a choice. In strength from magnitude, we look at cooperative or competitive sums of forces (sum of the evidence? In strength from multitude, we count the number of reasons in favor of a decision. For both of these strategies, "strength" is a measure of how likely a decision is to succeed. Question: How are these two strategies related to what we know about people's decision-making heuristics?

18.7 What Is A Number?

Meanings reflect networks of ideas, and therefore are personal or idiosyncratic. "In order for two minds to agree perfectly, at every level of detail, they'd have to be identical."

Math is a domain where agreement between different meanings might be best, but this agreement is still not perfect. "Even something as impersonal as `Five' never stands isolated in a person's mind but becomes part of a huge network.

"The really useful `meanings' are not the flimsy logic chains of definitions, but the much harder-to-express networks of ways to remember, compare, and change things. A logic chain can break easily, but you get stuck less often when you use a cross-connected meaning-network; then, when any sense of meaning fails, you simply switch to another sense."

(NB: On the one hand, you get a connectionist sense of linear, serial logic vs. parallel, redundant (and therefore robust) webs of ideas. But this notion of switching meanings -- which strikes me as a control issue -- is decidedly not connectionist!)

18.8 Mathematics Made Hard

Minsky describes science as practicing logical reasoning, but argues that this is not good psychology. "In real life, our minds must always tolerate beliefs that later turn out to be wrong. ...Why do so many schoolchildren learn to fear mathematics? Perhaps in part because we try to teach the children those formal definitions, which were designed to lead to meaning-networks as sparse and thin as possible."

18.9 Robustness and Recovery

Why do our minds continue to function, even as they change? Most machines that we build would certainly not work! Minsky suggests several reasons: duplication (see also Medler & Dawson, 1994!), self-repair, distributed processes, and accumulation.

Of these, accumulation strikes me as a new concept. With it, "each agent tends to accumulate a family of subagents that can accomplish that agent's goals in several ways. Later, if any of those subagents become impaired, their supervisor will still be able to accomplish its job, because other of its subagents will remain to do that job, albeit in different ways."

(NB: Lots of this stuff could have come right out of the PDP'86 bibles, with the exception of accumulation -- connectionists rarely talk about alternative strategies, with a system capable of switching from one to another!)

Notes on Chapter 19, "Words and Ideas"

19.1 The Roots of Intention

Words can't be the substance of our thoughts. "We must discard the usual view that words denote or represent, or designate; instead, their function is control; each word makes various agents change what various other agents do." We often think in words, with no conscious sense of how or why this happens. Our introspective abilities are too weak to answer questions about language.

(NB: Again, Minsky attacks our intuitions gained from introspection. In general, lay psychology can't be right, because we don't have enough access to our cognitive machinery.)

19.2 The Language-Agency

"As far as consciousness can tell, no sooner do we hear a phrase than all its meanings spring to mind -- yet we have no conscious sense of how those words produce their effects." However, language does play an important role in consciousness. Minsky's first approach to language is to divide the system into three different regions: 1) agents concerned with words; 2) agents concerned with how words engage mental processes; 3) agencies affected by words.

(NB: Right now, I don't see where this is going -- what is the motivation for this separation? How does it relate to other carvings of language, such as syntax vs. semantics?)

19.3 Words and Ideas

How do words evoke complex states of mind? "Many people have tried to explain language as though it were separate from the rest of psychology." Minsky argues against this. His view is that language must be treated as part of thought. Minsky's move is to integrate language and thinking by exploiting two new kinds of agents. 1) Polynemes -- a type of K-line that sends the same simple signal to many different agencies, each of which must learn what to do when that signal is received. 2) Isonomes -- controls a short-term memory in each of many agencies.

(NB: These building blocks strike me as being very associationistic in nature. Does Minsky's notion of semantics boil down to experiences associated with specific words?)

19.4 Objects And Properties

Some aspects of word meanings can be captured via lists of properties. But what is a property? 1) It is a stable characteristic. 2) It is independent of other characteristics. 3) It reflects the nature of reality.

"We derive a wonderful power from representing things in terms of properties that do not interact: this makes imagination practical. It lets us anticipate what will happen when we invent new combinations and variations we've never seen before." Minsky's move is to let different agencies represent different properties. "That way, a single word can activate many different kinds of thoughts at once! Thus the word `apple' can set your Color agency into a `redness' state, put your Shape agency into a `roundness state' ...etc."

(NB: On the one hand this reminds me a lot of the old semantic network vs. set theoretic debate in cognitive psychology; because each of these approaches really viewed objects as a collection of properties. On the other hand, this stuff makes me wonder -- what in the world is a "word", in the sense of what "word" properties make it possible for all of these properties to be activated?)

19.5 Polynemes

A polyneme is 1) an agent, 2) sends messages to a number of different agents, and 3) can have different effects on different agents. How is this last property possible? Because controlled agencies must have learned their particular response to particular signals. "To understand a polyneme, each agency must learn its own specific and appropriate response. Each agency must have its private dictionary or memory bank to tell it how to respond to every polyneme." What mechanism could account for this? K-lines. Meanings, then, are being treated as selected (context dependent) memories.

(NB: Here, another puzzle is brought to mind. Previously, it was noted that each agency on the receiving end of a polyneme gets the same signal. What in the world is a signal? Do different polynemes send different signals, or is it assumed that the same kind of signal is always sent? If the latter case is true, then how are different polynemes differentiated by receiving agencies?

19.6 Recognizers

How do we recognize objects? By verifying that it has certain properties, which could be done with an AND gate evaluating the presence of all properties. Problem, though, is that this method is not going to work very well because it is too demanding. A better move is to merely require that enough properties have been detected.

(NB: How does this fit into all of the stuff in chapters 12 through 14 about abstract properties, or idiosyncratic bridges between structure and function? Can objects really be identified?)

19.7 Weighing Evidence

Minsky provides an account of weighing evidence that is similar in spirit to feature weight summation. What kind of machine could do this? A perceptron! Minsky back peddles a bit, then ... "All feature-weighing machines have serious limitations because, although they can measure the presence or absence of various features, they cannot take into account enough of the relations among those features." (NB: I don't think this claim is right, because finding such relations are exactly the purpose of hidden units. At any rate, though, Minsky's sympathy with connectionism is coming through.)

