Notes and reflections on "The emotion Machine" by Marvin Minsky

Draft version downloaded from http://web.media.mit.edu/~minsky/

Paul Krueger, Ph.D.

This paper presents short notes about the content of "The Emotion Machine" by Marvin Minsky. I'm very much impressed by this book and it has helped me codify some of my own ideas. I found it very useful to record a synopsis of Minsky's ideas as well as my own reactions to those ideas as I was reading the book. I started reading in in November, 2008. In 2009 I transcribed my journal into electronic form and this is that transcription. My references to chapter numbers refer to the publicly available draft version of the book and may not be accurate references to the published version.

I'll differentiate my thoughts from Minsky's by prefacing his with "M:" and mine with "K:" wherever there might be confusion.

The text of the journal starts now ...

11/15/2008

K: Reading the first few sections, this really resonates with ideas I've had about general purpose reasoning. Minsky provides a nice structure within which I can fit my ideas about representation using fuzzy conceptual graphs and an overall control structure based on something similar to catastrophe theory.

Minsky asks questions that are great. Some of them I think I can address with my research and some help me to expand the scope of my thinking.

I'm going to start over with the book and record my thoughts about each part here.

Chapter 0

M: "We also do a remarkable thing that no other creatures seem able to do: whenever our usual ways to think fail, we can start to think about our thoughts themselves – and if this 'reflective thinking' shows where we went wrong, that can help us to invent new and more powerful ways to think."

K: This is exactly my notion about self-awareness and the idea that the brain monitors its own activity in the same way that it monitors the external world. We have a sixth sense that permits us to "see" what is going on and react to progress or lack thereof in appropriate ways.

Minsky describes emotions as processes which constrain how we think. I might describe it more as providing a focus of attention (both on external and internal stimuli), but I think we essentially agree on what happens: some "resources" are activated and others are suppressed.

Minsky implicitly accepts the notion that the brain monitors itself:
"For example, whenever a problem *seems hard to you*, then your mind will start to switch among different ways to think ..."
My idea is that we need to make that self-perception more explicit, both in terms of what is perceived and then how we react to that sensory input to alter how we think. And I'll argue that the best way to represent all that is using fuzzy conceptual relationships which can approximate control surfaces which may at times result in catastrophic jumps to entirely new strategies.

I need to expand this notion with simple examples. Fuzzy rules which collectively guide inference steps smoothly and continuously under normal situations, but which cause a catastrophic jump to a new control surface when a dead-end is reached.

Discussion about the involvement of the physical body in emotional states ...
It strikes me that there are times when we deliberately act in ways that *cause* a change in emotional state in order to change our own mode of response to future events. You do this when you decide to follow a cheerleader; when athletes gear themselves up for a game. It's not only the case that reasoning affects our actions; we can also act in ways that affects our physiology and hence our future mode of thought. "Fake it 'til you make it." is a popular strategy for acting as if you were thinking in some way in order to trigger changes that lead to thinking that way in actuality.

Ah – I now see Minsky says almost the same thing: "... by regarding our body parts as resources that our brains can use to change (or maintain) their mental states!"

Chapter 1
1-2
On the complexity necessary for models of thought -
K: Einstein said something to the effect that theories should be as simple as possible, but no simpler. I think that captures my feelings. Perhaps theories and models of thought need to be complex because they encompass too many layers of abstraction today. If we tried to merge all physical science from quantum physics to biology into one big theory, it would be pretty complex. Maybe the apparent complexity of theories about reasoning should tell us something about how we're missing a few layers of abstraction.

1-3 Moods & Emotions

1-4 Infant Emotions
M: Sudden changes in behavior are common. Think of this as an artifact of the simple pre-existing control structures that dominate infant reasoning.

K: There are few behaviors that are available, so catastrophic jumps within the control space are common. As we add nuance to the reasoning process we begin to avoid such jumps. The control space is more continuous and transitions are less abrupt. Of course, even adults can "fly off the handle" when presented with a situation that drives them to a catastrophic state transition.

Minsky's "Rule-Based Reaction Machine" is better served with fuzzy rules. They allow for control functions that can be progressively refined over time, and fuzzy inference processes blend rules (that might somewhat conflict) in ways that are at least reminiscent of the ways that people behave.

Note to self: Check out "The study of Instinct" by Nikolaas Tinbergen.

1-5 Mind as a cloud of resources
K: M talks about primitive associations between emotions and behavior:
Emotions (Anger, threat, fear)
Actions (affection, alarm)

These ideas seem self-evidently true to me. If I introspect a bit on what's going on in my own mind when I say that, it's that there is a sort of harmony between what is said here and ideas that I've come to accept on my own. I can blend the ideas easily with little or no dissonance. I think that fundamentally the emotions of happiness or satisfaction are reflections of this sort of harmony. Whether that is harmony between our goals and our current states or between our thoughts and someone else's is immaterial. I expect this them to recur later in the book. Dissonance is the result of the brain having no where to go with some information it has gathered. The result is invariably a catastrophic jump in control space to a new mode of thinking/ behavior.

If, for whatever reason, we are compelled to stay in one mode of thought, even in the presence of large amounts of dissonance, then we experience stress.

People seek out a philosophy that lets them deal with their environments with a minimum amount of dissonance. That may be religious beliefs for some, science for others, mythology of various sorts for some. Usually this helps us reconcile the disconnect between our goals and our perceived state

and thus reduces the dissonance.

1-5 Mind as a cloud of resources
K; This whole idea of achieving minimum cognitive dissonance is consistent with M's observation that "... if all our resources were active at once, then they would often get into conflicts."

M is vague at this point about how resources that compete "win out" over each other and become dominant. Perhaps that's clarified later in the book. For myself, I'd simply say that this choice is governed by another resource. It would be tempting to call this a meta-resource, but I would deliberately avoid that because I think the mechanisms are ll the same, it's just the subject of the reasoning that differs and there are probably not just two layers of abstraction, but many more. As M says, those strategies for what resources are applied are refined as we grow and learn what is effective. I would characterize this as the addition of more nuanced fuzzy associations that blend together to make a more subtle control surface on which we move.

1-6 Adult Emotions
M: Critic-Selector Rules differ from if-do rules "... because the resources called Critics can recognize, not just events in the external world, but problems or obstacles inside the mind."

K: This gets closer to what I had in mind. I'd describe it in terms of an additional sense mechanism that our brains have for sensing its own state rather than as a fundamentally different sort of process, but perhaps that's just a slight difference.

1-6(?) Levels of Mental procedures
M:
"Self-conscious Emotions
Self-reflective Thinking
Reflective Thinking
Deliberative Thinking
Learned Reactions
Instinctive Reactions"

K: It will be interesting to read more about these. It's not obvious to me that this sort of taxonomy is either complete or correct, but it doesn't seem obviously incomplete or wrong either; which I guess means that it's a good start.

1-7 Emotion Cascades
K: "Cascade" might be the appropriate word for what occurs here, but I would conjecture that a catastrophe, in the formal mathematical sense, might be a closer analog to what is going on, at least in some cases. There is a large discontinuous change in behavior triggered by a catastrophic jump on a high-level control surface which exerts influence on many resources.

I might conjecture that an extreme and dysfunctional form of such a catastrophe might manifest as manic-depression or schizophrenia, with their attendant large abrupt behavior changes.

1-8 Questions

M: How could machines understand what things mean?
K: I believe that when people say that machines can never truly understand that they are comparing their own sense of what is going on within their own mind to their understanding of how machines work and simply can't conceive of a machine that could adequately represent and sense its own state the way that humans can. I don't just know something, but I also know that I know it. When asked a question, not only do the appropriate information retrieval and inference resources activate, but also my internal brain activity sensors register that activity, allowing me to "know that I know".

