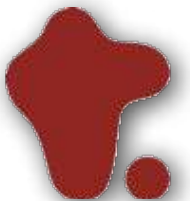# Cognitive Media Processing #1

Nobuaki Minematsu

Self-introduction

- Lecturer
  - Nobuaki Minematsu
    - Full professor @ Department of Electrical Engineering and Information Systems (EEIS)
    - mine@gavo.t.u-tokyo.ac.jp (ext.26662)
    - Specialty: speech science and speech engineering
      - How to build "human-like" speech machines?
    - http://www.gavo.t.u-tokyo.ac.jp/~mine/japanese/CMP/class.html
      - All the slides with some useful information are available on this web.
- Schedule of "Cognitive Media Processing"
  - Divided to three terms (three sub-themes).
  - **Human processing of media information**
    - 4 lectures on 10/3, 10/10, 10/17, and 10/24
  - **Speech communication technologies**
    - 4 lectures on 10/31, 11/7, 11/14, and 11/22(Wed)
    - Friday classes are give on 11/21.
  - **A new framework to build "human-like" speech machines**
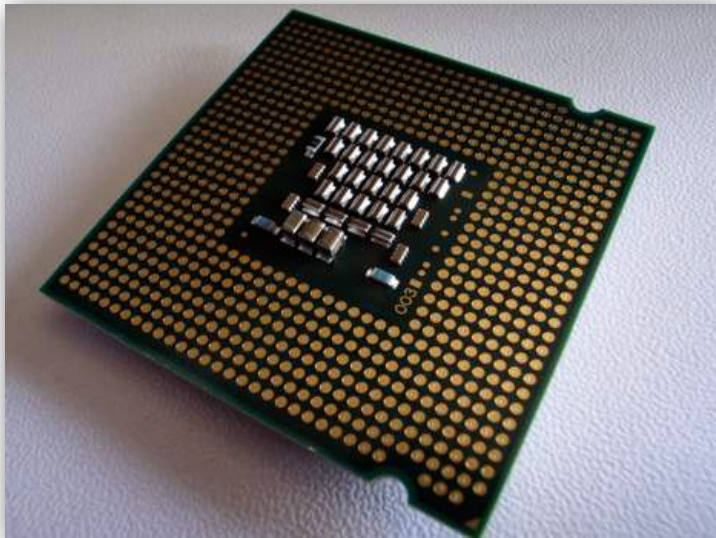   - 12/12, 12/19, 12/26, 1/9 (no class on 12/5 and 1/2)

Aim of this class

- Syllabus on the web
  - Cognitive processing of multimedia information by humans and its technical processing by machines are explained and compared. Then, a focus is placed on a fact that a large difference still remains between them. This lecture will enable students to consider deeply what kind of information processing is lacking on machines and has to be implemented on them if students want to create not seemingly but actually "human-like" robots, especially the robots that can understand spoken language.
  - The lectures are divided into three parts. The first part explains the multimedia information processing by human brains. Here, some interesting sensory characteristics of individuals with autism (自閉症), synesthesia (共感覚), and dyslexia (難読症) are shown as examples. The second part describes the current technical framework of spoken language processing and its drawback. The last discusses what kind of new methodology is needed to create really "human-like" robots that can understand spoken language. Then, a new framework is introduced and explained.

What I hope for you to acquire

- What I hope for you to acquire through the lectures
  - Human media processing, which can be a good model for computers or not.
  - What has been implemented on computers so far as media processing technology.
  - I hope for students to have a good sense to compare them and bridge the gap between them.
    - Similarity or difference, which is larger?
    - Your "conscious" world might be illusions created by your brain?

Title of each lecture

- Theme-1
  - Multimedia information and humans
  - Multimedia information and interaction between humans and machines
  - Multimedia information used in expressive and emotional processing
  - A wonder of sense - synesthesia -
- Theme-2
  - Speech communication technology - articulatory & acoustic phonetics -
  - Speech communication technology - speech analysis -
  - Speech communication technology - speech recognition -
  - Speech communication technology - speech synthesis -
- Theme-3
  - A new framework for "human-like" speech machine #1
  - A new framework for "human-like" speech machine #2
  - A new framework for "human-like" speech machine #3
  - A new framework for "human-like" speech machine #4

abcdefg
hijklmn
opqrstu
vwxyz

Credit

- Assignment will be given after each of the three themes.
  - **Human processing of media information**
  - **Speech communication technologies**
  - **A new framework to build "human-like" machines**
  - Students have to submit each assignment to me by ITC-LMS.
    - New students should be accustomed to using ITC-LMS for taking classes.
    - Japanese students are allowed to write their assignment in Japanese.
- Your grade (mark) depends only on the assignments.

Multimedia info. and humans

Nobuaki Minematsu

Today's menu

- The term of "information" used in human communication.
  - Two kinds of definition of information (C. Shannon vs. this lecture)
  - Data and information - intention of a sender and interpretation of a receiver -
- Various forms of information in human communication
  - Classification of media information
  - Context dependency of information
- Information and knowledge
  - From data to information
    - Knowledge-based cognitive processing
  - Unconscious processing
    - Your brain creates your world but you cannot be aware of the brain's processing.
  - Various forms of information and conversion between them
    - Recognition and synthesis: abstraction and embodiment
  - Logical information and expressive (感性, KANSEI) information
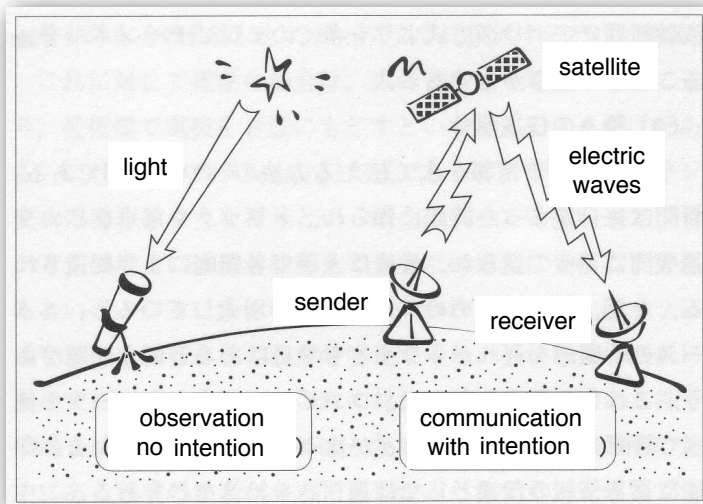   - Behaviors and information processing of autistics