19.8 Generalizing

"How do we make generalizations from fragmentary bits of evidence?" No one has solved the problem of generalization. Minsky's slant on this is uniframes, level bands, and now polynemes. (NB: For connectionist systems in general, generalization is a real problem. If Minsky is buying into a connectionist model, then his concern with generalization is justified.)

19.9 Recognizing Thoughts

"How do we recognize our own ideas?" No mystery because similar mental states can arise for very different reasons. Input from memory can have the same result as sensory input.

Q: What does it mean to recognize an idea? Does it mean to recognize it as an idea, or rather to understand the idea's content?

19.10 Closing the Ring

Now, Minsky elaborates his notion of the language agency by including recurrent connections. It turns out that this is a definitive property of his SOM architecture.

Why recurrence? "If you start with enough clues to arouse one of your applenemes, it will automatically arouse memories of the other properties and qualities of apples and create a more complete impression, `simulus', or hallucination of the experience of seeing, feeling, and even of eating an apple. This way, a simple loop machine can reconstruct a larger whole from clues about only certain of its parts!"

Notes on Chapter 20, "Context and Ambiguity"

20.1 Ambiguity

Not only words, but thoughts themselves, are ambiguous. One reason for this is because we only have partial access to the machinery representing thoughts. A second reason is that thinking about thoughts changes them, "the problem is that our states of mind are usually subject to change".

If all of this is ambiguous, then why do we have a clear sense of most sentence meanings? 1) Context. 2) We are adept with coping with ambiguity, because we always deal with the ambiguity of our own thoughts.

(NB: Just a thought -- decreasing ambiguity, by instantiating soft constraints, is a very common use of connectionist networks.)

20.2 Negotiating Ambiguity

Ambiguity can reveal itself even in very simple sentences. Context can help clarify such ambiguities, but how? 1) "Spreading activation" among specific agents, 2) sharpening of activation by recurrent connections. Problem: with this notion, won't everything at some point activate? Solution: no -- prevent this with cross-exclusion mechanisms.

(NB: This looks very much like a recurrent version of a semantic network. What kind of model? Well, the IAC network (for instance) that we demonstrated with the Sharks/Jets data set!)

20.3 Visual Ambiguity

Ambiguities are common in vision, too, but they are resolved so quickly that we often are not aware of their existence. Ambiguities might be resolved by low or high level processes. Minsky leans towards the "New Look" of ambiguity resolution, favoring constraints imposed by higher-order processes. (NB: Ambiguity reduction in vision, as exemplified by solving problems of

underdetermination, is often accomplished via constraint satisfaction imposed by recurrent connectionist networks!)

20.4 Locking-In and Weeding-Out

Ambiguity could be viewed as the activation of more than one polyneme. But context activates other agents, which will support some polynemes and inhibit others. Ambiguity is thus weeded out.

When ambiguity is gone, an interpretation is locked in. Sometimes, though, this will lead to a mistake, which must be corrected. Minsky's solution is to inhibit currently active agents (e.g., a B-brain could do this!), and then start the weeding-out process again. (NB: This procedure is very, very much like the kind of processing that Grossberg describes in a particular type of connectionist architecture, ART!)

20.5 Micronemes

Classifying things via property lists won't work all the time, because properties interact with context. To deal with this, Minsky introduces the term "micronemes", "those inner mental context clues that shade our minds' activities in ways that we can rarely express. There is a somewhat different microstructure to each person's thoughts; indeed, their inexpressibility reflects our individuality." Micronemes will be implemented using K-lines. The sense given here is that they are agents that produce only small or subtle signals, which in turn mitigate which agencies eventually turn on.

(Q: What is the relationship between Minsky's micronemes and Smolensky's subsymbols?)

20.6 The Nemeic Spiral

"Our polynemes and micronemes grow into great branching networks that reach every level of every agency." This is an approximate hierarchy, but there are lots of messy cross-connections too.

Minsky alludes to a spiral of activity where a B-brain could choose the level of agents that are most active (i.e., focus on small details when things are going well, but move back when things are going poorly).

(NB: If there are many recurrent connections in the system, then the system could not be described as hierarchical!)

20.7 Connections

"To learn the proper use of a single word must involve great numbers of connections between the agents for that word and other agents." This raises two issues -- what causes these connections, and how are they physically instantiated?

(NB: These are core issues, and after raising them, Minsky skirts them saying only that they don't need to be direct!)

20.8 Connection Lines

Minsky briefly describes Mooers' scheme to get a high degree of functional connectivity with a small amount of physical connections. "The trick is to make each transmitting-agent excite not one, but five of those wires, chosen at random from the available ten. Then each receiving-agent is provided with an AND-agent connected to recognize the same five-wire combination.

How might such connections arise? Essentially, Minsky views the receiving agents as perceptrons, which means we wind up with something like association via the delta rule!

(This is getting more and more connectionist in spirit!)

20.9 Distributed Memory

Minsky brings to our attention the fact that when agents are interconnected, each agent might have multiple inputs and multiple outputs. Finally, he makes a move towards explicit connectionism. We now get an essentially explicit statement (with his talk of Boltzman machines) that SOM boils down to a connectionist network, that has local encoding (possibly in connections, as mentioned at the end of the section), and recurrent processing.

(NB: My hypothesis: Minsky's 1986 revolution was recurrent, local connectionism!)

Notes on Chapter 21, "Trans-Frames"

21.1 The Pronouns of the Mind

"Pronouns do not signify objects or words; instead, they represent conceptions, ideas, or activities that the speaker assumes are going on inside the listener's mind. But how can the listener tell which one of the activities is signified when there are several possibilities?" Minsky

suggests that pronomolization is not language specific, and introduces the term "pronome" -- "temporary `handles' for taking hold of, and moving around, those active fragments of mental states."

(NB: Do pronomes serve the same function as FINSTs? Minsky's argument for the requirement of pronomes shares some properties with Pylyshyn's.)

21.2 Pronomes

Sentence comprehension requires paying attention to roles, such as time, place, trajectory, actor, etc. The importance of roles is reflected in the fact that they are made explicit in syntax. Roles are rooted in metaphor and analogy (i.e., in terms of when they can be applied), such as parallel relations between time and space. "Many of our language-grammar `rules' embody or reflect some systematic correspondences -- and these are among our most powerful ways to think."

(NB: All of this makes me think of Jackendoff's analysis of conceptual structure, which makes the things that Minsky calls roles explicit, and which assumes that such structures are made explicit in syntax.)