Incidentally, this also allows one to know that "the answer is right on the tip of my tongue" even when I can't actually retrieve the answer. In effect, I know that I know something even when I can't produce any proof of that by coming up with the answer. That occurs because resources are

activated because of my history with the topic and I sense that activation. We differentiate this situation from one where nothing ever activates in response to a stimulus ("I know nothing about it") or from the situation where nothing initially activates, but can be made to activate with additional stimulus ("I had completely forgotten about that until you reminded me about ...").

M: Why have multiple models of selves?
K: I think the simple answer to this is that we could never make one big model that was completely consistent. Trying to do so would result in huge amounts of cognitive dissonance and we would spend all our time in high anxiety trying to resolve the discrepancies. Having multiple models that are kept separate allows us to stay sane. This is fundamentally the same reason why formal logic (of any order) isn't suitable as a model of reasoning. Any inconsistency would permit the deduction of any formula. We live in a world of inconsistency and hold many mutually contradictory propositions simultaneously. We set up control structures to keep them from activating simultaneously, which would cause dissonance that we would be forced to resolve. Most political campaigns these days are designed to point out inconsistencies the actions of opponents in order to raise concerns in the voting populace (to cause dissonance in their models of the candidate which makes it unpleasant to think about them).

M: How do we develop new goals and ideas? "Initially we copy from parents and others."
K: Yes! This is a very powerful idea. It will be interesting to see this expanded.

2-1 Playing with Mud

2-2 Attachments and Goals

M: Emotions impact learning
Accomplishment: learn process
Shame: Adjust goals (suppress them)
Pride: Enforce Goals

2-3 Imprimers
M: "An Imprimer is one of those persons to whom a child has become attached."
K: There are a host of problems that can occur if this impriming relationship goes bad in some way. A child may imprint on too many different people and struggle to please them all. A child may imprint on nobody and become a sociopath. An imprimer may just be a bad person in some way: malevolent, vindictive, dictatorial, or have any number of other character flaws that may be passed on to the child via the impriming process. There must be a whole field of psychology somewhere that deals with such pathology.

2-4 Attachment-Learning Elevates Goals
M: Bootstrap goal formation process by initially just copying from an imprimer. Over time learn imprimer's process for generating goals.

"Negative Experience: When a method fails one learns not to use that subgoal.
Positive Experience: When a method succeeds, one learns to use that subgoal
Aversion: When a stranger scolds, one learns to avoid such situations.
Attachment Censure: When an imprimer scolds, the child devalues her goal.
Attachment Praise: When an imprimer praises, the child elevates that goal."

Questions:
"To what should each new goal be attached?
When and how should it be aroused?
What kind of priority should it have?
How long to pursue it, before giving up?"

2-5 Learning and Pleasure
M: Pleasant feelings help you remember successes.
K: If I'm correct that pleasure results from a lack of conflict within the brain, then this brain harmony

should also be conducive to the formation of new memories. Conversely, perhaps this explains why trauma victims often can't remember the trauma or events leading up to it. Something about the brain's discordance inhibits memory formation.

M: Credit assignment – How do we learn relevant associations?

K: I think this is bound up with the idea of *causality* and explanation. And these are facilitated by our visualization processes. What is really learned by the 'picking up mud with a fork' example is that liquids flow around things and must be contained to prevent such flows. Spoons can be seen to be satisfactory containers while forks are not. Learning is more than the strengthening of associations; it is a strengthening of *relevant* causal associations.

2-6 Conscience, Values, and Self-Ideals
M: Children build internal models of imprimers over time. Lets them predict what goals an imprimer would set, even in their absence. This becomes the "conscience". "Exploring, explaining and learning are a child's most purposeful urges and goals."

K: And why should this be so? Having a causal explanation allows a child to predict what will happen. When predictions are accurate there is little dissonance between the brain's internal model of what is going on and what it senses. I believe that is the essence of happiness and the ultimate goal of us all. We can achieve happiness in multiple ways. We can learn to keep our minds quiet and avoid dissonance by not trying to predict anything at all; i.e. by taking a mystical, zen-like approach to life. We can adopt a single universal causal agent and ascribe everything to God's will. Or we can choose to adopt a scientific approach which seeks ever deeper levels of understanding. I think most scientists would recognize the feelings of unhappiness that accompany a lack of understanding and the happiness that results when the source of cognitive dissonance is removed by the acquisition of a model that is more explanatory.

We have models of many aspects of our lives: ourselves, our immediate environment, groups with which we associate, etc. To the degree that our model matches reality, we are happy. When our model is flawed, we experience dissonance and unhappiness.

Building accurate models is not our only goal. We have more basic needs that also generate goals. So even if we accurately model ourselves as out of work and in a state of poverty, that alone doesn't make us happy; we would still have dissonance generated by our inability to satisfy our other needs.

2-7 Attachments of Infants and Animals

M: We choose imprimers who respond quickly and intensely to us.

2-8 Who are our Imprimers?

M: Can change as we age
K: We can choose role models at some point. Then we model their behavior. We may create wholly fictitious models of behavior that is not visible to us. This accounts for people who assume that famous people are also smarter and better in other ways. Finding out that such a model is incorrect leads to dissonance and unhappiness.

We don't just accept some imprimers. Over time we may adopt our imprimer's imprimers. This is the basis of social organization and behavior.

2-9 Self-Models and Self-Consistency

M: "... we learn to represent things in extremely simple, yet useful ways."
Can imprint on organizations or other social groups. Cults can replace previous goals with new ones by isolating a person from previous imprimers and then replacing old values with new. "... will discuss some conditions in which a process could appear to have motives and purposes of its own."

K: A strong shared model sets the behavioral agenda for all who accept it. There is likely a pattern of reinforcement that goes on between members which causes the model to increase in importance and dominate their behavior. In some sense the model can take on a life of its own if it influences members to recruit new members. This isn't necessarily bad. Our social fabric in the US likely depends on our common acceptance of a democratic model and the notion of individual freedoms. That is a very strong model that we will fight to preserve.

2-10 Public Imprimers

M: Famous people become imprimers.
"... all our attachments are made to fictions; you never connect to an actual person, but only to the models you've made to represent your conceptions of them, no matter whether they're parent or friend or merely transient attractions."

K: Right. And I think we make our models to be as consistent as possible with other models that we hold. This reduces the cognitive dissonance that would result from impriming on people with conflicting values. Thus we're happier.

3-1 Being in Pain (11/19/08)
M: Pain causes a cascade of resource recruitment in the brain.

K: What I've been calling 'happiness' in my previous comments might better be called 'contentment'; as contrasted with the discontent that accompanies cognitive dissonance. I want to differentiate because 'happiness' connotes a more active brain state where specific resources are engaged to create some degree of euphoria. I believe Minsky is correct that pain and pleasure engage similar machinery.

3-2 Prolonged Pain leads to Cascades

M: Prolonged pain focues the mind on a single objective, namely removing the pain. "The primary function of Pain is to make one remove whatever may be causing it. To do this, though, it needs to disrupt most of one's other usual goals. Whenever this leads to a large-scale cascade, then we use words like 'suffering' to describe what remains of its victim's mind."

K: Mental anguish is caused by a dissonance between the actual state of our mind and our goal/ model for our mental state. This sort of reflexive evaluation of mental state continues to occur even when most of the brain's resources are co-opted to deal with the stimulus pain.

The Machinery of Suffering
M: Not well understood: physical vs. mental 'pain'
Are physical and mental pain the same?
They can have similar effects.

3-3 Feeling, Hurting, and Suffering

M: Feelings are complex, intricate processes. Define:
Pain: sensations from injuries
Hurting: How we describe pain's early effects
Suffering: states resulting from large-scale cascades
"... what we call 'feelings' are attempts (by various parts of our minds) to describe large-scale aspects of mental conditions."

3-4 Overriding Pain
M: Distract the brain from one pain by introducing another. focusing on a pain tends to increase its intensity.

Prolonged and Chronic Suffering

Chronic pain is an unpleasant artifact of normally evolved injury avoidance mechanisms.