Info. in human communication

- "Information" defined by information theory (C. Shannon)
  - Self-information of an even e_i that occurs with probability of p_i : $I(e_i) = -\log_2(p_i)$
    - Amount of "surprise" when e_i happens.
  - Expected self-information of a set of events $E = \{e_i\}$: $H(E) = -\sum_i p_i \log_2(p_i)$
    - "Information" can be treated not as quality but as quantity.
- "Information" defined by this lecture
  - Existence of a sender and a receiver
    - Information = something to be informed from a sender
      - Messages or data
    - A receiver receives messages or data via the five senses
  - What is "media"?
    - Physical media and social media
      - Frequency of light, frequency of air particle vibration, pressure on the skin, etc
      - Mass media such as books, newspaper, radio, TV, internet, etc.
    - Other kinds of media
      - Some properties are transmitted from parents to kids via. genes.
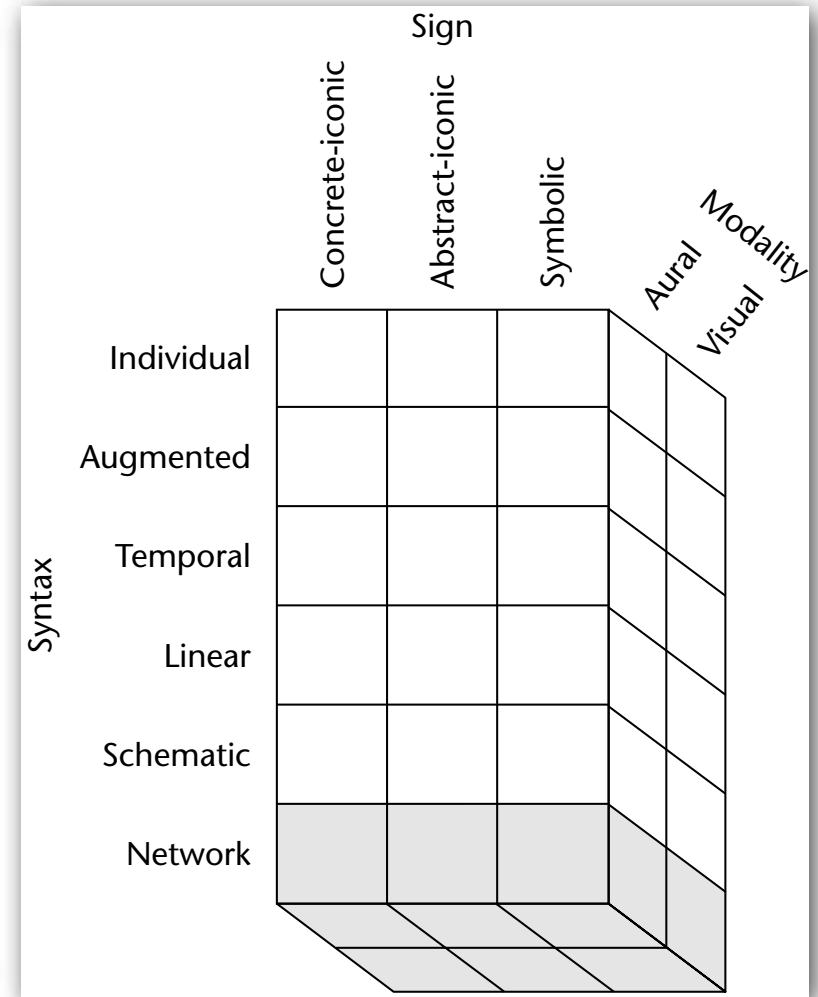
Sender ⎯⎯ ~~physical media~~ ⎯⎯⎯⎯⎯⎯⎯⎯ ~~physical media~~ ➜ Receiver

message

Info. in human communication

- Intention of a sender and interpretation of a receiver
  - Observation and communication
    - In both cases, a receiver receives messages or data and tries to interpret them properly.
      - Messages/data can become information only when a receiver can interpret them properly.
    - Intention of a sender
      - No intention : observation,   with intention : communication
  - Communication and miscommunication
    - Proper interpretation of both messages/data and the intention of a sender is needed.
    - Reading the mind of a sender is often needed.

satellite

light

electric waves

sender

receiver

observation
(no intention)

communication
with intention

if only i could read your mind

message

Sender ——— ~~physical media~~ ————————————— ~~physical media~~ → Receiver

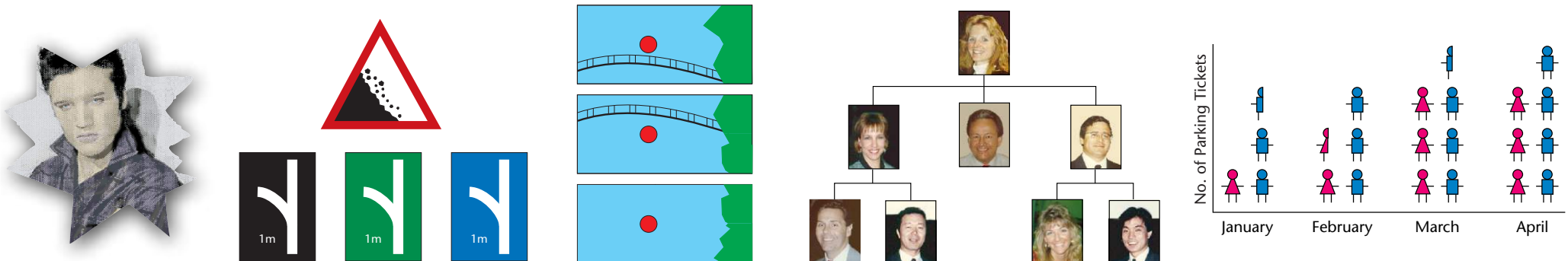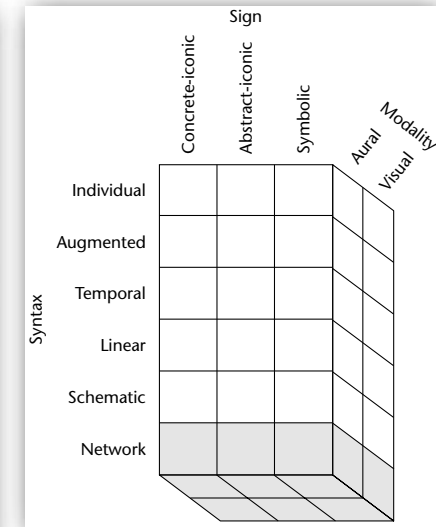# Forms of info. in human communication

- Classification of info. in terms of its physical media
  - Physical media is needed to transmit a message.
    - Sounds, letters, still and moving images, etc and their combinations
- Definition of multimedia
  - "Defining Multimedia" [G. Davenport'98]
  - Three dimensions - sign / syntax / modality
    - Concrete -- abstract
    - Temporal and/or spatial organization of signs
    - Sensation
  - Systematic understanding of the existing methods to represent multimedia info.
    - Finding and creating a new method

Forms of info. in human communication

- Definition of multimedia
 - Three dimensions - sign / syntax / modality in the case of visual modality

	Concrete-Iconic	Abstract-Iconic	Symbolic
Individual	Any photograph	An iconic road sign (see Figure 4)	Any written word
Augmentation	A shaped photograph, such as a star-shaped photo of a popular singer (see Figure 2)	Road signs whose color provides additional information (see Figure 5)	A word whose font provides additional information, such as a fast-food restaurant called *Express*, where
Temporal	A continuous rolli such as a film of a		
Linear	Any film		
Schematic	A taxonomic diag diagram depicting hierarchy of an or the relationship b are represented a: (see Figure 3)		

Forms of info. in human communication

- ## Definition of multimedia
  - ### Three dimensions - sign / syntax / modality in the case of aural modality