21.3 Trans-Frames

Minsky offers the trans-frame as a compact way of representing conceptual dependencies. The same trans-frame can be used in a variety of semantic domains. Why? Because learning efficient, chain-manipulating skills provides a general reasoning skill.

(Q: How is the trans-frame to be translated into agents?

21.4 Communication Among Agents

How can an agent like "Get" know what to get, seeing as "get" must by definition be simple/stupid? Move is to make "Get" a middle level manager that signals other agents; trans-frame provides context for the control structure. Agents can operate without explicit messages via pronomilization of thought.

(NB: Here, Minsky is working to show how Thurstone's rate is not lost in thought!)

21.5 Automatism

The simpler the agent, the simpler the message needed to control it. Still, reality must enter the system at some point. How? Agents in touch with reality look for real-world objects that have properties consistent with those represented properties (agents) that have been primed by context. I.E., context focuses attention to properties of 1 small part of the world, and this part becomes the "referent" for other agents.

21.6 Trans-Frame Pronomes

Here, Minsky elaborates his trans-frame to be more general (i.e., to include components consistent with Jackendoff's notion of conceptual structure. Minsky argues that this is a general structure prominent in many styles of thinking. "I suspect that Trans-like structures have a special prominence in how we think. One reason is that some sort of bridge like scheme seems indispensable for making those all-important connections between structures and functions."

(NB: For something so prominent, I have little sense of what its structure is, even with the supplied figure!)

21.7 Generalizing Witt Pronomes

Once agents are properly connected to world referents, a control structure must be engaged to guide appropriate actions in the world. Minsky's solution is to define actions over role agents in the trans-frame, not on the objects that they refer to. As a result, one control structure will apply to many, many different situations.

(NB: Does a similar argument find its way into the FINST literature, or into Ullman's account of the objects that visual routines act upon?)

21.8 Attention

"When several objects move at once, it's hard to keep track of them all." Minsky equates attentional limits with STM. Then he ties this in with the development of object permanence.

(NB: Pylyshyn and Storm's data, along with Pylyshyn's FINST model, would indicate that this might be a very dated notion of attention. How much would Minsky's story change here if the notion of a single "attentional spotlight" were abandoned?)

Notes on Chapter 22, "Expression"

22.1 Pronomes and Polynemes

Because of its temporary nature, a pronome must be a type of STM -- a temporary K-line. In contrast, polynemes are permanent K-lines.

Question: What is the physical difference between temporary and permanent K-lines? Is this difference qualitative, or merely a matter of degree?

NB: In connectionist terms, STM = unit activity, while LTM = connection weights!

22.2 Isonomes

What is the difference between a polyneme and a pronome? "The difference is that a pronome has essentially the same effect on each of its recipients -- namely, to activate or to assign a certain short-term memory-unit. I'll introduce a new word -- 'isonome' -- for any agent that has this sort of uniform effect on many agencies."

Isonomes would be expected to exist because different agencies might have many structural similarities. Pronomes are only one type of isonome. "The power of polynemes stems from how they learn to arouse many different processes at once, while isonomes draw their power from exploiting abilities that are already common to many agencies."

NB: The move to name different types of processes with respect to this property (i.e., same vs. different effects) is striking. Does it suggest that other crucial architectural differences must also be true of isonomes vs. polynemes?

22.3 De-Specializing

How are skills despecialized (or generalized)? One approach would be to replace polynemes with less specific isonomes. But this replacement -- which is an example of abstraction -- must be constrained to prevent absurdities.

"What we call 'generalizing' is not any single process or concept, but a functional term for the huge societies of different methods we use to extend the powers of our skills. No single policy will work for all domains of thought, and each refinement of technique will affect the quality of the generalizations we make."

NB: Generalization is incredibly important for connectionist networks. But connectionist studies of generalization ignore Minsky's insight in the quote above. What would happen to connectionism if it thought of different kinds of generalization, and different techniques for exploring or improving these different kinds? For example, why not consider generalization as a learning effect (e.g., measuring savings) instead of sum of squared error to untrained examples?

22.4 Learning and Teaching

"The power of what we learn depends on how we represent it in our minds. We've seen how the same experience can lead to learning different action scripts by replacing certain polynemes with isonomes." As a result, the educational issue for Minsky becomes 'How do kids acquire representational skills?' "Each child learns, from time to time, various better ways to learn -- but no one understands how this is done."

NB: In this section, Minsky makes the interesting move of equating metacognition with the B-brain. I doubt that many developmentalists would agree with this move! However, it would be useful to speculate how B-brain concepts might alter the developmental study of metacognition.

22.5 Inference

Trans-frames can be used to modify chains of reasoning to permit new types of deduction in the absence of logic. But this move requires recognition that trans-frames can be applied to terms that are not perfect matches. "By learning to manipulate our isonomes, we become able to combine mental representations into structures that resemble bridges, chains and towers. Our language-agencies learn to express these in the form of compound sentences."

NB: All of this requires what Minsky calls 'conceptualization' -- treating mental processes as though they were object-things. Is this kind of reflexiveness possible in other (or even most other) architectural proposals?

22.6 Expression

Language gives sense of thoughts as ordinary things. "Why do we 'thing-ify' our thoughts? One reason is that this enables us to reapply the wonderful machines our brains contain for understanding worldly things." E.g., the method of loci mnemonic technique uses the brain's ability to recognize and represent familiar locations to remember ideas placed at these (mental) locations. "I suspect that, as they're represented in the mind, there's little difference between a physical object and an idea." Indeed, the (at least) temporary permanence of ideas is required for thought to proceed. "No mind can work without some stable states or memories."

NB: This parallels the kinds of observations that Lakoff and Johnson make about pervasive metaphors. Also, the notion of different kinds of permanence of ideas -- related back to previ-

ous distinctions of STM and LTM -- are reminiscent of Newell's (1990) discussion of the temporal nature of mental properties. (i.e., the more permanent something is, the more it looks like architecture!!)

22.7 Causes and Clauses

We seek causes for most changes that we observe. "I wouldn't be surprised to find that brains have built-in tendencies to try to represent all situations in certain special ways: THINGS, DIFFERENCES, CAUSES, CLAUSES." These tendencies are built into our very language, suggesting to Minsky that our brains make us seek to represent dependencies.

