The Prehistory of Cognitive Science

#week-1   #ch-1   #my-notes-755

Contents

(1) - The reaction against behaviorism in psychology

Behaviorists think that,

- psychologists should avoid speculating about unobservable mental states,
- and instead focus on nonpsychological mechanisms linking particular stimuli with particular responses. (S-R associations)
- These mechanisms are the product of conditioning.
- eg- Pavlov's dog being conditioned to salivate at the sound of bell.

Some of the ideas that proved important for the later development of cognitive science by looking at 3 landmark papers:

- Learning without reinforcement:
  'Insight' in Rats (1930) - by Tolman and Honzik
- Cognitive Maps in Rats?
  Studies in Spatial Learning (1946) - by Tolman, Ritchie, Kalish
- Plans and Complex bahaviors:
  The problem of serial order in behavior (1951) - by Karl Lashley

'Insight' in Rats (1930)

Two standard behaviorist assumptions about learning.

- All learning is the result of conditioning.
- Conditioning depends upon processes of association and reinforcement.

We can understand these two assumptions by thinking about a rat in a Skinner box.

> In classical conditioning what is strengthened is the association between a conditioned stimulus (such as the typically neutral sound of a bell ringing) and an unconditioned stimulus (such as the presentation of food). The unconditioned stimulus is not neutral for the organism and typically provokes a behavioral response, such as salivation

Tolman and Honzik were interested in how rats learned to navigate mazes. They ran three groups of rats through a maze.

- g1: received a reward each time they successfully ran the maze.
- g2: never received a reward.
- g3: was unrewarded for the first ten days and then began to be rewarded.

OBSERVATIONS:

- g1 learned first
- g2 never learned
- g3 when started receiving rewards, they learned far more quickly than g1

CONCLUSIONS:
Behaviorism can explain g1 and g2:
since g1 received reward their SR-mapping became strong, and g2 were acting randomly
But they failed to explain g3:
had they been correct, then after g3 started getting reward they should have taken the same amount of time to learn as g1 (and not faster). This means they must have been learning all this while (and it was happening even without any reward).

> Tolman and Honzik argued that the rats must have been learning about the layout of the maze during the period when they were not being rewarded.
> This type of latent learning seemed to show that reinforcement was not necessary for learning, and that the rats must have been picking up and storing information about the layout of the maze when they were wandering around it, even though there was no reward and hence no reinforcement.

QUESTIONS: *But are the rats simply remembering the sequence of moves?*

Tolman and his students and collaborators designed many experiments during the 1930s and 1940s to try to decide between **place learning** and **response learning** accounts of how rats learn to run a maze. Some of these experiments were reported in a famous article in 1946.

Studies in Spatial Learning (1946)

One experiment used a cross-maze with four end points (North, South, East, West)
Rats were started at North and South on alternate trials.
g1: rewarded by the food at the same end point (say East) - **place learning**
g2: rewarded by the food at the same turning (East for North headed rats and West for South headed rats) - **response learning**

OBSERVATION:
g1 learned to run the maze much more quickly than g2.
This implied, place learning > response learning.

CONCLUSION:
Nature of animal learning in general:
It was easier for animals to code spatial information in terms of places rather than in terms of particular sequences of movements.

RESULT

- Tolman took his place-learning experiments as evidence that,
  Animals form *high-level representations of how their environment is laid out* – what he called **cognitive maps**.
- Tolman's cognitive maps were one of the first proposals for explaining behavior in terms of **representations (stored information about the environment)**

The Problem of Serial Order in Behavior (1951)

COMPLEX BEHAVIORS:

- Behaviorisms: They have to view them as linked sequences of responses
- **The standard behaviorist** view is that rats learn to chain together a series of movements that leads to the reward.
- **Tolman** showed that this is not the right way to think about what happens when rats learn to run mazes.

- **Lashley** made the far more general point that this seems to be completely the wrong way to think about many complex behaviors.

