The Emotion Machine
Marvin Minsky
2006

Introduction.

"Each of our major 'emotional states' results from turning certain resources on while turning certain others off – and thus changing some ways that our brains behave" pg 4

"If you 'understand' something in only one way, then you scarcely understand it at all – because when you get stuck, you'll have nowhere to go. But if you represent something in several ways, then when you get frustrated enough, you can switch among different points of view, until you find one that works for you!" pg 6

Chapter 1. Falling in Love. (We are born with many mental resources)

When a person you know has fallen in love, it's almost as though someone new has emerged-a person who thinks in other ways, with altered goals and priorities. It's almost as though a switch had been thrown and a different program has started to run.

Whenever you think about your "Self," you are switching among a huge network of models, each of which tries to represent some particular aspects of your mind-to answer some questions about yourself

In everyday life, we expect our friends to know what we mean by Pleasure or Fear-but I suspect that attempting to make our old words more precise has hindered more than helped us to make theories about how human minds work. So this book will take a different approach, by thinking of each mental condition as based on the use of many small processes.

It would be naive to try to describe a mind as nothing more than bundles of If then Do rules. However, ... They can account for a remarkable range of what animals do.

In the infant brain, only one "way to think" can work at a time, so that not many conflicts can arise. Adult brains have "Critic Selector Machines" which can recognize a problem type and activate a way to think.

Emotional cascades occur when a newly aroused resource proceeds to activate yet other ones.

Chapter 2. Attachments and goals. (We learn more from interacting with others)

This chapter considers how we acquire new kinds of goals.

An Imprimer is one of those persons to whom a child has become attached. In an attachment based learning scheme: If you detect praise and an Imprimer is present, Then "elevate" your present goal. Attachment depends on the quickness with which the person responds, and the intensity of that interaction.

"In the course of each person's development, we tend to evolve certain policies that seem so consistent that we (and our friends) can recognize them as features or traits-and we use these to build our self-images. Then when we try to formulate plans, we can use those traits to predict what we'll do (and to thus discard plans that we won't pursue). Whenever this works, we're gratified, and this leads us to further train ourselves to behave in accord with these simplified descriptions. Thus, over time our imagined traits proceed to make themselves more real." pg 61

At first, those "Imprimers" must be near to us, but once we make "mental models" of them, we can use those models to "elevate" goals even when those Imprimers are absent-and eventually, these models become what we call conscience, ideals, or moral codes. Thus, attachments teach us ends, not means-and thus impose our parents' dreams on us.

Chapter 3. From Pain to Suffering (Emotions are different Ways to Think)

Pain makes you focus on the body parts involved. It makes it hard to think about anything else. Pain makes you move away from the cause. It makes you want that state to end, while teaching you, for future times, not to repeat the same mistake.
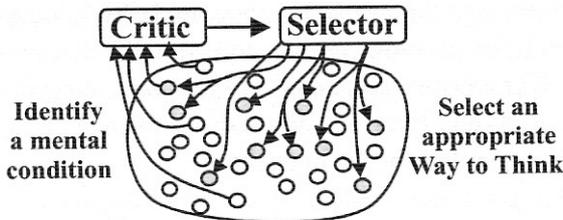

A CRITIC SELECTING A SET OF RESOURCES

Suffering is the frustration that comes with the loss of your options.

Is an expert some-one who knows what to do – or what not to do?

A Corrector declares that you are doing something dangerous. A Suppressor interrupts before you can begin the action you're planning. A Censor acts yet earlier, it prevent that idea from occurring to you.
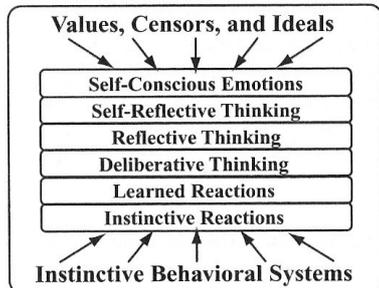
Chapter 4. Consciousness. (We learn to think about our recent thoughts.)

Consciousness is a suitcase like word that we use to refer to many different mental activities, which don't have a single cause or origin.