NB: The classes of representation noted here (things, differences, causes, clauses) remind me a great deal of some of the basic components of Jackendoff's conceptual structure. Jackendoff makes a great deal of use of the idea that the world offers (through our interaction with it) particular natural kinds; these overlap with Minsky's.

22.8 Interruptions

Toleration of interruptions requires using agents that control STM. Minsky describes sentence clauses as interruption. In particular, relative clauses interrupt main clauses to provide new information. Importantly, this is all described in terms of agencies -- Minsky provides an example of an interrupting clause (and a main clause) having the structure of a Trans-frame. Finally, Minsky proposes that certain wh words in English are signals to engage STM, and act to interrupt the language agency and at the same time have it store its present pronome assignments temporarily.

One neat thing here is what Jackendoff would call "the cognitive constraint" -- use of language to signal mental processes. From another perspective, consider garden path sentences. One (standard) approach to account for them would be to talk about parsing of the sentence. Minsky's alternative perspective might be to consider conflicts in processing -- the garden path syntax, for instance, would not signal the necessary interruption of processing, leading to problems when different agencies need to be coordinated.

22.9 Pronouns and References

How do we deal with ambiguous pronouns? One view would be that grammar provides constraints on pronoun assignment. But grammar by itself won't do this job all the time. Psychological factors (i.e., context, expectation) are also required.

"What does 'expectation' mean? At each point in a dialogue, both parties are already involved with various concerns and desires. These establish contexts in which each new word, description, or representation, however ambiguous, gets merged with whichever short-term memory best matches it.

How can speech be designed to be easily received? By basing it on the properties of the speaker's minds, and assuming that the listener's mind will have similar properties.

NB: How might this relate to work on pragmatics, a la Grice? Do researchers on language context view shared mental properties as being part of what they do? This reminds me too of some work on creativity, in which the aesthetic process involved reconstructing (in the appreciators mind) the representations that were originally constructed in the artist's mind -- the name of the author of that stuff is escaping me right now!!

22.10 Verbal Expression

People communicate effortlessly, but this simplicity is an illusion. (NB: The irony of this position in Minsky's book is the story -- confirmed by him in a Scientific American profile published in recent years -- that he viewed vision as being so simple that he assigned it as a summer computer programming project to an undergraduate student.)

Communication really seems to involve constructing structures in the minds of listeners. How? First, build a new version of the structure in your own mind. Second, with each step in 1), say the corresponding verbal expression so that the listener can copy the steps!

"To be able to do that, Mary must have learned at least one expressive technique that corresponds to each frequently used mental operation. And Jack must have learned to recognize those expressive techniques -- we'll call them grammar-tactics -- and to use them to activate some corresponding isonomes and polynemes.

NB: This turns the table on critics of protocol analysis! Minsky is saying that communication depends on our ability to create or interpret thinking aloud protocols.

22.11 Creative Expression

Minsky's theory of communication can lead to imperfect reconstructions in listener's

minds -- but this can be good. They can provide new insights, which (for instance) can lead to the solutions of problems.

"When we try to explain what we think we know, we're likely to end up with something new. All teachers know how often we understand something for the first time only after trying to explain it to someone else."

NB: The notion of reconstruction in language is central to the old empirical work of Bransford, Barclay, and Franks. One thing that this work points out is that when people hear or read sentences, veridicality is rarely the result -- understanding really does seem to be almost exclusively reconstructive, along the lines that Minsky has been discussing in Chapter 22.

Notes on Chapter 23, "Comparisons"

23.1 A World of Differences

Ordinary thought depends on recognizing differences. "This is because it is generally useless to do anything that has no discernable effect." Many familiar mental activities can be represented in terms of the differences between situations: PREDICTING, EXPECTING, EXPLAINING, WANTING, ESCAPING, ATTACKING, DEFENDING, ABSTRACTING.

One also can think about differences between differences (interactions). Minsky equates this to reasoning by analogy. "The ability to consider differences between differences is important because it lies at the heart of our abilities to solve new problems. This is because these 'second-order-differences' are what we use to remind ourselves of other problems we already know how to solve."

NB: The notion of detecting changes, and in particular high-order changes, is crucial. For example, change detection is central to lots of connectionist networks (like Grossberg's ART, or winner-take-all networks). The ability to detect higher-order relationships (though not necessarily higher-order differences) is the hallmark of New Connectionism -- it's what hidden units do. (From the Epilogue to the 1988 edition of Perceptrons, I doubt that Minsky would endorse this last point.)

23.2 Differences and Duplicates

The requirement to notice differences poses a problem for psychology: The Duplication Problem: The states of two different agencies cannot be compared unless those agencies themselves are virtually identical. Otherwise, the system is swamped by irrelevant differences. Problem is that this appears to require "duplicate brains".

NB: Duplicate brains may be an advantage, not a problem. For instance, from the selectionist standpoint championed by some researchers (such as Changeux in Neuronal Man) you might want lots and lots of structures that are quite similar, but not identical. Is this a problem for Minsky because he is essentially instructionist in his position?

23.3 Time Blinking

"Our senses react mainly to how things change in time." This is the key to Minsky's solution to the duplication problem. "Any agent that is sensitive to changes in time can also be used to detect differences. For whenever we expose such an agent, first to a situation A and then to a situation B, any output from that agent will signify some difference between A and B." As a result, only one agent is required to detect a difference.

NB: I'm quite interested in this as I write these notes (Aug. 16/98). For a long time now I've thought that there is an important relationship between motion detection mechanisms and learning in PDP networks, and I've been exploring some ideas which should lead this fall to the development of a new architecture. Here's the cool idea -- finding a way in which a network could learn to detect differences!

23.4 The Meanings of More

A term with different shades of meaning must still engage isonomes. Two key mechanisms will be involved -- a procedure for signaling that two pronomes must be compared, and a procedure for specifying the kind of difference that should be detected.

NB: Relating this back to motion detection -- if "more" can have many meanings, and depends on comparisons -- should there be different qualities of motion detected by the visual system? Perhaps all of the work on different sorts of motion perception (e.g., short- vs. long-range a la Braddick) could be reformulated by considering different qualities of motion that could be analyzed! In other words, pay attention to the motion, not to the mechanisms!

23.5 Foreign Accents

"Why do adults find it so hard to learn how to pronounce new word sounds? ... I suspect this particular disability is caused, more or less directly, by a genetically programmed mechanism that disables our ability to learn to make new

connections in or between the agents we use to represent speech sounds."