Grief

3-5 Correctors, Suppressors, and Censors
M: Experts know both what to do and what NOT to do.
"Critics – each of which learns to recognize a certain particular kind of mistake."

Critics contradict the line of inference that might otherwise have proceeded.

K: This is a key indicator of how potentially contradictory conclusions are avoided or reconciled. We always (or almost always or often) have reason to conclude opposing or contradictory or inconsistent things. Minsky's critics are highly valued heuristic associations that are activated in appropriate contexts to enable or (more likely) disable sets of other associations that could be used to make inferences.

K: Making this interlocking inference process explicit and precise will be a major part of my research. This is also where catastrophe theory will enter into the picture. The value of some functions define what sets of functions can be used to make further inferences. It's as if each association has a structure something like the following:
<context> <condition set> <conclusion set>
where a conclusion set might be the context for other inferences. You could possibly merge the conditions and context into one set of pre-conditions, but I expect that associations between the condition set and conclusion set might in fact be more like equality relations or associations than like implications; i.e. they work both ways. It's more like the rule says "In this context, these things co-occur."

Minsky talks about critics that pay attention to the internal state of the brain and I couldn't agree more. This is what I wrote about in my paper about self-awareness. We do have a privileged position that permits us to sense the internal state of our own brain and react to what we sensed.

Excessive Switching

M: Bi-polar disorder results from too many critics firing. When facing a new problem you might: 1) shut off most critics to gen options, 2) turn on critics to narrow options, 3) select one and pursue until critics complain about lack of progress. Minsky conjectures we do this on short time-scales, unaware of what we're doing.

Learning from Failure
M: We learn from failure as well as from success. Learning processes that only reward success and never punish failure may not work as well. "... we may need to endure some suffering to make larger-scale changes in how we think."

"Reinforcement can lead to rigidity."
Rules become too strong with constant reinforcement.

"Dependency leads to side effects."
Overgeneralizing the context in which a rule is applied can cause problems if the rule is later modified.

"Negative Expertise"
Negative associations can more precisely define the context within which a rule applies. Only failure can help to identify necessary conditions that need to be added to the critic.

"Radical Learning"
Positive reinforcement only lets you make incremental changes; for really big changes you may need major discomfort and disappointment.

"Papert's Principle"
When two or more methods fail, don't look for a compromise or blend them; go find something new.

M: Need to know what failed and <u>WHY</u>

M: Varieties of Negative Expertise
Creativity - Just enough new to be useful
Decisiveness - Results when we turn off critics that were comparing alternatives
Pleasure - Can shut off other processes. Feels good, but precludes other processes.
(K: Too much of a good thing!)
Parenting - Requires people to give much and give up much.
Beauty - Liking comes from critic suppression
Mystical Experience - Turn off all critics
(K: suspend disbelief)

3-6 The Freudian Sandwich
Freud recognized that the mind is a battleground for conflicting forces. Freud's "sublimation" is couching inputs in ways that don't trigger censors or critics. AKA rationalization.

3-7 Controlling our Moods and Dispositions
M: No critics ≡ Euphoria
K: Or equivalently, nothing for critics to react to. This lack of dissonance is what I called contentment previously.
M: Too many critics, you would see imperfections everywhere.

3-8 Emotional Exploitation
M: Fantasies can be used to trigger emotional states to support a desired goal.
Why does this work?
M: Because we constantly use imagination to fill in reality that we can't directly perceive. So we can fool ourselves into accepting an alternate reality which triggers the desired emotional response.

Competing emotions are useful. No one can always dominate. Need to resort to tricks (imagination) to change the weight in favor of a desired emotion over a 'natural' one.

Part IV Consciousness

4-1 What is the nature of consciousness?
M: No one single thing - lots of things

4-2 Unpacking the Suitcase of Consciousness
Great example and enumeration of actions that might be part of consciousness.

4-2.1 Suitcase words in Psychology
K: I finally found something to disagree with:
M: "... no part of a mind can 'see' much of what the rest of the mind does."
K: On the contrary, Minsky has already talked about critics that react to what is going on within the mind. That's only possible if there is some sensor network that lets part of the brain know what other parts are doing. This "self-awareness" is perhaps the most critical aspect of consciousness for me.

4-3 How do we recognize Consciousness?
M: Consciousness seems to be a serial process. Probably because the resources required are limited and can't do two things at once.

4-3.1 The Immanence Illusion
M: "For most of the questions you would otherwise ask, some answers will have already arrived before the higher levels of the mind have had enough time to ask for them."
Data gets to you before you recognize the need for it, so it feels instantaneous.

4-4 Over-rating Consciousness
M: Consciousness isn't a means for the brain to observe itself, but rather a way to deal with high-level problems.
K: Minsky wants to define consciousness as restricted to the solution of "practical problems". But clearly I am conscious of my own mood and many facets of internal brain state. I am conscious of my

own internal deliberations and can reason about whether I am making progress or not. I don't see any reason to restrict consciousness to external events.

4-5 Self-Models and Self-Consciousness
M: We have many models of self

4-6 The Cartesian Theater
M: Lots of active self-models inside us.
K: When I think about what my conscious thoughts are, they are the ones that I verbalize to myself. This being able to "hear" myself form statements is critical to what I think of as conscious thoughts. Is it possible that the vocalization/hearing mechanisms are the only way that disparate parts (models or resources) can communicate with each other? That consciousness evolved to facilitate coordinated activity between multiple resources? That would make the use of language critical to the ability to be conscious. It (language) may be necessary, but may not be sufficient for consciousness.

K: One of the problems for developing AI is the combination of facts - how do we pick which things to juxtapose to form new knowledge? We can't fully correlate all the facts we have because there would be a combinatorial explosion of facts, most of which are useless and irrelevant.

K: But suppose that when resource (A) is activated and produces some data which is also the subject matter of some other resource (B) operating simultaneously. Then through the conscious ability to hear what resource A produces, that becomes input to resource B and can be combined with resource B's internal data. B effectively learns from A because both were active at the same time and B could "listen" to what A had to say. Note that B doesn't absorb all the state of A, it just incorporates the relevant conclusion that A made. There isn't a complete unification of all facts, just a separate copy of a critical conclusion.

K: This model has properties that are similar to Hebbian learning in neural nets, where co-activation causes an association to be strengthened. But here it is co-activation that permits one resource to copy some information from another and combine it with its own. This is something like "unification facilitated by consciousness".

M: "... will raise serious questions about he extent to which our minds depend on a centralized workspace or bulletin board. We'll conclude that the idea of a 'cognitive marketplace' is a good way to start to think about thinking, but that when we look more closely we'll see the need for a great deal more architectural structure."

K: Yep. As we associate disparate things via conscious activation we learn associations that can activate in the future without conscious facilitation. This is consistent with learned expertise where we initially have to think hard about how to do something which become 'sub-conscious' over time. We learn so well that we may no longer be able to explain how or why we made the association because it's no longer something that we do consciously.

4-7 The Serial Stream of Consciousness
M: Basically we can only have one stream of conscious thoughts
K: Even those with multiple personality disorders only activate one at a time.

4-8 The Mystery of 'Experience'
M: Consciousness is an emergent property and we have to look at lower levels to learn more.

M: "... will argue that it is a mistake to suppose that you are 'aware of yourself' - except in a very coarse everyday sense."

K: Minsky argues that we use models of ourselves rather than sensing ourselves. It seems to me that we do both. It may be that most forms of self-awareness occur sub-consciously, but some can certainly rise to the level of consciousness on occasion. I would agree that most of what we directly sense is not the details of our reasoning, but more like sensing the level of activity of various

resources.

K: The strong statement that Minsky makes seems easily disprovable. I can think of a number and know that I'm thinking of it and no outside entity can have that same awareness. I do, however, think that there are limits on what we can be self-aware of, namely, we can be aware of some aspects of our physical state, including levels of brain activity, and we can be aware of things we verbalize to ourselves. The may also be non-verbal aspects of consciousness. I can imagine a color or a sound and see/hear it in my head.