The Organism Principle: When a system evolves to become more complex, this always involves a compromise: if its parts become too separate, then the system's abilities will be limited-but if there are too many interconnections, then each change in one part will disrupt many others. This surely is the reason why the bodies of all living things are composed of the distinctively separate parts we call "organs."

Aaron Sloman 1992: "I, for one, do not think defining consciousness is important at all, and I believe that it diverts attention from important and difficult problems. The whole idea is based on a fundamental misconception that just because there is a noun 'consciousness' there is some 'thing' like magnetism or electricity or pressure or temperature, and that it's worth looking for correlates of that thing. Or on the misconception that it is worth trying to prove that certain mechanisms can or cannot produce 'it,' or trying to find out how 'it' evolved, or trying to find out which animals have 'it,' or trying to decide at which moment 'it' starts when a fetus develops, or at which moment 'it' stops when brain death occurs, etc. There will not be one thing to be correlated but a very large collection of very different things."

The Immanence Illusion: For most of the questions you would otherwise ask, some answers will have already arrived before the higher levels of your mind have had enough time to ask for them.

Drew McDermott 1992: "The key idea is not just that the system has a model of itself, but that it has a model of itself as conscious. A computer might have a model of its environment, in which it models itself as a piece of furniture. It wouldn't be conscious on that account."

Daniel Dennett 1991: "[This concept assumes that] there is a crucial finish line or boundary somewhere in the brain, marking a place where the order of arrival equals the order of 'presentation' in experience because what happens there is what you are conscious of. . . . Many theorists would insist that they have explicitly rejected such an obviously bad idea. But ... the persuasive imagery of the Cartesian Theater keeps coming back to haunt us - laypeople and scientists alike - even after its ghostly dualism has been denounced and exorcised."

Chapter 5. Levels of Mental Activities. (We learn to think on multiple levels.)
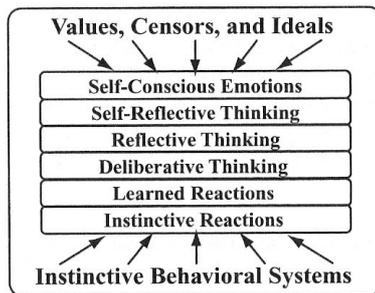
An IF-Then-Do model has far to many exceptions in the real word to be functional. Goals and planning are required. If-situation-And-Do-action-Then-result.

Learned reactions in animals are based on random search and followed by reinforcement/punishment.

Deliberation requires assessing alternatives and choosing.

A forward search of 20 binary steps has 2^{20} alternatives. Searching both forward 10 steps and backward 10 steps to find a common intermediate involves 2^{11} alternatives. Inserting a likely sub-goal further reduces the alternatives to 2^7.

Reflective thinking requires some short term memory and partial self-models at many levels.

Self-Reflection goes beyond thinking about recent events to thinking about the entity that had those thoughts.

Self-conscious Reflection enables us to think about higher values and ideals.

Structures as complex as a brain will not be adequately explained by one simple model with minimal assumptions like in physics.

Imagination. Computer modeling of visual systems have incorporated Image filters, Feature finders, Region finders, Object finders, Scene analyzers, and Scene describers. We recognize things by being reminded of familiar objects that could match incomplete fragments of evidence.

Envisioning Imagined scenes. It is much more efficient to imagine changes at higher semantic levels (more abstract). Coin the word "simulus" to refer to the various levels that people construct synthetic perceptions in their heads.

Prediction Machines. The ability to disengage if-then loops from the real world inputs and outputs.

Chapter 6. Common Sense. (We accumulate huge stores of commonsense knowledge.)

Out of date AI programming approaches include: lack of commonsense knowledge, no explicit goals, insufficient resourcefulness.

"Common sense is the collection of prejudices acquired by age eighteen" Albert Einstein

A Panalogy (parallel analogy) is the ability to switch easily between different ways to think about an object idea or situation. A thing or idea can be viewed as having multiple meanings in alternate realms (physical, social, emotional, visual, mnemonic ...)

Common-sense is a huge collection of scraps of knowledge, tricks and known dead ends.