|	Concrete-Iconic	Abstract-Iconic	Symbolic
Individual	A recording of a brief, atomic sound, such as a car ignition	A brief, atomic synthesized sound, such as a "whirr" from a computer	A brief, atomic, symbolic sound like a doorbell
Augmentation	A recording of a brief, atomic sound, whose volume is significant, such as a door slammed in anger	A brief, atomic synthesized sound whose tone is significant, such as a desktop trash can that produces a "clunk" that decreases in tone as it fills	A brief, atomic, symbolic sound whose tone is significant, like an error "beep" that changes in tone according to the nature of the error
Temporal	A continuous recording representing a single concept, such as the sound of waves on a beach	A continuous synthesized sound representing a single concept, such as the sound of gunfire in a violent arcade game	A continuous symbolic sound, such as a fire alarm
Linear	A sequential recording of sounds representing a story, such as the build-up, height, and conclusion of a storm	A sequence of synthesized sounds, such as a train's approach, passing, and department that has been synthesized rather than recorded	A sequence of symbolic sounds, such as a computer "humm" that changes in pitch depending on the load on the network
Schematic	A recording of a sound comprising different frequencies, such as a car crash involving breaking glass and severe body damage	A synthesized sound track for an animated cartoon comprising different frequencies, such as a cat howling as it hits a solid wall	A complex sound where differing frequencies have different interpretations, such as a two-tone fire alarm indicating both location and severity of the fire
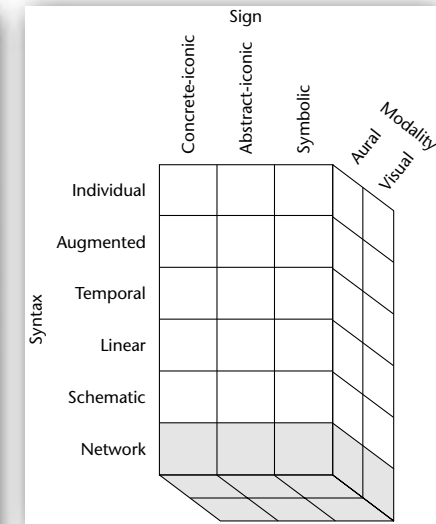
Table 3. Visual and aural modality examples in the network arrangement.

|	Concrete-Iconic	Abstract-Iconic	Symbolic
Visual modality	Interactive video, such as a video story where the reader chooses the story line	Interactive animation, such as an animated version of a video story	Hypertext, as an online thesaurus with links between related entries
Aural modality	Interactive audio of concrete-iconic audio recordings, such as a collection of different bird songs that can be selected individually	Interactive audio of abstract-iconic sounds, such as a collection of synthesized sound effects that can be selected individually	Interactive audio of symbolic sounds, such as touch-tone menus of recorded spoken information

Forms of info. in human communication

- Classification of info. in terms of its physical media
  - Physical media is needed to transmit a message.
    - Sounds, letters, still and moving images, etc and their combinations
- Definition of multimedia
  - "Defining Multimedia" [G. Davenport'98]
  - Three dimensions - sign / syntax / modality
    - Concrete -- abstract
    - Temporal and/or spatial organization of signs
    - Sensation
  - Systematic understanding of the existing methods to represent multimedia info.
    - Finding and creating a new method
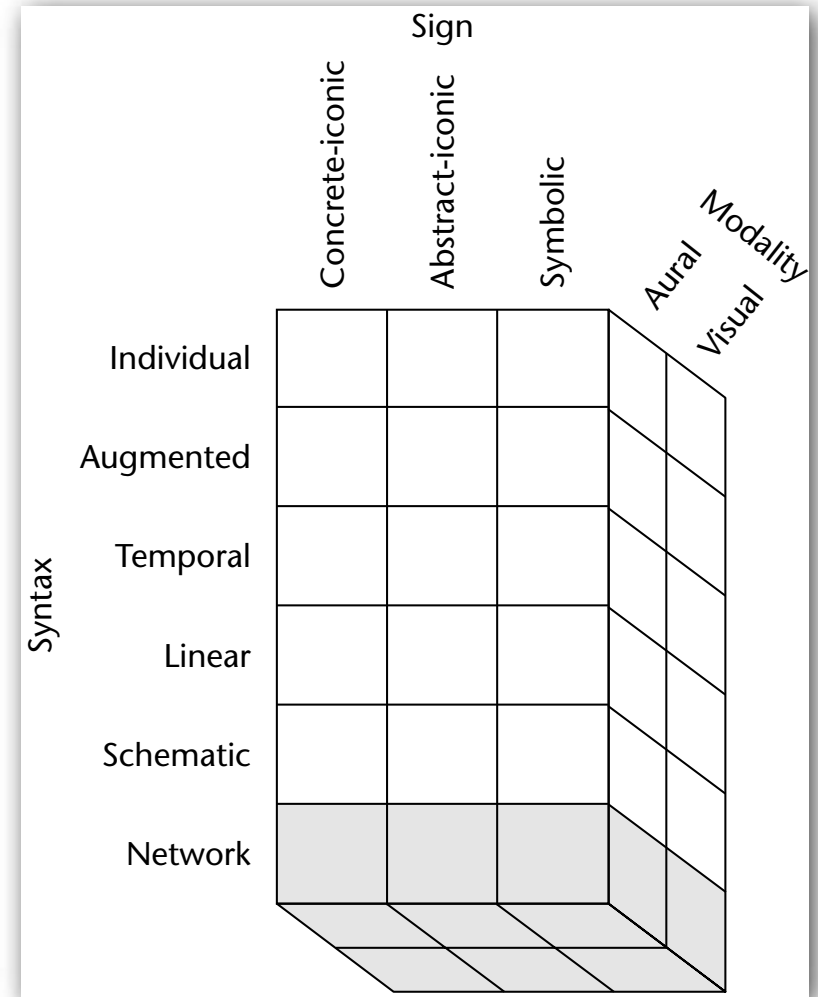
Forms of info. in human communication

- Information and symbols -- semiotics (記号論) --
  - Symbol, object, and concept
    - A symbol indicates a real object but often means a concept of that object.
      - Verbal expression, gestural expression, etc.
    - Messages or data are often composed of a set (sequence) of symbols.
      - Adequate understanding of symbols sent by a sender is important.
      - Understanding is based on cultural and/or common knowledge on the concepts.
      - It also requires good understanding of a sender's intention.
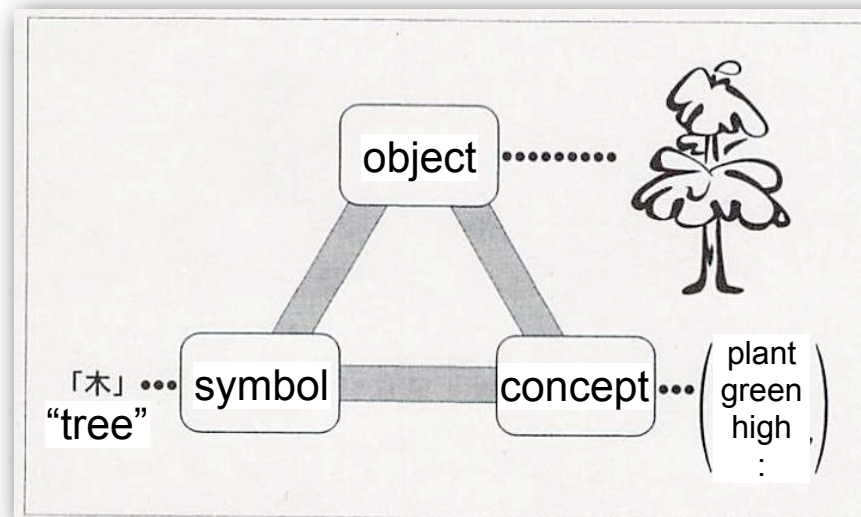
Forms of info. in human communication

