Marvin Minsky - The Society of Mind - 1986

Tom Rochette <tom.rochette@coreteks.org>

August 30, 2025 — 861fb9d0

0.1 Context

0.2 Learned in this study

0.3 Things to explore

- 28.8 What are the boundaries of the mind?

1 Overview

Marvin Minsky's *The Society of Mind* is the epitome of connectionism[1].

2 Thoughts

2.1 Unsorted

- Tasks are executed by low-level, specialized cognitive functions
- The brain is organized hierarchically, where more abstract processes are at the top and low-level functions are at the bottom. In between are *functionalities* that become more and more complex and abstract as they go to the top of the hierarchy
- Higher level functions manage low level functions to ensure they are doing their work properly as well as going in the right direction (doing the *right* thing)
- Higher level functions are competing with one another for resources and being the *main* active process
- Hierarchy of processes
- How are computers different from us?
- They follow instructions
- They don't start things by themselves (do we start things ourselves or if that simply the result of interacting with others?)
- Subconscious -> competing processes that are running in the background
- Learning is the process of replacing existing software
- To understand how something works, it helps to know how it can fail
- Detect and eliminate/externalize loops (prevent repetitiveness)

2.2 7.6 Reinforcement and Reward

- Reward the few moments before success. Works only for short solutions.
- What happens in an architecture with minimal memory? What happens when we gradually increase that memory?
- Only able to execute parts of a big loop
- Procedures encoded as part of a species' genes

---

[1] https://en.wikipedia.org/wiki/Connectionism

2.3 7.7 Local Responsibility

- Global vs local reward policy/scheme
- Global reward leads to slower learning

2.4 7.10 Genius

- Genius: knows how to manage what he learns
- Learn to rearrange how to think
- Learn better ways to learn to learn

2.5 8.1 K-lines: A Theory of Memory

- Tag all the tools that helped you solve a problem
- How do you know which tools to use? Won't you end up taking too many tools? Is it then a question of removing tags from tools that did not contribute to the solution?
- Tags = fragments of ideas

3 Notes

3.1 7.3 The Puzzle Principle

Puzzle Principle: We can program a computer to solve any problem by trial and error, without knowing how to solve it in advance, provided only that we have a way to recognize the problem is solved.

3.2 7.4 Problem Solving

The Progress Principle: Any process of exhaustive search can be greatly reduced if we possess some way to detect when "progress" has been made. Then we can trace a path toward a solution, just as a person can climb an unfamiliar hill in the dark - by feeling around, at every step, to find the direction of steepest ascent.

Goals and Subgoals. The most powerful way we know for discovering how to solve a hard problem is to find a method that splits it into several simpler ones, each of which can be solved separately.

Using Knowledge. The most efficient way to solve a problem is to already know how to solve it. Then one can avoid search entirely.

3.3 8.1 K-lines: A Theory of Memory

- How is knowledge represented?
- How is it stored?
- How is it retrieved?
- Then, how is it used?

3.4 17.6 Prerequisites for Growth

- One reason a skill may grow in steps is that it needs "prerequisites"
    - Some processes cannot be learned until certain other processes become available
- Many of Piaget's theories were based on his suspicion that certain concepts had prerequisites

3.5 17.8 Attachment-Images

- **The Oedipus complex:** rejecting one of the two parent models in order to simplify value-model learning.

3.6 17.9 Different Spans of Memories

- Parent-to-child and child-to-parents bonds based on certain types of memory
- The child (animals) learns through "imprinting" to recognize their parents
- The parents (animals) reject children with which they were not involved in bonding shortly after birth; it's an evolutionary disadvantage to raise the offspring of unrelated individuals
- Individuals either bond with a mate "for life" or to certain constant prototypes

3.7 17.10 Intellectual Trauma

- *I simply can't. I'm just no good at that.*
    - A learned way to avoid the shame and stress that came from social censure of failures in the past
    - It might also be a reaction to the nonsocial stress that came from having been unable to deal with certain ideas

3.8 18.1 Must Machines Be Logical?

- A logical system without a goal will merely generate an endless host of pointless truths

3.9 18.5 Strong Arguments

- Use several different arguments to prove the same point by putting them "in parallel"
- A chain (of arguments) can break with any single injury, but a parallel bundle cannot fail unless every one of its links has been broken

3.10 18.6 Magnitude from Multitude

Strength from Magnitude: When two forces work together, they add to form a single larger force. But when two forces oppose each other directly, their strengths subtract.

Strength from Multitude: The more reasons we can find in favor of a particular decision, the more confidence we can have in it. This is because if some of those reasons turn out to be wrong, other reasons may still remain.

3.11 18.7 What is a Number?