    - According to Lashley, we should think about many of these complex behaviors as products of **prior planning and organization**.
    - The **behaviors are organized hierarchically** (rather than linearly)

Lashley's essay contains the seeds of two ideas:

1. **hypothesis of subconscious information processing**
   Even though we are often conscious of our high-level plans and goals (of what goes on at the top of the hierarchy), we tend to be unaware of the information processing that translates those plans and goals into actions.
   - Example: consiously have intention to pick a glass, but calculations of the exact trajactory of hand is done far below the threshold of consious awareness.
2. **hypothesis of task analysis**
   understand a complex task (and the cognitive system performing it) by breaking it down into a hierarchy of more basic subtasks (and associated subsystems)
   - Example: Task to exploit previously acquired information can be broken down as-
   subtask: storing info - (sub-sub-task: long term, short-term)
   subtask: retrieving info

(2) - The idea of algorithmic computation in mathematical logic

---

- In 1936–7 **Alan Turing** published an article in the *Proceedings of the London Mathematical Society* that introduced some of the **basic ideas in the theory of computation**.
- **Computation** is what computers do and, according to many cognitive scientists, it is what minds do.
- What Turing gave us was a **theoretical model** that many have thought to capture the essence of computation.

Algorithms and Turing Machines

"On Computable Numbers, with an Application to the Decision Problem" (1936-7)

Computer Programs:

- Many computer programs are not defined for every possible input.
- They will give a solution for some inputs, the ones for which they are defined.
- But for other inputs, the ones for which they are not defined, they will just endlessly loop, looking for a solution that isn't there.

It is important to tell, whether the program is simply taking a long time to get to a solution, or is it stuck in an endless loop.

Halting Problem:

- *A fundamental concept in computer science and logic.*
- It refers to the problem of determining whether a computer program will eventually halt (terminate) or continue running indefinitely.

Solution to the Halting Problem:

- Given, a computer program and a given input(s).
- Can we tell, whether the program is defined for this input or not.
- The solution has to work both ways: "Yes" if defined and "No" if not defined.

Why the Turing Machine?

- Turing was looking for a *purely mechanical solution* to the Halting Problem.
- For this, he needed to make the solution more precise - an **Algorithm**

TURING MACHINE:

- An idealized machine that consists of:
    1. an infinitely long piece of tape (**Turing-Tape**) divided into cells
    2. a **Machine-Head**
    3. Set of instructions (**Machine-Table**)
- The "tape" being infinitely long => no storage restrictions => hence an ideal m/c
- Each "cell" in the tape can contain (1) a single symbol, or (2) be blank
- The tape runs through the "machine head", with a single cell under the head at a given instant
- The head can carry out a limited number of operations:
    1. delete a symbol in the cell
    2. write a symbol in the cell
    3. move the tape one cell to Left
    4. move the tape one cell to Right

- The *Turing machine* can be in any *finite number of different **states***.
- The "machine table" will determine, what the machine will do:
    1. when it encounters a given **symbol**
    2. in a particular **cell**
    3. depending upon its **internal state**
- **BEAUTY of TURING MACHINE:**
    - Its behaviour entirely determined by the machine-table (which can be mathematically represented by a *sequence of numbers*)
    - Hence, There is no ambiguity and no room for the machine to exercise "intuition" or "judgment."

OUTCOMES:

- Turing's paper contained a subtle proof that **Halting Problem cannot be solved**.
- **Church-Turing thesis:** Anything that can be done in mathematics by an algorithm, can be done by Turing machine.

CONTRIBUTION to Cognitive Science:

- A **"model of computation"** that looked as if it might be a clue to how **information could be processed by the mind**.
- As theorists moved closer to the idea that **cognition involves processing information** it was an easy step to think about **information processing as an algorithmic process** along the lines analyzed by Turing.

(3) - Linguistics and the Formal Analysis of Language

Language usage:

1. **hierarchically organized complex behavior** - Karl Lashley
2. **can be analyzed in algorithmic terms** - due to the emergence of tansformational linguistics and the formal analysis of syntax.

Syntactic Strutures (1957) - by Noam Chomsky

"why" (rather than "how") languages work the way they do

Structure of any language:

1. **Deep Structure** (phase structure grammer) - formed by basic parts of speech: nouns, verbs, adjectives. Similar in all languages
2. **Surface Structure** - The actual organization of words in a sentence.