- Qualitative aspect of information - intention and interpretation -
  - A message in the form of text
    - Interpretation often requires understanding the *context* of the message including a sender's intention as well as the (literal) *content* of the message.
  - "It's cold this morning."
    - From statement of a weather fact to the speaker's need of "I want a cup of hot coffee."
    - Proper interpretation of a message depends on the context where the message is made.
  - High-context language and low-context language
    - High-context : less verbally explicit communication, less written/formal information
    - "Can you pass me the salt?"   "Yes, I can."
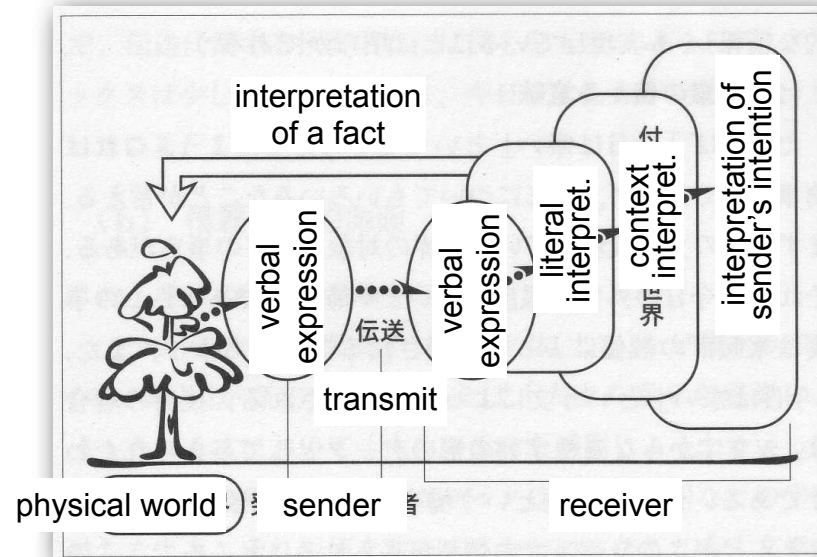
Forms of info. in human communication

- Context dependency of understanding a message
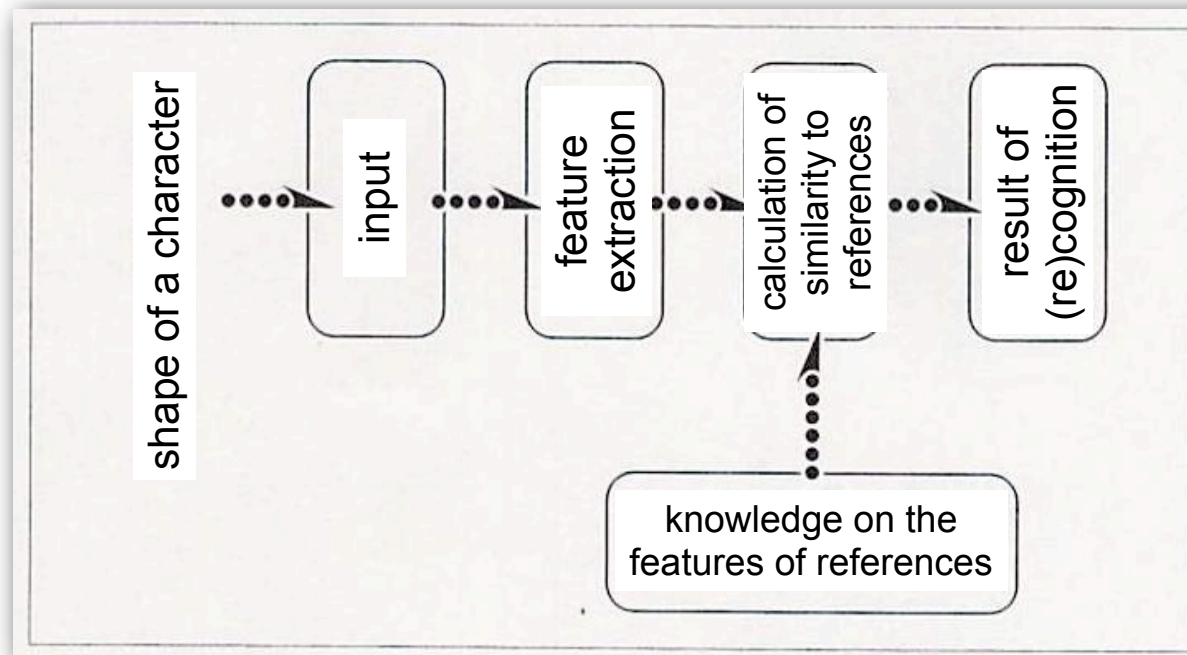 - "The lobster at no.18 is furious and about to burst into explosion."

Forms of info. in human communication

- Context dependency of understanding a message
  - "The lobster at no.18 is furious and about to burst into explosion."

The guest at table 18, who ordered a lobster, is very angry because the dish is not served yet.

Today's menu

- The term of "information" used in human communication.
  - Two kinds of definition of information (C. Shannon vs. this lecture)
  - Data and information - intention of a sender and interpretation of a receiver -
- Various forms of information in human communication
  - Classification of media information
  - Context dependency of information
- Information and knowledge
  - From data to information
    - Knowledge-based cognitive processing
  - Unconscious processing
    - Your brain creates your world but you cannot be aware of the brain's processing.
  - Various forms of information and conversion between them
    - Recognition and synthesis: abstraction and embodiment
  - Logical information and expressive (感性, KANSEI) information
    - Behaviors and information processing of autistics

From data to information

- Data (message), knowledge (memory), and information
  - Data (message) can become information only when it is interpreted adequately.
    - Interpretation of the context is also needed.
    - What makes interpretation possible?   Explicit and implicit *knowledge* is important!
- General framework of (re)cognition
  - Character recognition as example (a, a, a, *a*, *a*, **a**, etc)
  - We can perceive the abstract concept of "a" independently of font and glyph.

shape of a character ····▶ input ····▶ feature extraction ····▶ calculation of similarity to references ····▶ result of (re)cognition

knowledge on the features of references

From data to information

- How have we acquired knowledge?
  - Abstraction / generalization / induction from what is received as information.
    - A set of facts (instances) can be generalized into some (abstract) rules.
  - Information comes first or knowledge comes first?
    - Chicken-and-egg problem
  - All the required knowledge come from what one has experienced after birth?
    - Inheritance-based (inborn) knowledge and experience-based (acquired) knowledge
    - *Implicit* knowledge, which is often associated with *unconscious* processing
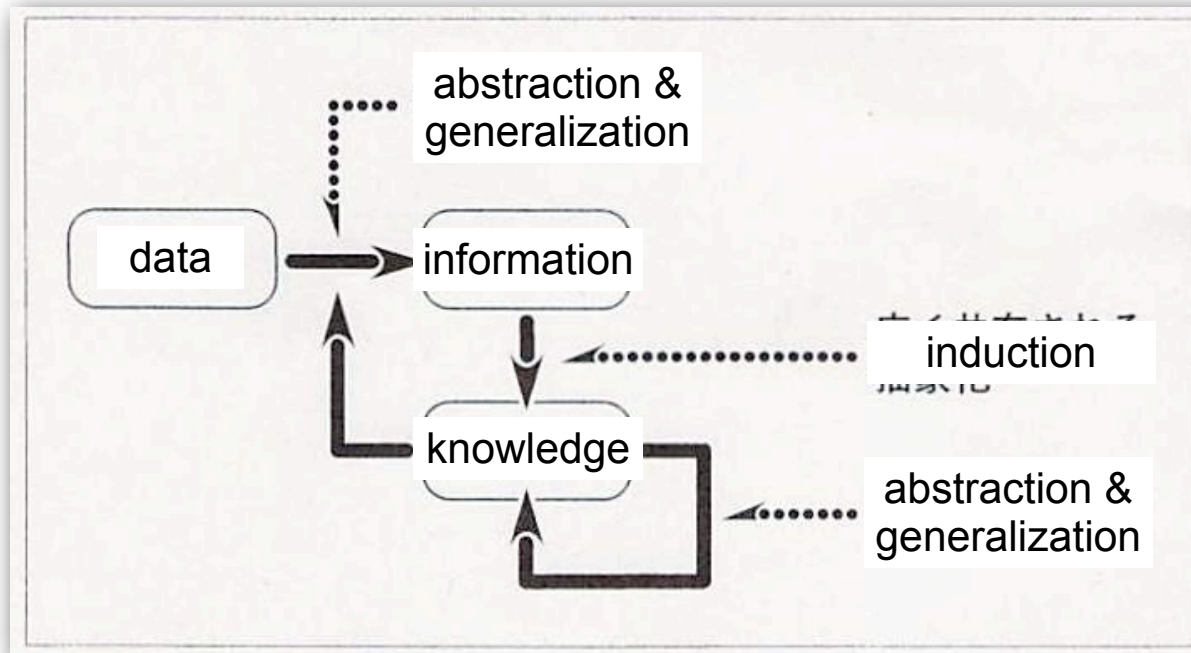
Implicit knowledge