- Every concept is part of a huge network
- Meaning is derived from the network upon which it is built
    - Different for each individual

3.12 18.8 Mathematics Made Hard

- The more cross-connected our common-sense ideas are, the more useful they're likely to be

3.13 18.9 Robustness and Recovery

Duplication: Duplicate functionality so that if some part of the agents duplicating this functionality are lost, the others can take over.
Self-Repair: Be able to repair a faulty section if needed.
Distributed Processes: Distribute the function over many area, such that if a portion is destroyed, not all of it is destroyed.
Accumulation: Work is based on previous experience, such that if newer experience is lost, we can still proceed from older experience.

3.14   19.1 The Roots of Intention

- Words are like notes, alone they're not much, but together they form sentences, and thus meaning
- Where do intentions come from? What do they emerge from?

3.15   19.3 Words and Ideas

- Polynemes are similar to neural networks which are encapsulated in their own little black box
  - Such black box might be called "Color", "Shape", "Smell", "Taste", etc.
  - When a signal is received by the black box, it will configure itself to output a specific signal ("Color" -> "red", "Shape" => "round", etc.)
- Isonomes are short-term memory controls for other agencies

3.16   19.4 Objects and Properties

- A property is something that does not change capriciously (often)
- The most useful sets of properties those whose members do not interact too much (orthogonal)

3.17   19.5 Polynemes

- Each agency must learn its own specific and appropriate response. Each agency must have its private dictionary or memory bank to tell it how to respond to every polyneme (Chinese room?)

3.18   19.6 Recognizers

- The simplest way to recognize something is to verify that it has certain properties
- Often times it is not possible to verify that all properties are present, thus we might use a system that "looks for evidence" (x out of y properties to become active)
  - This system has many issues (identifying objects improperly, unable to attain identification threshold, etc.)
  - One also has to verify their dimensions and relationships

3.19   19.7 Weighing evidence

- Recognizers start to look **a lot** like neural networks (inputs + weights + output)
- Introduction of the perceptron
- Perceptrons are unable to take into account enough of the relations among various features

3.20   19.9 Recognizing Thoughts

- Within the brain, various forms of inputs can lead to the same "final signal", in other words the same thought generated from seeing an apple or being told "thing about an apple"

3.21   19.10 Closing the Ring

- If enough recognition-agent are aroused to trigger a polyneme, that polyneme may in turn activate related sensors/K-lines in other agencies.
  - In other words, if you start with enough clues to arouse one of your apple-nemes, it will automatically arouse memories of the other properties and qualities of apples and create a more complete impression, "simulus", or hallucination of the experience of seeing, feeling, and even of eating an apple.
- This could be called *reminding*

3.22   20.4 Locking-in and Weeding-out

- To use the power of language, one must acquire many different ways to understand

3.23   20.8 Connection lines

- Bus-like arrangement between transmitting-agents and receiving-agents
- Transmitting-agents excite n wires on the bus
- Receiving-agents are and-gates and will only activate upon the right combination of activations (1 wire per required wire signal)
- The disadvantage of a bus-like design is that it can transmit only one signal at a time

3.24   21.3 Trans-frames

- Conceptual dependencies:
  - P-Trans: act on location (move, go)
  - M-Trans: represents mental acts (tell)
  - A-Trans: represents a transfer (give, take)
- https://en.wikipedia.org/wiki/Conceptual_dependency_theory
- Similar to the concept of transitivity

3.25   21.7 Generalizing with Pronomes

- Example of Trans-script
  - Activate apple-neme, Look-for, and Move
    * Then activate Grasp
  - Activate pail-neme, Look-for, and Move
    * Then activate Ungrasp
- Separation of concern through the creation of two types of scripts:
  - a pronome-assignment script
    * Assign the apple-neme to the Origin pronome
    * Assign the pail-neme to the Destination pronome
  - an action script
    * Activate Origin. Then turn on Look-for, Move, and Grasp
    * Activate Destination. Then turn on Look-for, Move, and Ungrasp

3.26   21.8 Attention

- It is difficult to keep track of multiple things at once
- Forced to focus on a few while losing track of all the rest

3.27   22.1 Pronomes and Polynmes

- A pronome is a short-term memory
- Pronomes are temporary K-lines
- A polyneme is a long-term memory
- Polynemes are permenant K-lines

3.28   22.2 Isonomes

- An isonome has a similar, built-in effect on each of its recipients
  - It thus applies the same idea to many different things at once
- A polyneme has different, learned effects on each of its recipients
  - It thus connects the same thing to many different ideas

3.29   22.3 De-specializing

- The process of learning to build process scripts with isonomes

3.30 22.4 Learning and Teaching

- The power of what we learn depends on how we represent it in our minds
- We should be less concerned about teaching how to acquire particular skills and more concerned with learning how to learn