- Analyzing sentences in terms of their phrase structure is a powerful explanatory tool.
- There are pairs of sentences that have very **different phrase structures, but are clearly very similar in meaning**.
  - eg- "John has hit the ball" and "The ball has been hit by John."
  - In most contexts these sentences are equivalent and interchangeable, despite having very different phrase structures
- The transformational principles of **transformational grammar** are examples of **algorithms**.
- They specify a set of procedures that operate upon a string of symbols to convert it into a different string of symbols.

CONTRIBUTION to Cognitive Science

Provided a very clear example of **how to analyze (in algorithmic terms)**, the bodies of information that might underlie certain very **basic cognitive abilities** (such as the ability to speak and understand a language)

(4) - Information-Processing Models in Psychology

The emergence of information theory can be pinned down to a single event – the publication of an article entitled "A mathematical theory of communication" by Claude E. Shannon in 1948.

Information-processing models became established in psychology through 2 famous publications:

1. "The magical number seven, plus or minus two"- *article by George Miller* (1956)
2. "The role of auditory localization in attention and memory span" - *paper by Donanld Broadbent* (1954).
   - The paper presented 2 inflencial experiments

- which led to the book: "Perception and Communication" - by Broadbent (1958)
    - one of the first *Information-Processing models in psychology*

The magical number seven, plus minus two (1956) - George Miller

The tools of information theory can be applied to the study of the mind. One of the basic concepts of information theory is the **concept of an information channel**. **Perceptual systems are themselves information channels**. Example: Vision is a medium through which information is transmitted from the environment to the perceiver.

> Human subjects are really rather **limited in the absolute judgments**
> (eg- naming a color, or identifying the pitch of a particular tone)
> – **as opposed to relative judgments**,
> (eg- identifying which of two colors is the darker, or which of two tones is higher in pitch)

EXPERIMENT

- Subjects were asked to assign numbers to pithces of tones
  Eg- (C <- 1, E <- 2, F# <- 3)
- Then they were presented a sequence of tones and asked to identify them in terms of the assigned numbers.
  Eg- If they heard the seq. (E-C-C-F#-E) then correct resp would be (2-1-1-3-2)

OBSERVATIONS

- When the sequence is only one or two tones long, subjects never make mistakes.
- But performance falls off drastically when the sequence is six or more tones long.
- *A similar phenonemon occurs when we switch from audition to vision and ask subjects to judge the "size of squares" or the "length of a line".*

CONCLUSION

- There seems to be an upper bound on the number of distinct items that can be processed simultaneously

RELATION TO PSYCHOLOGY

- Our **sensory systems are all information channels with roughly the same channel capacity**.
- The perceiver's capacity to make **absolute judgments is an index of the channel capacity** of the information channel that she is using.
- One way of increasing channel capacity - **chunk information**. (eg- 25 is easier to remember than 11001)
- Miller pointed out, to return to a theme that has emerged several times already, *natural language is the ultimate chunking tool*

"The Role of Auditory Localization in Attention and Memory Span" (1954) and "Perception of Communication" (1958)

One of the first models of **how sensory information is processe**d was developed by the British psychologist *Donald Broadbent* in his 1958 book "*Perception and Communication*"

The cocktail party phenomenon - *We manage to focus only on what we want to listen*

DICHOTIC LISTENING EXPERIMENTS

- The experiments were reported in his paper "The role of auditory localization in attention and memory span"
- Subjects are presented with different information in each ear. Involved presenting subjects with a string of three different stimuli (letters or digits) in one ear, while simultaneously presenting them with a different string in the other ear.
- The subjects were asked to report the stimuli in any order

OBSERVATIONS

Broadbent found that they performed best when they reported the stimuli ear by ear – that is, by reporting all three presented to the left ear first, followed by the three presented to the right ear