4-9 A-Brains and B-Brains
M: A-Brain is reactive sensing of external world
B-Brain is the deliberative brain.

M: B only gets A's description of things. B can influence how A works/reacts.

M: C-brain which is reflective and thinks about how B is working.

M: 6 levels:
Instinctive reactions - from birth
Learned reactions
Deliberative thinking
Reflective thinking
Self-reflection "Why and how we could think about such things."
Self-consciously - "Whether we ought to have done those things."

M: Levels provide the right amount of organization, neither too few interactions between data nor too many.

K: I think the objective is correct, namely compartmentalizing knowledge and providing ways for the separate compartments to interact. I'm less satisfied with the layers notion. Rather than using Minsky's simple rooted tree, I would opt for a graph structure where there are sub-graphs with highly interconnected sub-nodes and with less connectivity between different sub-graphs.

K: The disadvantage of layers is that they are fixed and immutable. I suppose it's possible that we evolved separate physical mechanisms that are suitable for each layer. It seems more plausible to me that we evolved a single mechanism that lets us create these complex graph relationships over time. It may be that if we took one of my graphs and analyzed it we would find a structure that is similar to Minsky's. Our human experience may dictate an evolution of the brain that falls into that pattern.

M: Argues for an evolved layered structure.
K: Certainly there must be some physical layering and specialized abilities. I'm not yet convinced that they correlate to Minsky's set. Until I see evidence to the contrary I'm inclined to think that some of M's layers are physically co-located and result from differentiated connectivity within that region.

Part V Levels of Mental Activities

K: Minsky has his taxonomy of thought processes (6 layers).

K: Questions:
Are there processes I can imagine that don't fit well within M's taxonomy?
Are there processes that seem to cross or blur the lines between layers?

5.1 Instinctive Reactions
K: Seems reasonable -- all living creatures possess some forms of instinctive reactions, even for quite complex behaviors like mating rituals.

K: Need more than If-Do rules. Need "If <situation> Do <action> Then <result>. Allows us to project

what would happen next. Can iterate into the future to form more elaborate plans.

5-2 Learned Reactions

M: 20th century psychology says random actions with positive outcomes are reinforced so they are more likely in the future. Not too good at explaining how we learn complex sequences of actions.
Need to answer:
How were successful reactions produced?
Which aspects of recent events should be remembered?

Need to talk about 'thinking'

5-3 Deliberation
M: Decisions require ability to predict outcomes
K: For entirely novel or very complex decisions this may be true. I suspect that many choices are compiled into heuristics.

M: Postulates rules of the form:
IF <situation> DO <action> THEN <result>

M: If multiple rules are applicable, outcomes can be compared experimentally to choose the more attractive alternative. We believe that we 'envision' scenes, but this is mostly illusory.

Planning and Search
M: Linking rules in chains allows prediction further ahead. Exponentially explodes unless you create stepping stones to constrain search.

Reason and Reliability
M: When reasoning fails, we think about what happened 'reflectively'.

5-4 Reflective Thinking
M: "... resources at each level make descriptions of what some lower-level ones recently did." A variety of partial self-models are employed for reflective thinking.

5-5 Self-Reflection
M: "The reflective systems we just described can think about some of their recent deliberations. *Self-reflection* does just a little more: it considers not only its recent thoughts, but also the *entity* that had those thoughts."

"To do this, she must use some self-representations – models she's made for describing herself."

".. it is when our usual systems fail that reflective thinking gets engaged."

5-6 Self-Conscious Reflection
"... that enables us to think about our 'higher' values and ideals."

K: I'm having a tough time making M's three levels of reflective activity really distinct. Compare, for example, to deliberative thinking. We could create types of deliberative reasoning that correspond to the types of reflective thinking that M postulates. Why does one type warrant the extra levels of abstraction and the other doesn't? For that matter, why exactly should we make thinking about models of ourselves and models of others into distinct types of thinking? Couldn't we have just as easily created a taxonomy that distinguished between thinking about what entities do and thinking about what goals they have and make that independent of whether we're thinking about "other" vs. "self"? Clearly we have more data about what our 'self' is doing/thinking than about what some 'other' is doing/thinking, but why is that more than just a matter of degree? Certainly we care more about ourself and can make direct changes, so we devote more resources to ourselves, but is there really a qualitative difference? M asks similar questions of himself and falls back on saying that models which are too simple don't work. For that argument to be ultimately acceptable he would

have to demonstrate some ability that is gained by having these levels in the model that is not possible without it. Perhaps later in the book ...

5-7 Imagination
M: Difficulty of vision systems. Needed hierarchical recognition system but even that didn't work well because info flow was in only one direction.

K: I think this is the correct diagnosis of the problem. Look at the capability of Bidirectional Associative Memories (BAMs) (see Kosko's book). Both high and low levels contribute by feeding back to each other. Going with my dissonance-driven idea, I'd say that each level of the hierarchy settles into a harmonious, low-energy state that is most consistent with adjacent levels of abstraction. And actually I'd resist the notion of strict levels and instead talk about a network of associations that mutually affect each other.

K: Research direction: Adapt the techniques of BAMs to networks and work on the process of acquiring network topologies to support better recognition. Sense when BAM computation is not converging and jump catastrophically.

K: In the overall network, which of the many possible inference rules are actually used? This set is defined by the current context. What we sense activates some rules (or at least makes them more prone to activation). Our current goals or focus of attention makes higher-level resources more prone to activation. In some sense this activation energy flows through the inference network in all directions and when waves collide they reinforce each other and we connect the chain. Various sorts of heuristic rules may take high and low-level data as input and make various intermediate resources more prone to activation, providing the intermediate islands needed to connect up the inference chain.

K: A network settles into a relaxed low-energy state until perturbed. Then it adjusts as efficiently as possible to the new situation. Certain types of visual recognition phenomena bear this out. In a series of pictures that morphs from one image to another, the point at which the pattern recognized switches is different in each direction (i.e. scanning from image A to B has a different switchover point than scanning from image B to A). This is a hysteresis phenomenon which is characteristic of functions that define surfaces with catastrophic transitions. In some sense we hold onto what we previously recognized until it is just overwhelmed by contradictory data that cause a perceptual catastrophe. I'm betting that we can construct networks of fuzzy associations that will exhibit this same behavior.

5-8 The Concept of a "Simulus"

M: Simulus is a high-level abstraction of a scene. Can make significant changes efficiently by changing at the right level of abstraction.
Simulus ≡ simulate + stimulus.
"... a counterfeit perception caused by changing a mental representation."

5-9 Prediction Machines
M: Use If/Do/Then rules but suppress imagining the "Then" state. Iterate to plan several steps.
K: M ignores a bunch of complexity here. When we imagine, we don't necessarily shut off all external stimuli. And we "know" that we're imagining; we don't trick ourselves into believing the state is real. And we can imagine multiple alternative future states and compare them. We have the ability to generate multiple internal representations of the same situation. What physical mechanisms (brain or computer) can support this?

Part VI Common Sense (notes taken 12/01/08)

6-1 What do we mean by Common Sense?
M: "... things that we expect other people to know and regard as obvious."

M: The Telephone Call

The Concept of a 'Panalogy' ==> "Parallel Analogy"
"... special machinery that links corresponding aspects of each view to the same 'role' or 'slot' in a larger-scale structure that is shared across several different realms."

M: Sub-realms of the Telephone World: Spatial, Dominion, Procedural, Social, economic, conversational language, sensory & motor, kinesthetic, tactile, haptic, cognitive, self-knowledge

6-2 Commonsense knowledge & reasoning
M: Lots of facts required to understand ordinary conversation

M: How much does a typical person know?
– thousands of words each linked to a thousand others
– hundreds of uses and properties of thousands of different common objects
– tens of people
– hundreds of things about hundreds of people
– tens of items about 1000 people
∴ ~ 1 million things in each realm
~ a few dozen realms
∴ a few dozen millions of items of knowledge

M: Could we build a baby machine?
M: Many previous programs failed because "... those programs failed to develop a *good new way to represent knowledge.*

M: Must solve the credit assignment problem to avoid getting bogged down with irrelevant information.