- Unconscious processing
  - Difficult to notice consciously what is being done in the brain "unconsciously".
    - Ebbinghaus illusion
      - When you pick up one of the circles, is the distance bet. the two fingers different bet. the circles?
      - Your mind is easily tricked but your fingers *in action* are not tricked.
    - What-pathway and how-pathway in the vision system of the brain
      - A brain damage in the visual cortex makes "conscious" experiences of seeing impossible.
      - But blind individuals can behave properly according to the visual characteristics of nearby objects!
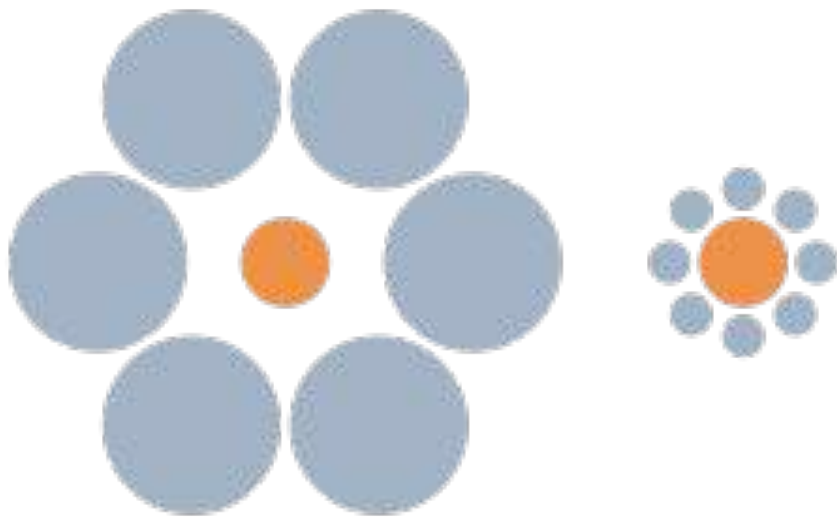
Ebbinghaus illusion
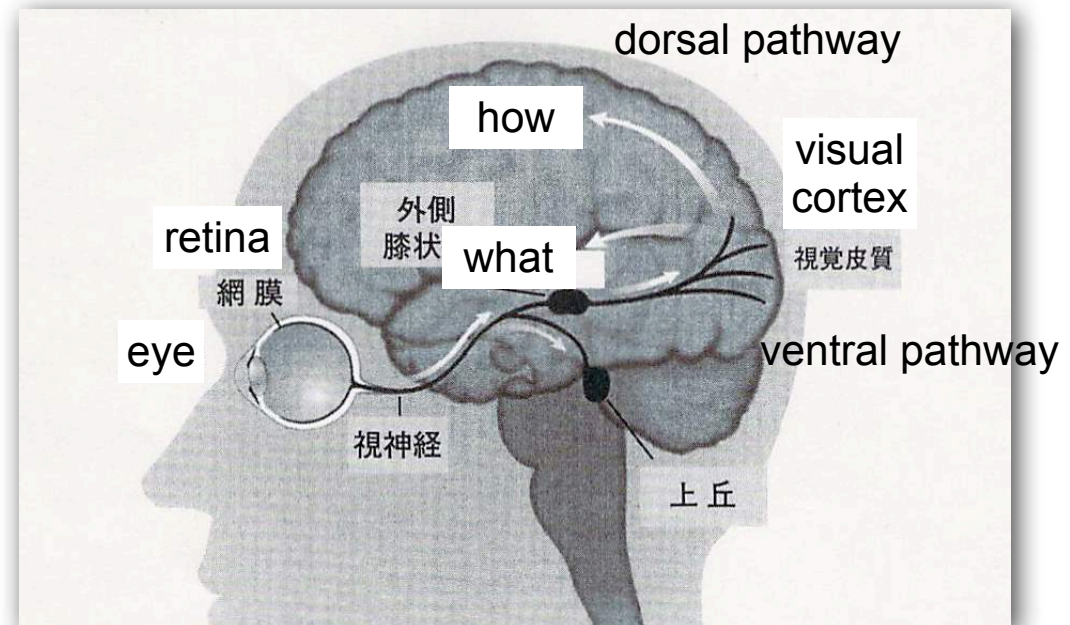
Implicit knowledge

- ## Unconscious processing
  - ### Blind sight [L. Weiskrantz'86]