NB: This amounts to a version of a "special biology" view of critical periods, which is common in the literature. However, it raises some other important issues -- in particular, what is the specific link between biology and agency?

Notes on Chapter 24 “Frames”

24.1 The Speed of Thought

What could explain the blinding speed of sight? Expectations! Something like stereotypes could provide mechanisms for efficiency. "The moment you sense the presence of a person, a whole world of assumptions are aroused that are usually true about people in general. At the same time, certain superficial cues remind you of particular people you've already met. Unconsciously, then, you will assume that this stranger must also resemble them, not only in appearance but in other traits as well." For Minsky, these expectations are provided by frames.

NB: Relating this to material in Chapter 23, comparison is at the base of this too. As well, assuming that few differences exist between agencies is here too. Presumably, to break away from a frame for a particular person, difference detection would have to kick in. One wonders, then, if you could represent people as a stereotypical frame, duplicated, with specific differences made explicit...

24.2 Frames Of Mind

A frame is a skeleton of terminals, where terminal values are used to represent a particular entity. "In principle, we could use frames without attaching their terminals to anything. Normally, though, the terminals come with other agents already attached -- and these are what we called 'default assignments' when we first talked about level-bands."

Frames define expectations, based on past experience, to facilitate processing the present. But frames won't make perfect predictions, so we must learn to adapt them.

NB: How do you detect imperfect predictions? By detecting a difference between an activated frame and an agency representing the present state of affairs!

24.3 How Trans-Frames Work

How might a frame actually work? Imagine a current state of affairs represented by a pronome agent, and another agent representing a frame. An AND-agent would link the two. "According to this simple scheme, a frame could consist of little more than a collection of AND-agents, one for each of the frame's pronome terminals." How would a frame learn which polynemes should fill its terminals? You could have a "virgin K-line" as the thing directly connected to the terminal; eventually the terminal would be connected to whatever the K-line learns.

NB: Crucial point here -- frames are not primitives, because they are built out of subcomponents!

24.4 Default Assumptions

Optional details in frames are easily displaced by reality. But perception of reality requires these default assumptions. "But why use default assumptions at all, instead of simply seeing what's really there? Because unless we make assumptions, the world would simply make no sense. It would be as useless to perceive how things 'actually look' as it would be to watch the random dots on untuned television screens."

NB: How does this relate to natural constraints on vision? Does this agency approach treat bottom-up constraints and top-down expectations with essentially the same mechanism?

24.5 Nonverbal Reasoning

A general form of reasoning involves replacing particular things with typical things. Memory agents do this by 'moving memories around'. "Children must develop complex skills, not merely to replace one representation with another, but to compare the two representations and then move around inside them, making different changes at different levels. These intricate skills involve the use of isonomes that control the level-band of the activities inside our agencies." But we don't know much about how such processes work.

NB: Again, difference detection seems to be at the heart of this. Perhaps encoding differences, instead of "facts", is the key to knowledge representation.

24.6 Direction-Nemes

How do many agencies communicate about places and shapes? Minsky's hypothesis is that "many agencies inside our brains use frames whose terminals are controlled by interaction-square arrays. Only now we'll use those square arrays not to represent the interactions of different causes, but to describe the relations between closely related locations." These are agents called "direction-nemes".

NB: The basic idea here is that different agents are laid out in a grid, and the activation of one agent in the grid will represent a direction relative to the center of the grid. I wonder if this kind of scheme is practical, though. Lots of the work in visual cognition argues that there are many, many different spatial relations that could be computed -- too many to compute automatically. This is why Ullman and Pylyshyn appeal to a level of processing called visual cognition. Wouldn't it be better in Minsky's framework to let different agencies develop for computing these different relationships when required? (Or am I missing the point -- do the direction-nemes work automatically or not??)

24.7 Picture-Frames

"We have each accumulated enough room-frames to represent most rooms we're likely to see". But Minsky goes one step further than this, and proposes a generic frame that fits almost any room, with terminals corresponding to ceiling, floor, and walls. Then, each of these terminals is represented by a subframe that includes direction-nemes to identify different regions of the surface. The point of this is to use the direction-nemes to encode the spatial relations of objects that might be seen in different views.

But what about mistakes -- like misidentifying an object type? This kind of mistake won't be catastrophic if related frames share the same terminals. The frame representation can be easily revised, preserving the other correct representations that have been built.

NB: This approach is interesting, but it buys into a specific viewer-centered coordinate system. How might the agency approach be used to represent, say, the object-centered coordinate systems of the objects in the room? Would the same kind of direction-neme strategy work?

24.8 How Picture-Frames Work

Minsky modifies the trans-frame approach to create a picture-frame. In this system, the frame terminals are attached to a set of nine direction nemes instead of the pronomes in the trans-frame.

With K-lines attached to this kind of system, visual learning becomes possible. "Imagine that you're looking at some real-world scene. Your eyes move in various directions, controlled in some way by direction-nemes. Now suppose that every time you move your eyes, the same direction-nemes also activate the K-lines attached to the corresponding terminals of a certain vision-frame. Suppose, also, that those K-lines are ready to form new memories. Then each time you look in a different direction, your vision-system will describe what you see -- and the corresponding K-lines will record what you see when you look in that direction!"

NB: Cool idea hear is the idea of cross-modal associative memories. In other words, particular visual information will be recalled by the appropriate eye-movement cue. This makes me think of other possible functions served by parietal cortex. Instead of computing mappings between coordinate systems in different modalities, might it also be working on establishing associative links between these modalities?

24.9 Recognizers and Memorizers

The activation of a frame is equivalent to recognizing familiar situations or things. But how do frames become activated? "We'll simply assume that every frame is activated by some set of recognizers. We can regard a recognizer as a type of agent that, in a sense, is the opposite of a K-line -- since instead of arousing a certain state of mind, it has to recognize when a certain state of mind occurs."

NB: As Minsky notes later in this section, the big issue here is fleshing out the mechanisms for frame recognition. Does New Connectionism provide any insights into how this problem might be solved?

Notes on Chapter 25, "Frame-Arrays"

25.1 One Frame at a Time?

Why do we see only one interpretation at a time when viewing an ambiguous figure? Minsky says our agencies can tolerate only one interpretation at a time. Question: why?