M: Problems such systems face:
"The Optimization Paradox: The better a system already works. The more likely each change will make it worse."

"The Investment Principle: The better a certain process works, the more we will tend to rely on it, and the less likely we will be inclined to develop new alternatives."

"The Parallel Processing Paradox: The more that the parts of a system interact, the more likely each change will have serious side effects."

M: "... for a machine to keep developing, it must have ways to protect itself against changes with too many side effects."

M: Split system into parts that evolve separately.

K: Along with this, make reasoning itself be partitioned and restrict the interaction. We reason within fairly narrow contexts and switch contexts if necessary. Merging contexts is a much more difficult process because it requires reconciling inconsistencies. I suspect that we switch so quickly that we often don't even notice that we did so. That may result in us asserting contradictory or inconsistent conclusions from each context that may go unnoticed. Only when someone else points them out do we worry about reconciling them.

K: Minsky's levels are a way of partitioning, but I think this is still too structured. I suspect there are many more independent contexts. Reconciling separate, inconsistent context models is often the work of science or philosophy or any rational thought process. There is more predictive power in more complex models of larger scope, but getting to such models can be challenging.

M: Descriptions need to be abstract to handle different situations
K: Abstraction is critical to the process of combining separate models. You have to find points of correspondence and without abstraction that would be hopeless.

K: We probably don't lose one specific context when we combine it into a more comprehensive one. It may still have specificity that was abstracted away that is useful within the specific context in the future.

M: Remembering

M: Recalling relevant memories requires a lot of machinery.
M supposes that retrieval is relatively fast due to pre-existing usage links that are created as part of the process of memorizing something.

K: I'm not so sure I buy this. There are just too many possible usages for a memory to expect all usage links to be pre-compiled. I suppose the alternative is some form of associative search to find relevant memories. I'm personally more inclined to this direction.

6-3 Intentions and Goals

M: A powerful goal can push aside other goals. Some resources can make choices in ways that the rest of the mind can't control.

K: This is consistent with my model. There can be many conflicting conclusions drawn by different resources. Only when they are roughly balanced does a conflict occur. Normally we settle on one or the other and suppress the loser.

Difference Engines
M: We can define "'having a goal' to mean that a difference engine is actively working to remove those differences."

K: 'Difference' = 'Dissonance'. A conflict between current state and a goal state causes a jump on a control surface to a point where such differences can be resolved. This notion really needs more formalism and some simple examples to demonstrate the idea.

Goals and Subgoals
M: "... every difference it needs to reduces becomes another subgoal for it!"

K: If we have a process for achieving a goal, that defines a control surface along which we can move smoothly, as long as all pre-conditions are in place. If not, we jump to another surface defined by a process to achieve the pre-condition. *This is too much hand-waiving, I need to define how multiple rules/associations which define different control surfaces fit together. It's something like using the most highly activated surface until there is no further you can go with it, at which point the next most active surface takes over.*

6-4 A World of Differences

M: Our sensors react mainly to changing stimuli.

Rhythmic and Musical Differences (Notes taken 12/02/08)

M: Why is music so ubiquitous? Possibly to provide "... 'virtual worlds' for refining difference-detectors that we can use for condensing more complex events (in other realms) into more orderly story-like scripts."

Difference Networks

M: Postulates networks of concepts such that you could find something similar to what you want by partially matching some close alternative (e.g. looking for 'chair', match 'bench').

K: How might we implement this? Imagine that we activate the 'chair' concept which activates the various sub-concepts used to define it. They, in turn, partially activate other concepts where they

play a part. So activating 'chair' activates 'leg' and 'seat' which are also sub-concepts of 'bench' and therefore 'bench' partially activates. But if we find a real chair, it activates more fully and suppresses 'bench'. In the absence of finding a 'chair', 'bench' might be the most activated concept and serve as an acceptable substitute. This would suggest that similar concepts should mutually suppress each other to allow the network to make a clear choice.

K: Humans can also recognize a chair and a bench sitting side-by-side and if asked could imagine others in the same scene. So a change of focus (spatially, temporally, meta-physically) allows different items to be correctly recognized. An AI must be able to build up composite concepts from smaller ones and keep everything sorted out.

K: M wants to link concepts based on differences. I'm guessing that there may be more similarity links than difference links, but probably both are useful.

6-5 Making Decisions

M: Free Will is a process that stops other deliberative processes so that the most positive choice can be made at that moment.

6-6 Reasoning by Analogy

M: Analogy is a common process. Find a similar situation; find a process that worked there; adjust based on situational differences.

M: Geometric Analogy Program
Solves geometric similarity puzzles of the form "A is to B as C is to ...".
Shows value of exploring alternative descriptions of situations.

6-7 Knowledge needs Multiple Representations
M: Humans have diverse forms of representation. No one is always best.

K: Minsky seems to conflate the formal representational mechanism with the various models that are represented using that mechanism. The advantages that he describes come from using different models. The underlying mechanism used to implement each model could be any number of things that are formally equivalent. It is likely that some models are better suited to some mechanisms, but it's not clear how important that is. I still like the notion of using something like conceptual graphs that express fuzzy relationships. That could be used to represent any number of different models.

Part VII Thinking

M: Complex modes of thinking replace previous simpler ones. Sometimes we can no longer even access previous ones.

7-1 What selects the subjects we thing about?
M: Critics

7-2 Emotional Thinking
M: Obstacles interrupt orderly process and you have to find something else to do. May lead to a cascade of changes. M enumerates various ways that we can trigger changes in thinking. Some goals are very short-term, some can last a lifetime (called characteristics or traits). We describe large-scale cascades as changes of emotional state.

7-3 The Critic-Selector Model of Mind
M: Some changes in how we think are almost imperceptible to us. Some are made only after diagnosing the problem and deliberating about how to best correct it.

K: I would probably argue that the activation of a critic happens when the current mode of thinking fails in some fashion. In some sense the critic is always partially activated, but as long as the current

mode of thinking is more activated, the critic is suppressed. Dissonance can cause the current mode of thinking to be suppressed, which makes the critic relatively more activated. So it takes over and may result in further suppression of the previously active mode of thought (so it doesn't get immediately reactivated) and the increased activation of some alternative. Some critics can do all of this 'sub-consciously' and some result in more conscious awareness of what's going on.

M: selection of critics can be controlled by higher-level critics.

K: 'Higher-level' critics are just other associations that increase or suppress other critics. There could be many forms of these. Some might be very situation-specific. Some may activate in response to internal sensors that tell something about he progress being made. But again, these critics only fully activate when the currently active mode is suppressed for some reason.

7-4 What are some useful "Ways to Think?"
M:
Knowing How
Extensive Search
Reasoning by Analogy
Divide and Conquer
Planning
Simplification
Elevation (make concepts more concrete)
Reformulation
Self-Reflection
Contradiction
Use External Representations (e.g. documents)
Simulation
Correlation
Logical Reasoning
Wishful Thinking
Impersonation
Cry for Help
Resignation

7-5 What are some useful critics?

M: Innate reactions & built-in alarms
K: These are things that raise the activation level of some resource to the point that it dominates thinking. This is through no fault of whatever process was previously the most active.
M: Learned reactive critics
K: Basic behavioral rules. These have the effect of suppressing those innate, infantile reactions.
M: Deliberative critics
K: More complex combinations of associations that result in relatively convoluted control surfaces. It normally takes some time to traverse the surface to a solution; which makes the process feel more deliberative than reactive.
M: Reflective critics
K: I think these are an illusion. Failure results in negative feedback which suppresses the current mode of thinking, causing a catastrophic jump to a different process. That process may itself be just a deliberative critic that requires some time to activate an alternative mode of thinking.
M: Self-reflective critics
K: Like any process, deliberative critics can fail to make progress under some conditions. As this causes them to be suppressed, other forms of deliberation become active and may result in a different set of things being activated.
M: Self-conscious critics
K: These are modes of thinking that affect our emotional state. I would guess that these cause physiological changes that in turn are sensed and cause the mode of thinking to be altered. So the effects of these critics on thinking are somewhat indirect.