posting task

correct direction

By visual inspection

Through action of posting

D・F Controls
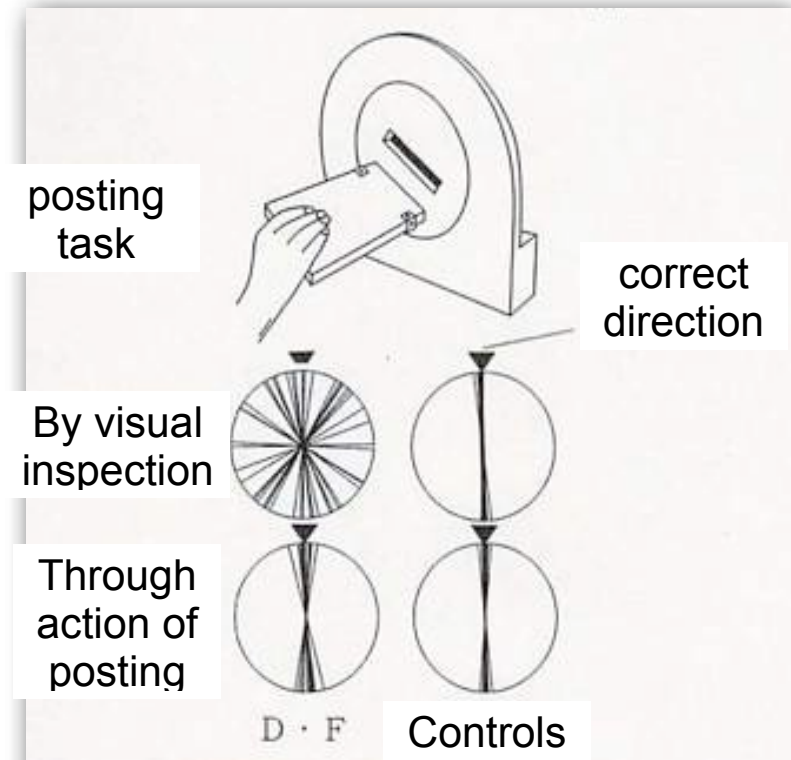
図 1・5 Unconscious action

D.F. has a severe brain damage on the visual cortex but no damage on the cortex associated with handling things. She cannot guess (consciously) the hole direction by visual inspection but can guess (unconsciously) through action of posting.
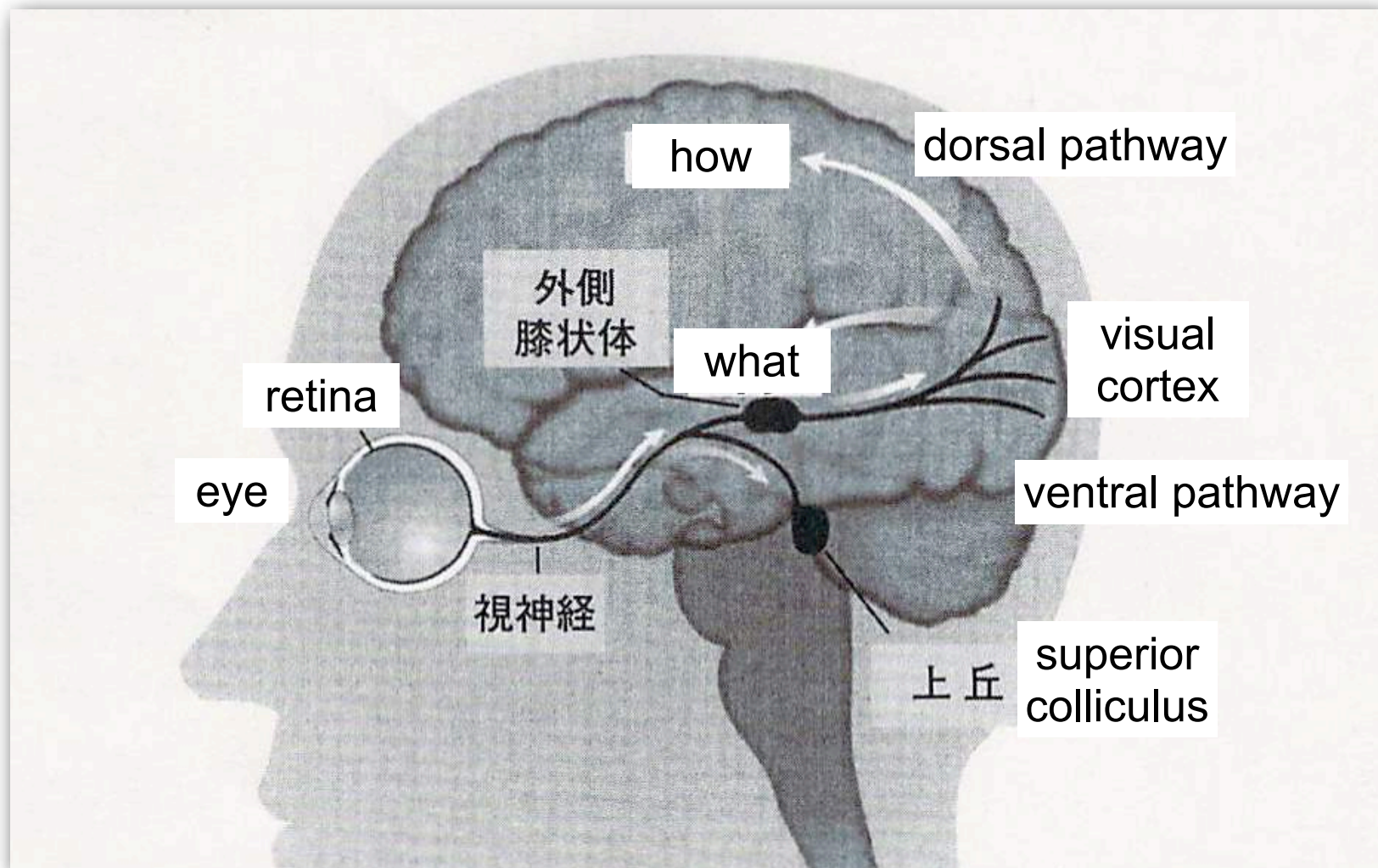
Implicit knowledge

- ## Unconscious processing
  - ### A blind monkey can see everything [N. Humphrey'72]

Implicit knowledge

- Unconscious processing
 - A possible mechanism to make "blind sight" possible.

Implicit knowledge

- Unconscious processing
  - Conscious world = what is created by the brain
    - If the brain has some damage, the conscious world changes drastically?   Yes!
  - Phantom limb (phantom leg, arm, and finger) [S. Mitchell1871]   (phantom = 亡霊)
    - A man who dose not has his thumb shouts "my thumb itches!"
    - Plasticity of the brain   (Plasticity = 可塑性)
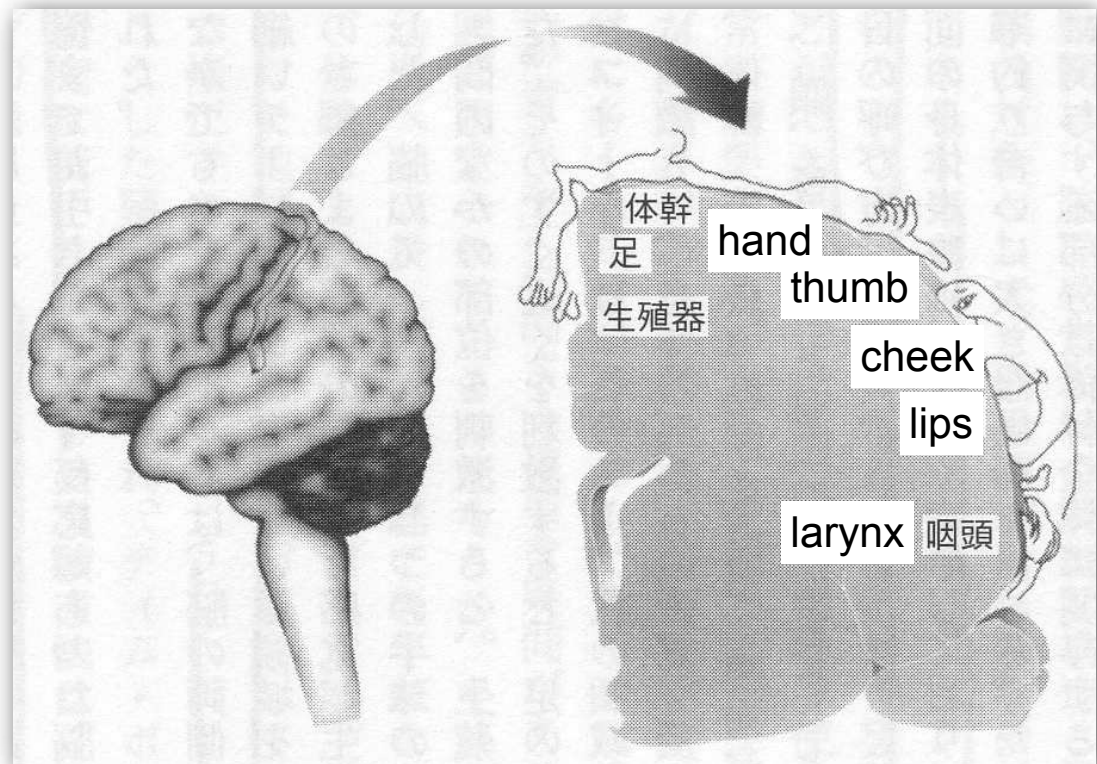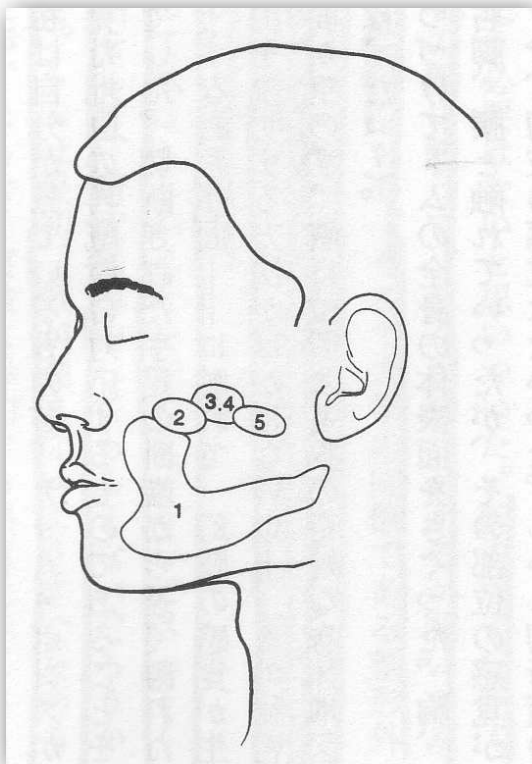
Implicit knowledge