Attempts to build a baby-machine have been unsuccessful because of:
● The optimization paradox: the better a system works, the more likely a change will degrade performance
● The investment principle: the better a principle works the more we invest in it to the exclusion of alternatives
● The complexity barrier: the more the parts interact the more likely each change will have unexpected side effects. (development of organs in organisms has mitigated this)

Evolution has taken millions of years and has developed high level structure which enable knowledge to be represented.

"Most systems still classify things in terms of how those things have been described instead of in terms of the goals that they can help us achieve." pg 183

Difference engines compare the current description with the desired state and act to remove the largest difference. This captures the everyday concept of a goal. However self reflection is required to allow a difference engine to overcome local minima. Every difference becomes a sub-goal. Successful paths are turned into sequential scripts. See Newell et al 1960b "A Variety of intelligent learning in a general problem solver"

"most of our external sensors react only to rather rapid changes in conditions" pg 196 Music vision and language narratives are perceived in increasingly large physical/time scales. We frequently use analogy to move between realms as in personification of an orchestra.

Once a repetitive pattern is detected in information

flowing up thru the brain then the signaling direction may reverse to filter out the irrelevant information.
Patrick Winston proposed that Difference-Networks could be used to find substitutes to achieve a particular goal. A Difference link is "A is like B, except for difference D".

Bengamin Franklins (1772) decision making technique involved writing down pros and cons over 3-4 days, considering their respective weights and then striking out any 1 or more pro's that balanced any 1 or more cons.

"If I had eight hours to chop down a tree, I'd spend six hours sharpening my axe." Abraham Lincoln.

Douglas Lenat 1997: " Analogy works because there is a lot of common causality in the world, common causes which lead to an overlap between two systems, between two phenomenon or whatever. We as human beings, can only observe a tiny bit of that overlap; a tiny bit of what is going on at this level of the world ... [So] whenever we find an overlap at this level, it is worth seeing if in fact there are additional overlap features, even though we do not understand the cause or causality behind it."

George Polya 1954: "We can learn the use of such fundamental mental operations as generalization
, specialization, and the perception of analogies. There is perhaps no discovery, either in elementary or advanced mathematics, or, for that matter, in any other subject, that could do without these operations, especially without analogy."

"Notice that to make and use analogy, one must work on three different levels at once: (1) descriptions of the original objects, (2) descriptions of their relationships, and (3) descriptions of the differences between those relationships." pg 209

Chapter 7. Thinking (We switch among different ways to Think.)

This chapter is concerned with goals and the idea that we have many ways to think. The many different ways are the result of evolution and the need to be adaptable.

Mostly the subjects that we think about are directed unnoticed due to every-day events. The mind reacts to cognitive obsticals with the critic-selector. The selector tries to start up a set of resources that will likely be successful on the current problem. The most basic critic-selectors will be simple if-then rules. More sophisticated will employ prioritization among different conflicting selectors. And more sophisticated still more reflective policies.

Damisio suggested that the reduced emotions played a role in Eliots decision making failures. However, Minsky thinks that is was the changes to Eliots critics (ways to think) that reduced his emotions (cascades).

Some useful ways to think include: knowing how, searching, analogy, divide and conquer, reformulating, planning, simplifying, evaluating, changing the subject, wishful thinking, self-reflection, impersonation, logical contradiction, logical reasoning, external representations, imagination, cry for help, ask for help, and Resignation.

Some useful critics include: Innate reactions and built-in alarms, learned reactive critics, deliberative critics, reflective critics, self-reflective critics, self-conscious critics.

We do not know how new critics and selectors are learnt. Some argue that emotional embodiment drives thinking (ie your brain reads the physical signs of your innate emotions). Some focus on unconscious processes like preparation, Incubation, revelation, evaluation and collaboration.

Our mind needs to be able to store cognitive contexts to that when distracted from a task one does not have to re-start from the beginning. Memories appear to become progressively permanent. Retrieval involves significant reconstruction. We can only think about a few things at once at a high level but many other happen continuously and automatically.

Each way to think must have some suppressive ability on others to enable persistence.