To answer this question, Minsky turns to problems of vision: Why do we see objects as being composed of features? How are features grouped into objects? "Our vision-systems are born equipped, on each of several different levels, with some sort of 'locking-in' machinery that at every moment permits each 'part, at each level, to be assigned to one and only one 'whole' at the next level."

NB1: One possible way to view the machinery alluded to in the quote above is in terms of stable attractor states in a recurrent network.

NB2: Minsky will expand this hypothesis to a "locking in" of semantic interpretations.

NB3: Minsky is portraying vision as a hierarchy of agents solving problems of underdetermination a la Marr -- feature detection, object detection, object interpretation -- all constrained by a locking in of unique solutions. Must then the locking in mechanisms reflect natural constraints to be used to solve problems of underdetermination?

25.2 Frame-Arrays

How might the See agent work? The answer to this question is much more complicated than introspection would reveal -- you have to deal with object permanence issues and hidden surfaces, for example.

Minsky's approach to this problem is to use frame-arrays, in which different views of the details into the same structure. (NB: Cool thing here is that the object-centered "coordinate system" really becomes a concatenation of different viewer-centered perspectives!). "When you represent a thing's details with a frame-array, you can continue to move around yet 'keep in mind' all that you've observed from those different viewpoints, even though you've never seen them all at once. This gives us the wonderful ability to conceive of all of an object's different views as aspects of a single thing.

NB: From Pylyshyn's perspective, the role of a FINST would be to provide an index from a retinal cluster to terminal nodes of the frame array interpretation.

25.3 The Stationary World

Why does the world appear stable as we move through it? "This is because our higher-level agents don't 'see' the outputs of the sensors in our eyes at all. Instead, they 'watch' the states of middle-level agencies that don't change state so frequently.

Why then are "inner agencies" stable? Because they are implemented by frame arrays. This is accomplished in part by using directionnemes to select frames from frame arrays (e.g. to predict world-appearance as we move through it).

NB: The idea of internal stable reps (at least viewed in terms of time) is neat. Why might the visual system be sluggish, as in the visible persistence effect? Perhaps to build internal stability. Are early effects like visible persistence magnified by interior processes that construct intermediate representations?

25.4 The Sense of Continuity

Frame-arrays "let us 'visualize' imaginary scenes, such as what might happen when we move, because the frames for what we can expect to see are filled in automatically." Frame-arrays could be the heart of imagined visual scenes -- digital shifts underlying continuous changes of experience. We represent change as being continuous, and so experience it -- even if it is ultimately based on a digital representation.

NB: One very nice example of the continuous experience of digital information is the phenomenon of apparent motion.

25.5 Expectations

What does "expect" mean, particularly in vision? For Minsky, expectation is the content of filled-in nodes of frame-arrays (the defaults) -- but qualifies this as an oversimplification.

NB: What does expect mean -- a neat question! One wonders the impact of this question on the debate of the cognitive impenetrability of vision. To what extent might top-down visual effects be due to wired in mechanisms, and thus only seem "cognitive"?

25.6 The Frame Idea

Critics of frames described the idea as being too vague. "In retrospect, it seems those explanations were at just the right level-bands of detail to meet the needs of that time, which is why the essay had the effect it did. If the theory had been any vaguer, it would have been ignored, but if it had been described in more detail, other scientists might have 'tested' it, instead of contributing their own ideas." But frame array idea (frame systems) did not gain popularity. Minsky points out that lots of interesting questions about frames still remain to be answered.

NB: I wonder if this book (SOM) will receive the same kind of evaluation from Minsky after a while -- criticisms that I've heard of the book are very similar to the criticisms of frames that Minsky discusses here.

Notes on Chapter 26 Language-Frames

26.1 Understanding Words

People are adept at drawing complicated inferences; far more adept than they realize. "We call these understandings 'common sense'. They are made so swiftly that they are often ready in our minds before a sentence is

complete! But how is this done?" Minsky views such inferences as a natural consequence of activating frames; frames that have default assignments. "Such knowledge comes from previous experience." Note: this kind of inferencing ability underscores, in my mind, why Fodor describes high-level processing as being non-modular. In drawing such inferences, anything goes, so the system cannot be informationally encapsulated.

26.2 Understanding Stories

Frames can be used to give an account of story understanding. New information in a story is attached to the best spot in frames which have already been activated. "What makes a story comprehensible? What gives it coherency? The secret lies in how each phrase and sentence stirs frames into activity or helps already active ones to fill their terminals." Note: here, I'd like to see a more explicit translation of this stuff into an agency architecture. What does the society of mind do to elaborate the notion of a frame?

26.3 Sentence-Frames

The activation of frames per it's a great deal of information to be communicated using a small number of words. Minsky introduces the notion of sentence frames to reflect the roles of various words in a sentence. Note: there are two main points in this material. First, this reinforces the previous point of language being used to control existing structure in another's mind -- this is another reason for this kind of efficiency. Second, the notion of Satan's frames reminds me a bit of Jackendoff's conceptual structure, which is both syntactic and semantic in flavor.

26.4 A Party-Frame

"Dictionary definitions never say enough." Frames, like the party frame described here, place an enormous amount of structure on events. "We take our social customs for granted, as though they were natural phenomena. Few guests or hosts will ever wonder why their parties have those explicit forms or ask about their origins. As far as to any child can tell, that's just how parties ought to go; they always did and always will. And so it is with almost everything we know." Note: this last point raises Whorfian-style questions in my mind. How much of the world can we really understand in the absence of such frames?

26.5 Story-Frames

Understanding stories requires listening skills, and a storyteller must work to focus listener's minds. That is, the storyteller must activate the appropriate frames. For example, the phrase "once upon a time" raises a bunch of "expectation frames". Note: again, the underlying theme here is the use of language to control or direct the mental processes of the listener.

26.6 Sentence and Nonsense

Minsky introduces the notion of a grammatical sentence being one that is consistent with a sentence frame. This raises the issue of how sentence frames are fit, which will be dealt with later. For now, it is assumed that this is a task that can be accomplished by the language agency. "A word-string seems 'grammatical' if all its words have hit quickly and easily into frames that connect suitably to one another." Once this is done, conflicts with other agencies can emerge, conflicts concerning the semantics of grammatical sentences. Question: if you did a quantitative analysis of English, how many different sentence frames might you find?

26.7 Frames for Nouns

As we add more and more adjectives to a sentence, we do this systematically -- in an expected order. "This suggests that we use the frame like structures for describing nouns as well as verbs --that is, for describing things as well as actions."