3.31 22.6 Expression

- Why do we "thing-ify" our thoughts?
  - It enables us to reapply the wonderful machines our brains contain for understanding worldly things
  - It helps us organize our expeditions in the mental world, much as we find our ways through space

3.32 22.7 Causes and Clauses

- Things: What we represent the scene with
- Differences: When we discern a change or just compare two different things, we represent this as a difference thing. In languages, these often correspond to verbs.
- Causes: Source of the action, change or difference. In languages, causes often take the forms of things.
- Clauses: Whatever structures we conceived are dealt with like single things. In languages, this corresponds to treating an entire phrase as though it were a single word.
- Our brains want us to represent dependencies

3.33 23.1 A World of Differences

- It is generally useless to do anything that has no discernable effect
- To ask if something is significant is similar to asking "What difference does it make?"
- Several kinds of thinking:
  - Predicting
  - Expecting
  - Explaining
  - Wanting
  - Escaping, attacking and defending
  - Abstracting

3.34 23.2 Differences and Duplicates

The Duplication Problem: The states of two different agencies cannot be compared unless those agencies themselves are virtually identical.

3.35 23.3 Time Blinking

- Any agent that is sensitive to changes in time can also be used to detect differences. For whenever we expose such an agent, first to a situation A and then to a situation B, any output from that agent will signify some difference between A and B.
- Two descriptions can be compared simply by presenting them to the same agency at different times
- Similar to a system where you do an initial setup (initial difference from scratch) and then only the differences are *sent* (like in video games, you don't send all units position at every frame, you only send their delta IF they moved)
- Time blinking has some limitations:
  - It cannot directly recognize relations among more than two things at a time

3.36 23.5 Foreign Accents

- Most human languages use less than a hundred phonemes

- One or more of the genetically controlled mechanisms that brings on sexual maturity also acts to reduce the capacities of these particular agencies (that reproduce accents) to learn to recognize and make new sounds
- Hypothesis: The onset of the childbearing age is the biological moment when a person's social role changes from learner to teacher. The "evolutionary purpose" of suppressing speech-sound learning may simply serve to prevent the parent from learning the child's speech - thus making the child learn the adult's speech instead

3.37   24.1 The Speed of Thought

- Each perceptual experience activates some structure that we'll call frames - structures we've acquired in the course of previous experience.

3.38   24.2 Frames of Mind

- Default assumptions fill our frames to represent what's typical

3.39   24.4 Default Assumptions

- Unless we make assumptions, the world would simply make no sense

3.40   24.8 How Picture-frames Work

- Each time you look in a different direction, your vision-system will describe what you see - and the corresponding K-line will record what you see when you look in that direction
- If this same frame is activated at some later date, we should perceive the same scene through "simulus"
- Through this simulus and the perception of the actual scene, it should be possible to determine the differences between the two

3.41   24.9 Recognizers and Memorizers

- A recognizer is the opposite of a K-line
  - A K-line arouses a certain state of mind
  - A recognizer recognizes when a certain state of mind occurs

3.42   25.1 One Frame at a Time?

- Our vision-systems are born equipped, on each of several different levels, with some sort of "locking-in" machinery that at every moment permits each "part," at each level, to be assigned to one and only one "whole" at the next level.

3.43   25.2 Frame-Arrays

- **Frame-Arrays** When we move, our vision-systems switch among a family of different frames that all use the same terminals.

3.44   25.3 The Stationary World

- We can see because we're seeing all the time and keeping contact with the world
- Our higher-level agents don't "see" the outputs of the sensors in our eyes at all
- They (the high-level agents) "watch" the states of middle-level agencies that don't change state so frequently
- When we move, we record the distortion applied to whatever we are looking at while a direction-neme is active (to indicate the direction of movement)

3.45   25.4 The Sense of Continuity

- We perceive continuity because that is how we model it internally
- Consciousness comes not from ceaseless change of state, but from having enough stability to discern significant changes in our surroundings

3.46   26.1 Understanding Words

- How do we bring together scattered knowledge quickly?
  - Some words activate a frame in the reader's mind
  - Attached to that frame are certain memories of various concerns
  - Each concern is represented by a frame whose terminals are already attached, as default assignments, the most usual solutions to that particular kind of problem
  - Such knowledge comes from previous experience

3.47   26.5 Story-Frames

- To listen well, a child must acquire potent forms of self-control
- It is also the responsibility of the storyteller to fix the focus of the listener's mind

3.48   26.6 Sentence and Nonsense

- Part of what a sentence means depends upon its separate words, and part depends on how those words are arranged
- A word-string seems "grammatical" if all its words fit quickly and easily into frames that connect suitably to one another

3.49   26.7 Frames for Nouns

- The way we think must have a strong and universal influence on how we speak - if only through its influence on the sorts of things we'll want to say

3.50   26.8 Frames for Verbs

- Multistage theory where we initially fill the terminals of word-string frames with nemes for words and later on we learn to fill those terminals with other filled-in language-frames

3.51   26.9 Language and Vision

- Points to the fact that vision predates language
- Language-agencies would have evolved from variants of genes that first evolved in shaping the architecture of our vision-systems

3.52   26.10 Learning Language

- A word can only serve to indicate that someone else may have a valuable idea - that is, some useful structure to be built inside the mind. Each new word only plants a seed: to make it grow, a listener's mind must find a way to build inside itself some structure that appears to work like the one in the mind from which it was "learned".