Psychology should return from statistics to individual analysis.

Chapter 8. Resourcefulness (We find multiple ways to represent things.)

"If you understand' something in only one way, then you scarcely understand it at all – because when something goes wrong, you'll have no place to go. But if you represent something in several ways, the when one method fails you can switch to another."

When we estimate distances we use a variety of approaches including, size, shading, perspective, height, grain, parallax, and focal length.

Panologies allow you to switch rapidly between different realms of the same real word thing.

We don't see everything when we enter a room and our eyes dart around with saccades. The reality is represented as fixed models while we move. We rarely make an entirely new idea but modify an existing one.

We can learn rapidly from one example by using short term memory to repeat the example many times in our head to develop stable models.

Children learn new things <u>and</u> new thinking techniques.

Credit must be assigned to successes – but only to those aspects of the success which are sufficiently generic to be applied to future problems. Self-reflection has a multiplying role.

Creativity and genius is characterized by:
• high proficiency in a field
• high self confidence
• high persistence
• accumulate more ways to think
• habitually think in novel ways
• better systems for self control
• reject many popular beliefs and myths
• tend to keep thinking more of the time
• excel at explaining what they have done
• tend to make better credit assignments.

Knowledge is not composed of isolated ideas but interconnected ideas.

Modern researchers are using a number of ways to represent knowledge:
• describing events as stories or scripts
• describing structures with semantic networks (a collection of symbols with labeled links)
• using trans-frames to represent actions (a pair of semantic networks with differences due to actions)
• using frames to embody commonsense knowledge (represents the effect of an action before an after the action – ie a pair of semantic networks but with many slots to contain when where why info)
• learning by knowledge lines (K-Lines link to and activate a number of resources which have been successfully combined on previous problems)
• Connectionist and Statistical representations differ from semantic networks in that they do not have an explicit representation of what thy store. Minsky believes the the continued focus on connectionist networks and baby machines has retarded efforts on the higher level representations of knowledge.
• Micronemes for Contextual knowledge. Micronemes are the myriad of nameless clues that color and shade our thoughts about things.

Perhaps all of these need to be connected in some loose hierarchical way.

We learn from experience but we are able to learn due to genetics.

It is important to have many different ways of representing the same thing so that we have different ways to guess an answer.

Chapter 9. (We build multiple models of ourselves.)

There is no homunculus. The models people make of their friends are frequently better than the models they make of themselves. "Others often better express myself" Kevin Solvay.

We create multiple models of our-self and the first split is the mind-body split. A self is not a permanent unchanging entity but a collection of drafts of self-portraits or stories.

We need to make ourselves predictable. Self has a considerable measure of self-control.

Two part dumb-bell descriptions are common in describing personalities (dominant vs submissive, cheerful vs cranky etc). There is an abundance of two's and opposites in our environment.

We like the idea of self and build a model or representation just like we do for everything else. The idea of self is supported by a single physical body, private thoughts, moral responsibility, decision making, casual attribution, attention and focus, and social relations.

Pleasure is a suit-case word for satisfaction, exploration, goal-suppression and relief. Goal suppression and exploration are consistent with having multiple selves as some are being satisfied while others are not.

Feelings are hard to describe because the apparent directness of experience is an illusion that comes because our higher mental levels have such limited access to the systems we use to recognize, represent and react to our external and internal conditions.

Feeling pain is at first an interruption to whatever thinking is going on and it may grow to be all consuming if the balance is heavily tipped.

Philosophers argue that the basic feeling of redness is a primitive not experienced by a machine but the red light is measured by the brain by an instrument we call an eye.

The human mind develops from basic groups of resources to, if-then-do rules to, Critics and Selectors to, Symbolic trees. Each has virtues and faults. These capabilities are not arranges in a clean hierarchy.

Parts of the brain may frequently break down but it simply switches to a different way to think due to the redundancy.

Critical thinking skills are the only protection to the variety of harmful memes such as religions etc.

our fabulously capable brains have been hard won in evolutionary terms.

Three sources of resourcefulness were developed on vastly different timescales: Genetic Endowment, Cultural Heritage and Individual Experience.