"Many scientists have asked, indeed, why so many human languages used similar structures such as nouns, adjectives, verbs, clauses, and sentences. It is likely that some of these reflect what is genetically built in to our language-agencies. But it seems to me even more likely that most of these nearly universal language-forms scarcely depend on language at all but reflect how descriptions are formed in other agencies." Note: this is similar to the role of conceptual structure in Jackendoff's Erie. And

26.8 Frames for Verbs

Now are used as signals to build sub-frames for clauses. "Using frames this way simplifies the job of learning to speak by reducing the number of different kinds of frames we have to learn." Note: might this be a solution to connectionist encoding problems?

26.9 Language and Vision

Is the use of frames and language unique? No, answers Minsky. "In vision, too, there must be similar processes involved in breaking seems apart and representing them as composed of objects and relationships." Note: again, see Jackendoff who makes explicit links

between language, vision, and conceptual structure.

26.10 Learning Language

Some ideas survive by being incorporated as words into the lexicon. "Yet it is no paradox to say that even as we inherit those ideas from our culture, we each must reinvent them for ourselves. We cannot learn meetings only by memorizing definitions: we must also 'understand' them." We must also learn grammar for using words. "Some language theorists have suggested that children learn to use grammar so readily that our brains must be born with built-in grammar-machinery. However, we've seen that our visual systems solve many similar problems in even earlier years." For Minsky, the question is really why does language learning takes so long.

26.11 Grammar

Because we know so little, it is difficult to even speculate about the early stages of learning a language. Note: Minsky's speculation seems to require interactions between syntactic and semantic processes. This were minds me of Wexler and Cullicover's requirement not only for a universal basic grammar, but also for a language learner to grasp the meaning of incoming sentences.

26.12 Coherent Discourse

"Every discourse works on several scales. Each word you hear can change your state in a way that depends upon all the structures you have built while listening to the words that came before." For Minsky, there is no clear boundary between language and thinking. Note: this leads directly to the frame problem of artificial intelligence. "To understand what people say, we also exploit our vast stores of common knowledge, not only about how specific words are related to the subjects of concern, but also a how to express and discuss those subjects." Note: this strikes me as a natural consequence of viewing language as a system for control, and not as a system for transferring information.

Notes on Chapter 27 Censors and Jokes

27.1 Demons

Even basic things, like understanding pronouns, and are very complex. For the speaker, this means that the appropriate structures must be activated in the listener's mind. But how is appropriate information activated it real-time? As a student of Minsky, Charniak "suggested that when ever we hear about a particular event, specific recognition-agents are thereby aroused. These then proceed actively to watch and wait for other related types of events." For Minsky, this work raises a number of different questions, and indicates to him that understanding is a huge accumulation of skills.

27.2 Suppressors

Mistakes can be prevented by recognizing bad ideas early. "All communities in false some prohibitions and taboos to tell their members what they shouldn't do. That, too, must happen in our minds: we accumulate memories to tell ourselves what we shouldn't *think*." Minsky proposes new mechanisms for this. A suppressor-agent waits for a bad idea, and then makes you think of something else. Censor-agents tried to prevent bad ideas from even emerging by suppressing their precursors. Minsky thinks that modern psychology ignores these sorts of mechanisms, and that this is a serious mistake.

27.3 Censors

Censors are an attempt to increase the time efficiency of a system. The idea is to restrict many different possible states of bought from occurring. As a result, each censor may require a substantial amount of memory. This is a price to pay for using them. Note: first, censors strike me as being similar to natural constraints. Second, isn't the need for memory for a censor going to result in time problems because memory search will take time? Third, some of these ideas remind me of the different approaches to searching a problem space --the trade-off between using a heuristic and the cost of using the heuristic.

27.4 Exceptions to Logic

"Certainty seems always out of reach. This means that we have to take some risks to keep from being paralyzed by cowardice." As a result, Minsky thinks that we look for islands of consistency within which reasoning seems safe. We have to develop methods of thinking that realize that facts are seldom always true. Note: this is similar to arguments against the classical view of reasoning that tries to equate thinking with logic or probability.

27.5 Jokes

According to Freud, chokes our story is designed to fool the censors. But Freud's theory couldn't account for nonsense jokes. Minsky argues that the same mechanism works for both. "Once we recognize that ordinary thinking, too, requires censors to suppress ineffectual mental

processes, that all the different-seeming forms of jokes will seem more similar."

27.6 Humor and Censorship

Humor has a practical function in learning. This is because humor is involved with how our censors learn. Not only do we have to learn what states are desirable, but we also have to learn what states are undesirable. Minsky's theory is that humor activates mechanisms that create censors. Note: could something like an artificial neural network have two different types of learning?

27.7 Laughter

"The function of laughing is to disrupt another person's reasoning." The whole idea here is that the creation of censors is involved with the creation of processes that will disrupt trains of thought. That is, censor building must interrupt other thought processes.

27.8 Good Humor

"To ask how humor works in a groan-up person is to ask how everything works in a groan-up person, since humor gets involved with so many other things."

Notes on Chapter 28 the Mind and the World

28.1 The Myth of Mental Energy

The modern concept of the term energy needs something like "cellular fuel". But what makes it seem like we have mental energy? Minsky's view is that mental energy regulates the transactions between agencies.

28.2 Magnitude and Marketplace

We are often forced to choose between alternatives that we cannot compare. "Instead of comparing the items themselves, you could simply compare how much time they seem to be worth. We turn to use in quantities when we can't compare the qualities of things." So, a magnitude is assigned to each alternative. Agencies could do this as well, but what would they use as currency or magnitude? Minsky's hypothesis is that the courtesy is the "pleasure of success". Note: could this be related to work on magnitude assigned in perception?

28.3 Quantity and Quality

Quantitative descriptions conceal the structures that give rise to them because they are one-dimensional and featureless. Quantitative descriptions ignore qualitative differences. As a result, "we should never assume that the quality or character of their thought process depends directly on the nature of the circumstances that evil kid. There is no quality of 'sweetness' inherent in sugar itself."

28.4 Mind over Matter

Feelings are not inherent, they are engineered. They are engineered to anticipate problems before they do damage. "But what about the feelings of depression and discouragement?" They are similar to fatigue. They are designed to prevent the unproductive use of time.