3.53   27.2 Suppressors

- We can imagine two poles of self-improvement
  - We can try to stretch the range of the ideas we generate: this leads to more ideas, but also to more mistakes
  - We can try to learn not to repeat mistakes we've made before

- **Suppressor-agents** waits until you get a certain "bad idea." Then they prevent you from taking the corresponding action, and make you wait until you think of some alternative. If a suppressor could speak, it would say, "Stop thinking that!"
- **Censor-agents** need not to wait until a certain bad idea occurs; instead, they intercept the states of mind that usually precede that thought. If a censor could speak, it would say, "Don't even begin to think that!"
- It's easier to study what someone does - instead of what someone doesn't do

3.54   27.5 Jokes

- Freud suggested that children construct censors in response to prohibitions by their parents or peers. This explains why so many jokes involve taboos concerning cruelty, sexuality, and other subjects that human communities typically link to guilt, disgust, or shame.

3.55   27.6 Humor and Censorship

- When we learn in a serious context, the result is to change connections among ordinary agents. But when we learn in a humorous context, the principal result is to change the connections that involve our censors and suppressors.
- Scolding tends to produce suppressors, but laughing tends to produce censors
- Suppressors merely need to learn which mental states are desirable
- Censors must remember and learn which mental states were undesirable

3.56   28.2 Magnitude and Marketplace

- We turn to using quantities when we can't compare the qualities of things

3.57   28.3 Quantity and Quality

- Whenever we turn to measurements, we forfeit some uses of intellect. Currencies and magnitudes help us make comparisons only by concealing the differences among what they purport to represent
- Any substance or quantity whose availability is limited can be made to serve as a currency.

3.58   28.4 Mind Over Matter

- A person's ability to persist in circumstances we hadn't thought were tolerable need not indicate anything supernatural. (. . .) It is merely a matter of finding ways to rearrange our priorities.

3.59   28.7 Individual Identities

- At what point replacing parts of someone's identity makes them a completely different person? If we were to replace every cell in your brain by a new one, would it still be your brain?
- Modifying or replacing the physical parts of a brain will not affect the mind it embodies, unless this alters the successions of states in that brain.

3.60   28.8 Overlapping Minds

- Like members of a family, the different minds (in your brain) can work together to help each other, each still having its own mental experiences that the others never know about.

3.61   29.1 The Reals of Thought

- We need to be able to describe things at many levels of detail.

3.62 29.2 Several Thoughts at Once

- How can many different thoughts proceed at the same time without interfering with one another? The processes use agents that do not compete (are orthogonal to one another)

3.63 29.3 Paranomes

- Many of our higher level conceptual-frames are really parallel arrays of analogous frames, each active in a different realm.
- **Paranomes:** Pronomes operating in several different realms at once (in parallel)

3.64 30.1 Knowing

- The meaning of "knowing" depends on who is saying it. After all, no one knows everything about everything.

3.65 30.2 Knowing and Believing

- We classify our thoughts into different types:
  - facts: "the red object is on the table"
  - opinions: "I think the red block is on the table"
  - beliefs: "I believe that the red block is on the table"

3.66 30.3 Mental Models

- Does a book know what is written inside it? Clearly no. Does a book contain knowledge? Clearly, yes. But how could anything contain knowledge, yet not know it?
- Possessing knowledge amounts to saying that some observer could employ that person or machine to answer certain kinds of questions.
- A person or machine *contains* a model of knowledge itself. It is not perfect, but it allows us to query it and generate answers.

3.67 Appendix

3.67.1 2 The Genesis of Mental Realms

- In their first few days, human infants learn to distinguish people by their odors; then, over the next few weeks, they learn to recognize individuals by sound of voice; only after several months do they start to reliably distinguish the sights of faces. Most likely we learn to make each of these distinctions by several different methods, and it is probably no accident that these abilities develop in a sequence that corresponds to their increasing complexity.

4 See also

5 References

- Minsky, Marvin Lee. The Society of Mind. New York: Simon and Schuster, 1986.
- http://www.amazon.com/The-Society-Mind-Marvin-Minsky/dp/0671657135