7-6 Emotional Embodiment (Notes taken 12/03/08)
M: Quotes William James suggesting that "... the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur *is* the emotion."

M: Physical manifestations could provide a sort of feedback loop that helps to maintain the emotional state within the brain.
K: This idea is similar to what I said previously about the indirect effects of emotions leading to change of brain state. Clearly we sense what is going on within the body and this influences our mode of thinking.

M: Brain exploits body as a dependable external memory device.
Aristotle: "The emotions are all those feelings that so change men as to affect their judgments, and that are also attended by pain or pleasure. – Rhetoric, Book II"

7-7 Poincare's Unconscious Processes
M: Poincare's stages of discovery.
"Preparation: Activate resources to deal with this particular type of problem.
Incubation: Generate many potential solutions.
Revelation: Recognize a promising one.
Evaluation: Verify that it actually works."

M: Blind trial and error won't often suffice. You need to activate resources that will generate good results.

M: Revelation requires the ability to recognize a good solution. Symmetry, consistency, harmony are words that M uses.
K: Early stages of this process induce more mental dissonance. Only by forcing this to be explicit, do we create the pre-conditions necessary to find a solution. The brain is so good at minimizing dissonance that without forcing it upon ourselves we would likely act immediately to minimize it.

K: Sidebar: This is why our current educational theories that stress positive reinforcement and minimal downside for failures are critically flawed. We learn over time to resolve dissonance with increasingly sophisticated methods. If all dissonance is studiously avoided we never learn how to resolve it. The human brain matures by failing and learning how to overcome that failure.

M: Creativity is not the ability to generate entirely new concepts or points of view. Must be a novel extension or combination of existing concepts.

Collaboration
M: Takes advantage of different areas of expertise to create something better than either could do independently.
K: I suspect M is just setting this up as an analogy for what goes on between different mental resources.

Do we normally think 'Bipolarly'?
M: If durations of Poincare's stages weren't controlled, practitioner might be seen as manic-depressive.

7-8 Cognitive Contexts
M: Returning to an interrupted thought process requires that records of that context be kept somewhere ad retrieved on demand. "Cognitive Context"
M: We need to have several "cognitive contexts" that we can refer back to.
K: I would totally agree and add that we need different contexts for each point in time. And we need multiple physical/spatial contexts. This likely means that a context is not one big structure, but rather a composite of "pointers" to sub-contexts. In a network one could imagine activation links between a context object ad a set of sub-contexts. This means that a brain must be able to dynamically create structures that can activate arbitrary other brain resources. Since we don't grow new neurons quite that dynamically, we might posit mechanisms for creating activation paths between specialized

'types' of contexts. For example, visual contexts might be one type. Assume that the cerebral cortex is the locale for context neurons. Then we could posit the ability for any context neuron to activate its visual sub-context. Within the visual area of the brain we postulate all-to-all connections (may need a multi-layer network to implement such a crossbar) so that the visual sub-context can, in turn, activate the corresponding visual record.

K: Note that we could imagine varying amounts of detail available in sub-contexts depending on our focus at the time they are created. If we see a painting, we may remember it vaguely, but be unable to provide much detail unless we choose to "memorize" some of it. People with "photographic" memory likely cannot control the amount of detail in their sub-contexts and remember everything (or at least much more than the rest of us.)

M: Likely that short-term memories are copied to a separate long-term facility. Takes hours to do this. Can sometimes retrieve mostly lost memories from fragments.
K: Memory retrieval must be some form of associative process whereby the activation of enough components of the memory triggers its activation. With enough similar memories that are more easily activated, we can 'lose' a memory. But by activating more fragments and consciously suppressing falsely triggered memories, we may succeed in activating the target. Refer to Kosko's BAM for details on iterative restoration of partially recalled patterns.

How many thoughts can you think at once?
M: Hundreds at lower-levels: reactive. Only a few at high levels.
K: I'd contend that the issue is mostly resource contention. You need to arbitrate when a resource is demanded by multiple contexts. That requires a higher level of control. But as long as resource utilization is independent, they can operate simultaneously.
M: says almost the same thing ...

What Controls the Persistence of Processes?
M: Need a way to control which critics are active. Some extremes:
– If your set doesn't change: one track mind
– If some are on all the time: obsession
– All turned off: No questions to answer and no problems
– Too many active at once: depression - flaws everywhere
– Too many turned off: mentally dull (K: autistic?)

K: i don't think it's a matter of whether critics are on or off. It's more a matter of exactly how activated they are that determines whether they dictate our actions. That level of activation is partially determined by other resources, so the net effect might be similar to what M suggests.

8-1 Resourcefulness

M: Humans have lots of ways to model things.

8-2 Panalogy

M: Will describe machinery for switching quickly between resources

M: Why can't a person hold multiple possible meanings of a phrase simultaneously? M contends this is because they must be contending for some limited resource.
K: I suspect something a bit different is occurring here. Trying to hold two conflicting interpretations simultaneously results in dissonance that our brains work to resolve. We relax into the most plausible state. That doesn't mean that the alternative is totally inactive, just that it is relatively less active. If additional evidence presents we might change to the alternative. The mathematics of catastrophe theory likely describe this. The hysteresis effects account for the appearance of a reluctance to change a preliminary judgment until the evidence is overwhelming.

M: Postulates panalogies which are multiple models linked by matching similar concepts.
K: I think in terms of concept associations that arise naturally as we create new, larger composite

concepts. However we can identify analogous sub-concepts at some later date and create a concept that equates them.
M: Now says much the same thing: "Links" are made automatically when we learn new things.

8-4 How does Human Learning Work? (Notes taken 12/04/08)
M: Most important learning skill may be learning by being told.

How do we learn so rapidly?
M: Animals learn by multiple repetitions. One theory is that humans do too, but repeat mentally rather than physically as part of making a short-term memory persistent.
K: Repetition is a funny word to use for this process. Surely there must be a feedback process that guarantees fidelity between the short-term version of a memory and the corresponding long-term version. But I would liken this more to the iterative nature of the BAM algorithms than to some sort of enhancement by repetition of the copy process.

M: Transfers from short to long-term memory can take hours to a day. Likens short-term memory to computer cache.

M: Other reasons why it takes so long:
Retrieval: need to make relevant recall links
Location: need to find a place for the memory
Complexity: copying complex structures takes time

Learning by building "Knowledge Lines"
M: K-Line is a structure that activates some set of resources
K: This is very similar to the way I think about all mental states: namely a structure that can activate another set of resources.

M: When working on novel problems, start several K-Lines. Leads to conflicts, cascades, etc. Leads to a novel combination of resources that we can capture. That K-Line is then activated by new problems that have similar relevant attributes.

8-5 Credit Assignment
M: Without accurate credit assignment; if you gave credit to everything that happened to be active, you would need many repetitions with different environments before you'd give credit to the right thing. M suggests that we use higher-level self-reflective processes to decide how to give credit. Processes involved include:
– Choosing how to represent the situation will affect which future ones will seem similar
– Unless you select only what helped, you may learn too many irrelevant things
– Mental experiments to discover relevant features
– Need to connect relevant features

M: These reflective credit assignment processes can take lots of time and need more research.
K: Another technique we use, I believe, is a sort of retrospective simulation. We rehearse what we should have done and thereby change weights so that in future situations that are similar we will react differently. I think that language plays a big role in all this. We express what we should have done quite literally. We have very good language processing abilities that let us "hear" what we're saying to ourselves and turn that into associative memories that trigger when similar situations occur in the future. One of the benefits of using fuzzy conceptual graphs, as I intend to do, is that we can approximate any possible control surface as closely as desired just by adding more specific and clarifying statements. So instead of being constrained to the inflexible T/F world of standard conceptual graphs, we can get much closer to the shades of grey needed to provide intelligence.