- Unconscious processing
  - The blind spot on the retina and "filling-in" done by the brain
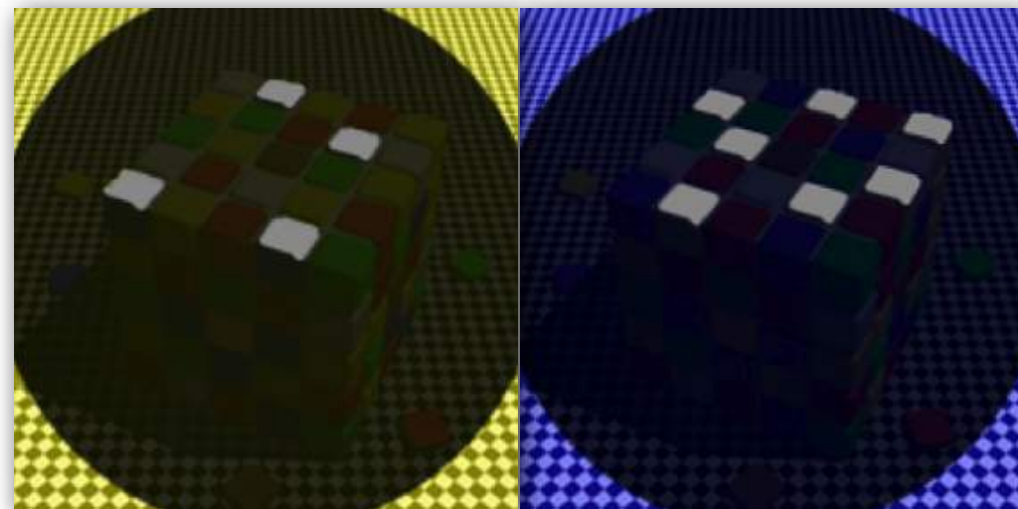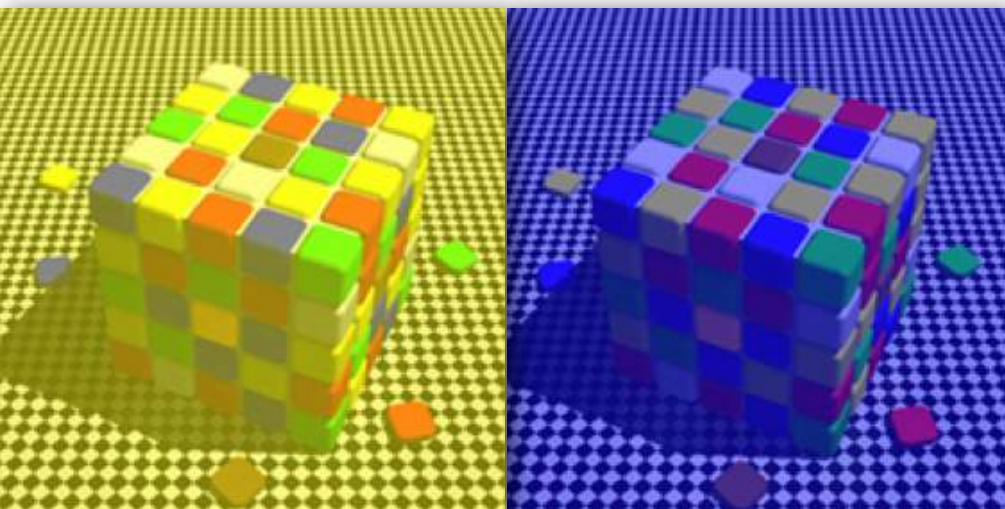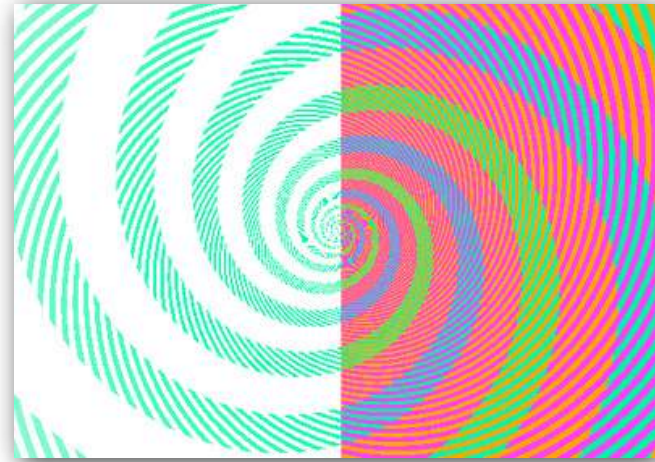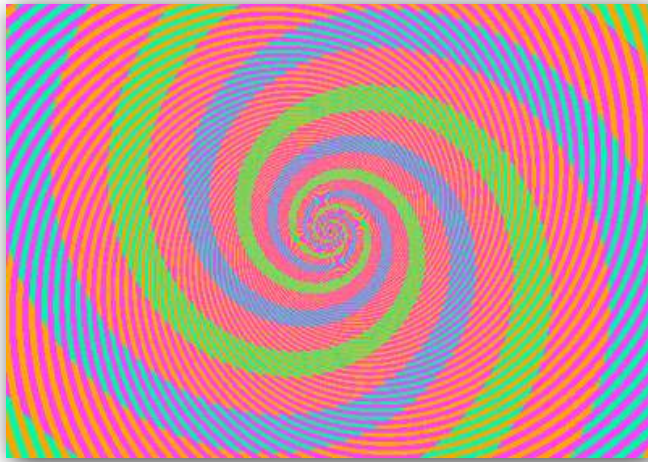   - Photoreceptors (視細胞) do not exist on a small region of the retina.
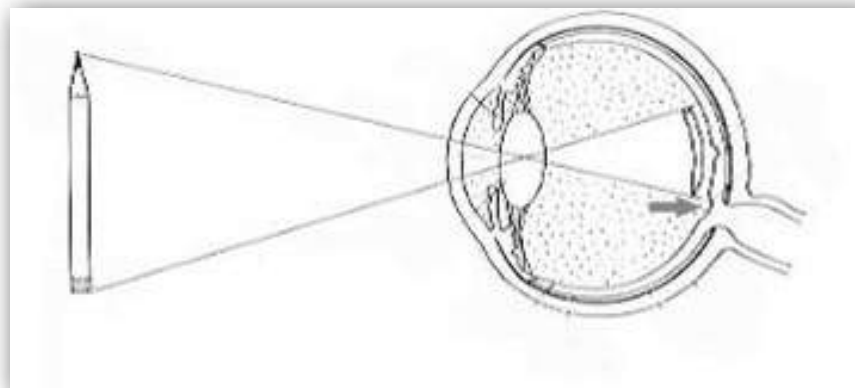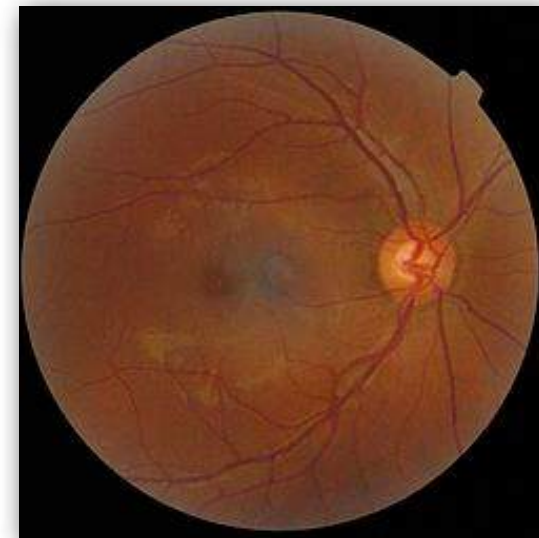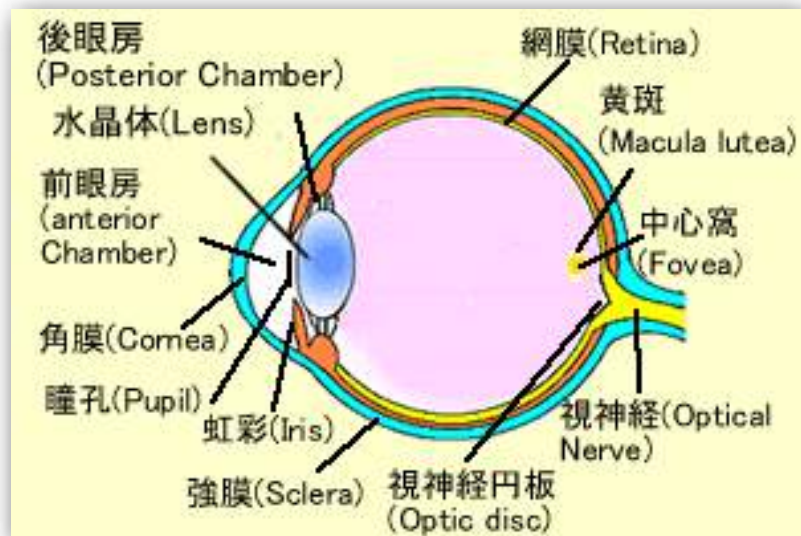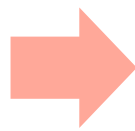