EXPLANATION using a Model
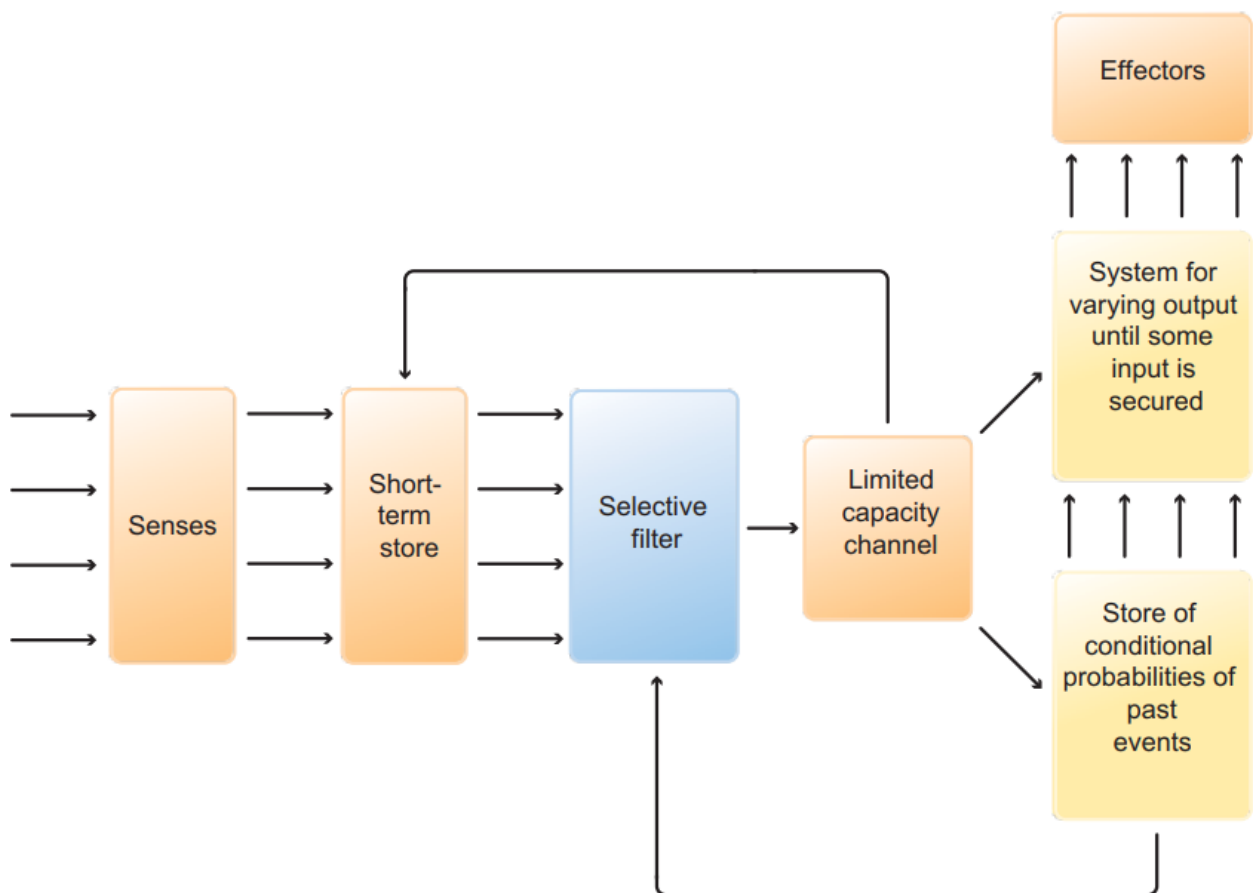
1. Information comes through the **senses**
2. and passes through a **short-term store**
3. then passes through a **selective filter**.
   - The selective filter screens out a large portion of the incoming information. Filtered-out info is quickly decayed.
     This is what allows us selectively to attend to only a portion of what is going on around us in the cocktail party.
   - Only information that makes it through the selective filter is "*semantically interpreted*"
4. Information that makes it through the selective filter goes into what Broadbent calls the **limited capacity channel**.
5. From the limited capacity channel information can go either into:
   - the **longterm store**, OR
   - on to **further processing and eventually into action**, OR
   - it can be recycled back into the **short-term store**

We can only attend to a single information channel at a time (assuming that each ear is a separate information channel) – and that the selection between information channels is based purely on physical characteristics of the signal.

the selective filter is "programmed" by another system

We are assuming that the system is pursuing a goal. What is programming the selective filter is information about the sorts of things that have led to that goal being satisfied in the past.