Transfer of Learning to Other Realms
M: Learning how to learn is critical to higher intelligence. Good credit assignment s critical to the ability to transfer from one domain to another.
K: I agree with this and might qualify it to say that those who have a deeper understanding of the causal relationship between what they did and the outcome will, as a result of that understanding,

have generalized the association in just the right way to make it useful in other contexts. Perhaps just having the right taxonomy of contexts permits retrieval of the strategy in a new context. So if the original association is

P_A can be solved by R_B

and we know that P_A is an instance of P_C, then when problem P_D arises, if it is also known to be an instance of P_C then P_D activates P_C and P_C partially activates P_A which partially activates R_B. R_B may also be partially activated just because it shares concepts with P_D. So the two partial activations raise the activation level of R_B to the point where it takes over. If it solves the problem, then the presence of a solution may activate it even more, to the point where it reaches consciousness.

8-6 Creativity and Genius
M: Experts/Geniuses have some common traits:
– Highly proficient in their fields
– More self-confidence
– Persistent
– More ways to think about things
– Habitually think in novel ways
– Often reflect on their goals and ideals
– Better self control
– Reject popular myths and beliefs
– Tend to keep thinking more of the time
– Excel at explanations
– Better at credit assignment

K: Few people are "geniuses" because thinking in a truly new way requires the ability to work through considerable amounts of cognitive dissonance between the new idea and more conventional ideas. Somewhere along the line you must have learned that this is a necessary pre-requisite to resolve the dissonance between current ideas and the real world. This likely means that a prerequisite for genius is to first be aware of the failings of current modes of thought. If there weren't any, or they weren't recognized, then there wouldn't be any problem to drive the whole process. People who routinely juxtapose concepts that are not normally considered together are more likely to uncover situations with inconsistencies (dissonances) that they then feel compelled to resolve.

8-7 Memories and Representations (notes taken 12/05/08)
M: We don't record 'events' themselves, we record what impact the invent had on our mental state.
K: Seems to me that sometimes we do record sense data with fairly high fidelity. We can sing back a song we hear with the right timing, tonal quality, phrasing, etc. That doesn't mean that there isn't some form of representation going on, because clearly there is or we'd never make a mistake repeating it back. But the degree of fidelity can, with repetitive presentations, be extremely high.

Multiple Ways to Represent Knowledge
M: Structures researchers have used:
– Narrative scripts: sequence of events in time
– Semantic networks
–Trans-Frames: A pair of semantic networks representing before and after images of some action. Can link in chains.

M: Frames of object properties are convenient. Frame slots have default values
K: I imagine that M's frames are groups of co-activating concept nodes. Each concept corresponds to a frame 'slot' and a default value is just the value concept that is activated by the concept in the absence of any other more strongly activated value. I imagine these values might be fuzzy.

K (side note): While I'm thinking about it – I think the notion of confidence in some assertion corresponds to the strength of its activation. If it is strongly activated it will suppress contradictory beliefs which is just what we'd expect if we were confident of that fact. In frame parlance this would

correspond to the slot value in a child frame overriding the default value inherited from the corresponding slot in a parent.

M: Picture Frames – More complex frames where slots themselves are interconnected.
K: This is where the idea of frames starts to feel a little kludgy. A network, otoh, is well-suited for representing such relationships. I imagine building up progressively larger and more interconnected fuzzy conceptual graphs.

M: Frames for Including Additional Slots – Sort of a frame with slots containing before and after networks (ala trans-frames) with additional slots for annotations.
K: I think all of this would fall out naturally in a FCG (Fuzzy conceptual Graph). As you add sentences that say whatever it is that you want to say about the concept, the FCG is expanded accordingly. Concepts are composed of sub-concepts, but can be arguments in others. A hierarchy naturally arises. See Sowa's work on conceptual graphs for examples. What needs to be added to FCG's is a way to activate or suppress individual concepts as needed. Some sort of 'spreading activation' model may be appropriate.

Connectionist and Statistical Representations
M: Work on frame and similar architectures stopped in 80's in favor of simpler systems that could learn on their own. Acquiring millions of fragments of common sense seemed daunting. New methodologies, primarily numeric.
K: M lumps fuzzy logic in with neural nets and other fundamentally numeric representations. I think this is a mistake and a fundamental misunderstanding of what fuzzy concepts can provide. They can give a much better computational representation of linguistic statements and do not require losing whatever led to the conclusion. I will, of course, need to provide appropriate evidence to support this claim. A FCG would be very similar to the semantic networks that M seems to like. M's aversion to anything numerical could certainly pose problems for his systems because without them rules/ resources are either active or not, there could be no partial activation. It would be impossible to pick one conclusion over another even if one was strongly supported and another was weakly supported. It would be impossible to use a rule unless all preconditions were strictly met, even if it might be advantageous to extend its scope by relaxing some precondition. By rejecting "numbers" M is implicitly restricting computation to 1 and 0, true or false, with none of the approximate reasoning that fuzzy can provide and which is a hallmark of human reasoning.

Micronemes for Conceptual Knowledge
M: Choice of meaning for a word depends on context
K: Precisely – and it may not be immediately obvious what that context is. Over time we may refine our understanding of the context and change our mind about which interpretation should be used. In the absence of fuzzy numerics you must adopt one or the other at any given time with no ambiguity tolerated. If you introduce confidence factors then you are adopting a cruder version of fuzzy numerics.

M: "Micronemes" "... myriad of nameless clues that color and shade how we thing about things ...". M wants the state of these micronemes to be context for other reasoning.
K: So my question would be whether these micronemes have values, logical or numeric. If they must be either on or off, then what he's doing seems equivalent to standard first order logic.

A Hierarchy of Representations
M: Postulates a hierarchical representation that might plausibly work:
– Narrative Stories
– Trans-Frames
– Frames containing semantic nets
– K-Lines and K-Trees
– Neural Nets
– Micronemes

K: As a working hypothesis I think that FCG's can do well for most of these levels if we use fuzzy inference methods to "activate" mental states. I even think that at the lowest level some form of fuzzy

computation underlies those neural networks.

How do we learn new representations?
M: "I suspect that, in the case of physical objects, our brains are already innately endowed with machinery to help us 'to compare, to connect, or to separate' objects so that we can represent them as existing in space."
K: M uses the term 'representation' somewhat ambiguously both to mean the particular abstraction of reality that is constructed and for the physical means used to instantiate that abstraction. He perhaps leans more towards the instantiation mechanisms: K-lines, frames, semantic nets.

M: Primitive structures built in but require time and effort to refine. Research needed on this process.
K: M asks – could any person ever invent a totally new kind of representation? Not clear what is meant by 'representation' in this context.
M: A new representation requires "... effective skills for working with it."
K: This language reflects a data/process dichotomy that is common in computer science, or perhaps the notion of an 'abstract data type' that also embodies legal actions that can be taken to affect it. The advantage of my FCG idea is that the result of adding new knowledge is immediately operational. As refinements in understanding become clear to us our processes are commensurately adjusted automatically.

M: Thinks it makes more sense to figure out what 'methods' work best and then pick a representation that works well with that method.
K: So clearly there are very specialized mechanisms used to interface with and do preliminary processing of sense data. But at somewhat higher levels of reasoning the power of the human brain comes from its ability to be very versatile in the ways in which it operates. Whether one calls these modes different 'representations' depends on whether we're talking about a new abstraction or a new mechanism for instantiating it. I believe that we have a single mechanism that is capable of handling fairly arbitrary new abstractions. We don't need to create a new mechanism for each new abstraction that we'd like to use.

Part 9 The Self

M: "The Single-Self view thus keeps us from asking difficult questions about our minds. We have multiple models of our 'self'".