28.5 The Mind and the World

We leave in physical, social, and mental realms. These all seem to obey different rules. What is the relationship between physical and mental? Minsky says there really isn't a mind-body problem, because minds are simply what brains do. Minds don't seem physical because a mind is always in transition from one state into another. Here, transitions reveal Minsky's version of functionalism.

28.6 Minds and Machines

Mine seen different from things because they aren't things. We have trouble classifying processes. Why? Because mind is a set of self-modifying processes. "The principal activities of brains are making changes in themselves." In short, mind is a complex process produced by a very complex machine, so it is no wonder minds and brains are hard to understand. Note: with this reminds me of a standard type of process that basically says when we have better measuring devices we'll have better theories.

28.7 Individual Identities

And what if all the physical components of an object are replaced by functionally equivalent components? Is it the same object? The issue here really is what we mean by "same". No: this is a standard version of an argument in favor of functionalism.

28.8 Overlapping Minds

There are many ways to draw in imaginary boundaries through brains. "If you agree that each person has both a last-brain mind and right-brain mind, then you must also agree that each person also has a front-brain mind and a back-brain mind!" The issue becomes does the bounded area have a mind? If it doesn't resemble us, then we are reluctant to say that it has a

mind. But brain locality does seem to exist indicating that small areas of the brain are functionally specific. "All this suggests that it can make sense to think that there exists, inside your brain, a society of different minds." Note: this move seems late. The brain locality argument could give strong support for Minsky's thesis. But I have a question: what is the difference between a society of minds and a society of functions?

Notes on Chapter 29 the Realms of Thought

29.1 The Realms of Thought

We view the world as being divided into different realms. We have a better understanding of things with in each realm that we have other differences between realms. A really big gap concerns the relationship between minds and brains.

29.2 Several Thoughts At Once

We can think in several mental realms at once. For example, we have words and their layers of meaning. "How can so many different thoughts proceed at the same time, without interfering with one another? I suspect that it is for the same reason that we have no trouble imagining an Apple as both round and read at the same time: in that case, the processes for color and shape used agents that do not compete." Note: are apt metaphors those that just the folk noncompeting agencies? Is cooperation between agencies important too?

29.3 Paranomes

Al Minsky moves to cooperative links. He starts by arguing that meanings in different realms can be very similar, and then argues that they can be so similar that there is no conflict. Additional agencies forge links between parallel frames in different realms. Note: this reminds me a lot of general thinking about how metaphors are understood, by having the meaning of the vehicle change the interpretation of the features of the topic.

29.4 Cross-Realm Correspondences

Metaphoric meetings of terms he panned on cross-realm correspondences. But these correspondences change process. "No sooner does a language agency assign some polynemes and isonomes to a phrase than various mind divisions proceed to alter a how they are used inside each different realm." Note: these sorts of moment by moment interactions would provide a good example of why this sort of system would have to be studied using simulation.

29.5 The Problem of Unity

Why isn't the mind unified? Because it isn't practical to do so. Different realms require different principles. We have to build separate systems to represent these different principles. Note: from all of what has come before, we get the sense of the system with complex cooperative and competitive interactions. Could this be an artificial neural network?

29.6 Autistic Children

"How do children start on the path toward distinguishing between psychological and physical relationships?" The key is to discover the different principles for each different realm. But we must also keep these different realms separate. Autism is a failure to do this. Note: another example of this is bilingualism. Are there any disorders in which the separation of language systems breaks down?

29.7 Likenesses and Analogies

"We always try to use old memories to recollect how we solve problems in the past. But nothing is ever twice the same, so recollections rarely match. Then we must force our memories to fit -- so we can see those different things as similar." How? We need to modify memory, or modify the representation of the current situation. This is easier to do if the agencies to be changed have weak attachments. Also, this is easier to do if it is easy to switch from one realm to another.

29.8 Metaphors

Our language is full of analogies. What is a metaphor? There are too many structural versions of metaphors to be unified under one out. Good metaphors transport intact uniframes from one well to another. Note: the pervasiveness of metaphor in ordinary thought and language is seen in Lakoff and Jackendoff's *Metaphors We Live By*.

Notes on Chapter 30 Mental Models

30.1 Knowing

What does "knowing" really mean? The answer to this question depends on who is saying the answer. If this is unknown, then default assumption will do the job. Even self-knowledge is relative. Note: this takes care of the "judge

criticism" of the Turing test. This is because the judge is a typical conversationalist.

30.2 Knowing and Believing

When we speak, we classified our thoughts into different types, including a variety of intentional terms. We can hold inconsistent beliefs under Minsky's multiple agent view. Why do we think that our beliefs are certain? At times, we must let particular agencies dominate if only for a while. Note: here, knowing and believing emerge from cooperation and competition amongst agencies.

30.3 Mental Models

Minsky says that knowing something must mean that we have a model of that something inside our head. Models are good to the extent that they are useful for answering questions. So a mental model is just a set of agencies that can be used to answer questions. Note: this approach show is the emphasis of control structure over format issues concerning representations. With the agency view, the key issue is what can be done with agencies, not what is there representational for.

30.4 World Models

A world model must include an additional component that stands for the entire world itself. As a result we wind up with meta-models. For instance, Mary can't have a model of the world, a model of herself, and a model of her model of the world, all in her brain.

30.5 Knowing Ourselves

If you ask someone to describe himself, you will get a whole bunch of answers to questions, and these can all be organized. These dancers should reveal two realms, body and mind. We cannot examine directly our self; we can only examine our model of our self. Can we have models of models of models? Maybe, but not to infinity. At some point we will lose track and the meta-model set up will fail.

30.6 Freedom of Will

Where does our sense of being in control come from? "Every action we perform stems from a host of processes inside our minds. We sometimes understand a few of them, but most life far beyond our ken." So, we talk of being free when faced with these unknowable processes. Note: you can compare this perspective to Brooks's statement about his view of his own components in *Fast, Cheap, And Out Of Control*.

30.7 The Myth of the Third Alternative

People postulate free will. Minsky seems to argue against a place for free will in science, but a need for free will in our views of ourselves. Free will is a central in our models of the mental realm. Note: does this mean that free will amounts to the control structure for societies of agents?

30.8 Intelligence and Resourcefulness

Intelligent stems from our vast diversity. Note: here diversity leads to redundancy and damage control, similar to standard connectionist arguments against classical artificial intelligence. Minsky is similar because his notion of control involves parallel processing.