9-1 How do we represent ourselves?
M: Gives examples of utility of multiple different models of a single system.
K: Note M seems to equate model with representation.
M: We need multiple models of ourselves
    We need multiple sub-personalities
K: Yep and Yep

The Sense of Personal Identity
M: Maybe the same thing that causes us to invent explanatory causes for things compels us to invent the sense of "I".

9-2 Personality Traits
M: Convenient short-hand descriptions of behavior.

Self-Control (Notes taken 12/06/08)
M: Need to learn techniques to ensure dependability and persistence. Adopt admiration for these qualities from imprimers (including society).

Dumbbell Ideas and Dispositions
M: We tend to view traits as one of two opposites: solitary/sociable, Dominant/Submissive, Tranquil/Agitated, etc. Left/right brain differences arise over time as one side assumes control for both. Side that has language tends to get other control resources because language is so crucial to many of them and physical proximity makes connections easier.

Many things have opposites.

Intensities and magnitudes
M: Adjectives such as 'slightly', 'largely', or 'extremely' can be applied to almost every emotion word.
K: Fuzzy Logic (FL) has a nice way to realize these modifiers for fuzzy values.

Structural vs. Functional descriptions
M: Distinction is often useful

Inborn Brain-Machinery
M: We may think in terms of pairs because our brains have mechanisms for comparing pairs.
K: My whole idea of dissonance driving our actions is based on the assumption that we can compare the activation levels on the two sides of any association. Is that strictly a binary comparison? I don't think that's necessarily the case, but may be much more prevalent than 1-to-many or many-to-many associations.

M: Why do people find it so hard to classify things into more than two groups? Is it because language doesn't have verbs for "tri-viding"? Lots of pairs experienced in early life. Many fewer triplets.
K: We divide groups in order to make a distinction that helps us model better in some way. If that model isn't adequate we might further divide one of the two groups previously identified. That effectively creates a "tri-vision" without ever having to think about it in the first place. So having one mechanism for dividing into two groups is always sufficient and from an evolutionary perspective we would be unlikely to evolve some more complex innate machinery for tri-viding. This doesn't stop us from creating higher level algorithms that we can adopt which can create more groups: it's just significantly more difficult for us to do so. We can learn to execute a K-means clustering algorithm, but it's not an instinctive thing to do.

9-3 Why do we like the idea of a Self?
M: People have varying senses about their own unity. "... even when we feel that we're in control, we recognize conflicts among our goals"
"Thus what we call 'self' in everyday life is a loosely connected collection of images, models, and anecdotes." We represent self the same way we represent everything external; with multiple models.

M: Single-Self model is useful in many situations. Maybe that's why we have it.
K: I think M is still dodging a crucial question here. I can (and do) accept that our brain functions are widely distributed. But I'd still like to understand what is different between a conscious state and an unconscious state. Clearly the brain is still active during the latter. If we think of dream states as unconscious states, then clearly there are fairly sophisticated sorts of things going on including language processing and generation as we talk to others in our dreams. It seems to me that what is often missing are the filters and self-reflection that would cause me to question what is going on. I can't say there isn't any dissonance because there can be enough to cause nightmares, but clearly some brain functionality is missing because I can be very amazed that my dreaming self never questioned all the absurdities of its environment. I think if we can come to understand what's going on in these non-conscious states we'll be able to explain brain functioning much better.

9-4 What is Pleasure and why do we like it?
M: Very difficult to describe pleasure and pain. We resort to analogies. It's hard because we can't further divide it into parts. M will argue there are many processes that make up 'pleasure'. M discusses multiple situations under which you experience pleasure.

The Pleasure of Exploration
M: Understanding something new and different can lead to pain and stress. Why do we do it?
Adventurousness: "... while some parts of your mind are uncomfortable, other parts of your mind may enjoy forcing those first parts to work for them."

9-5 What controls the mind as a whole?
M: Reiterates organization previously given.

Is a mind like a human community?
M: Human orgs tend to be hierarchical, tree-like, with a single root. Not a good brain model.

M: Popular belief of brain parallelism. Low-level functions are parallel, but much high-level reasoning/functioning is carried out serially.
K: Yes and no – I still think that there are lots of things going on at an unconscious level, just waiting to take over if the currently dominant process experiences some major dissonance.

Central and Peripheral Controls (Notes taken 12/07/08)
M: Many resources work without interrupting conscious processes. Alarms come from inside and outside. We could take a less centralized view "... in which the processes we call 'thinking' are affected by a host of other, partly autonomous processes."

M: Human processes frequently 'crash' but we recover so quickly we don't notice.
K: I agree. When dissonance reaches a level that is too high to sustain the current mode of thinking, that process is usurped by another that seems to be more capable of reducing dissonance. Dissonance may arise from lack of progress, increasing differences between current state and goal state, some external event, something equivalent to a timer interrupt, etc.

Mental Bugs and Parasites
M: If a mind could make changes in how it works it would risk destroying itself. Resource shouldn't be able to affect credit assignment or goal evaluation because it could increase beyond all bounds.
K: Sounds like the definition of obsessive/compulsive disorder.
M: Refers to certain non-specific pathologies that resemble this.

Why don't we have more bugs than we do?
M: Common bugs:
– Making generalizations that are too broad
– Failing to deal with exceptions to rules
– Accumulating useless of incorrect information
– Believing things because our imprimers do
– Making superstitious credit assignments
– Confusing real with make-believe things
– Becoming obsessed with unachievable goals
K: Most of these 'bugs' are very close to very useful behaviors. The ability to see generalizations can be a powerful part of learning. Becoming obsessed with a goal that everyone else believes is impossible may lead to a breakthrough if it isn't impossible for you.

9-6 What makes feelings so hard to describe?
M: Typically we describe feelings in terms of similarity to other feelings rather than more directly in terms of other primitives. One of the hardest philosophical problems is "Why do people <u>experience</u> events – instead of just processing them."

K: Following is a great insight from M ...
M: "... if your brain can begin to speak about some 'experience' it must already have access to some representations of some aspects of the event; otherwise, you would not remember it – or be able to claim that you have experienced it!" Sensations are extremely complex and involve many different parts of the brain.

K: I have argued that we have the ability to sense much of what is going on within our brains. But it may be that we don't have much, if any, ability to sense what is going on in our low-level sense processing resources. That makes it much more difficult for us to describe those sensory processes.

9-7 How do you know when you're feeling a pain?
M: We don't always immediately recognize that we're experiencing pain. I.E. if you're focused on

something else at the time. We may also think we have pain when we don't. " ... our self reflections reveal very little about the nature or causes of what we can see of our own mental activities."
K: I'd agree with that generally. I think we can sense the level of dissonance associated with whatever resources are currently active. And more critically, we can sense the linguistic utterances that are generated by any number of disparate resources. Indeed, I would postulate that the ability for one resource to hear and react to the utterances of another is essential to our sense of self. I literally hear myself say things that are never uttered aloud.

(Notes taken 12/08/08)
Feelings are hard to describe because they are complex!
M: Folk psychology says that sensations are irreducible. Philosophers maintain this is a hard problem because there are no physical properties that can be described. Clearly there must be a mechanism in the brain that detects sensations because we can report that the experience occurred. To understand how feelings work we'll have to look for complex processes, not simple ones. "... what we call 'qualities' might mainly reflect the ways we assess the relations *between* events in our brains."

9-8 The Dignity of Complexity
M: 30 Million centuries of evolution led to the human mind.

9-9 Some Sources of Human Resourcefulness
M: Genetic, Cultural, Personal Experience
Genetic: "Inherited systems in our brains help us survive the most common kinds of hazards and threats. Those mental resources were selected from variations that occurred over five million centuries."
Cultural: "The communal sets of beliefs called 'cultures' evolved over hundreds of centuries during which intellectual processes selected ideas from many millions of individuals."
Individual Experience: "Each year, one learns millions of fragments of knowledge from one's own private experiences."

M: Metaphors are everywhere in our language/thinking. We can create our own metaphors, but many are suggested via our language.

END OF BOOK