Wow! The star disappears!

Implicit knowledge

- Unconscious processing
  - Color illusions
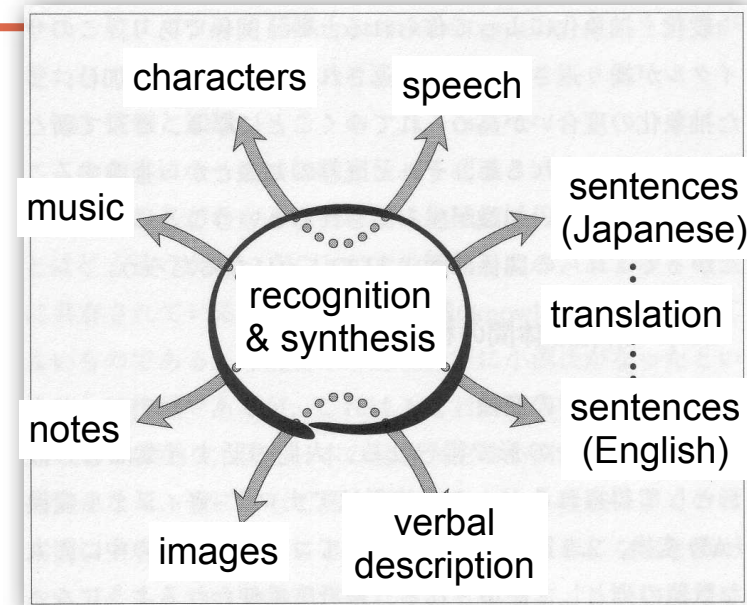
White and gold or blue and black?

Implicit knowledge

- How have we acquired knowledge?
  - Abstraction / generalization / induction from what is received as information.
    - A set of facts (instances) can be generalized into some (abstract) rules.
  - Information comes first or knowledge comes first?
    - Chicken-and-egg problem
  - All the required knowledge come from what one has experienced after birth?
    - Inheritance-based (inborn) knowledge and experience-based (acquired) knowledge
    - *Implicit* knowledge, which is often associated with *unconscious* processing
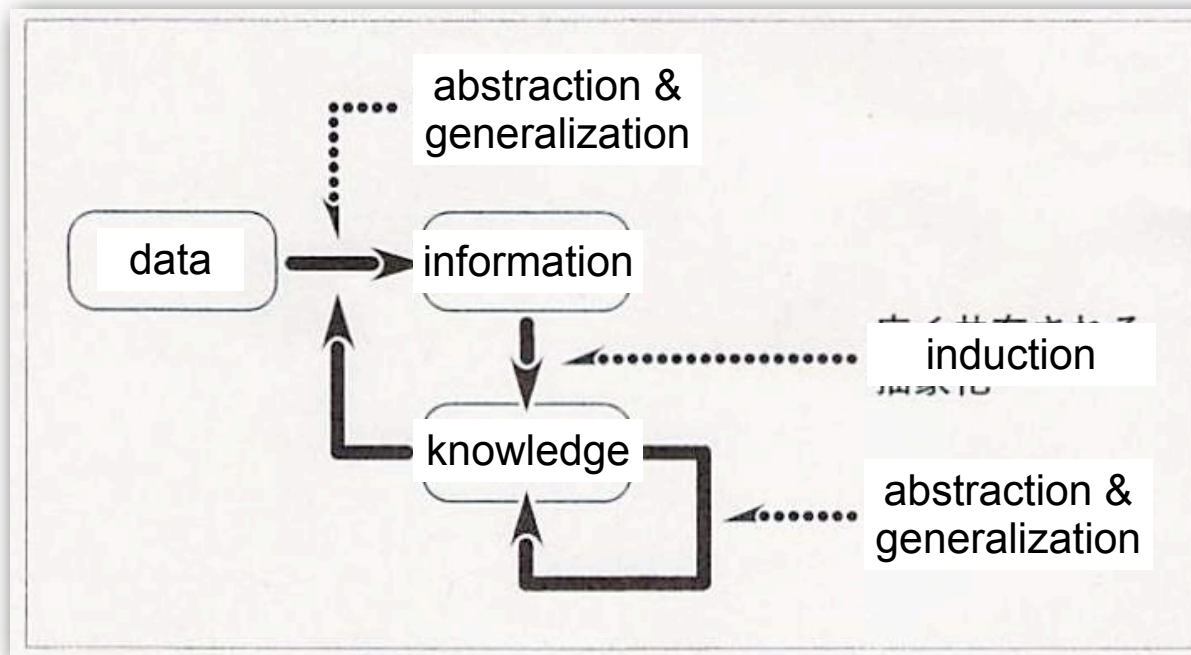
Various forms of information

- Media conversion for communication
  - character to symbol: character recognition
  - speech to symbol (character): speech recognition
  - symbol (character) to speech: speech synthesis
  - character to character: font conversion
  - sentence to sentence: language translation
  - speech to speech: spoken language translation
  - musical scores to music: automatic music performance
  - music to musical scores: automatic annotation of notes
  - real images to diagrammatic drawing: outline or edge extraction
  - real images to symbols: object recognition
  - diagrammatic drawing to images: automatic drawing of pictures or scenes
  - The same message can be represented in different ways.
    - The most effective use of media depends on the message and its context.

  - Conversion is done through two processes of recognition and synthesis.

characters   speech

music

sentences
(Japanese)

recognition
& synthesis

translation

notes

sentences
(English)

images

verbal
description

Various forms of information

- Media conversion for communication - abstraction and embodiment -
  - (Re)cognition or identification (ex: character recognition)
    - includes a process of removing irrelevant attributes attached to instances
    - Abstraction
  - Generation or synthesis (ex: character synthesis)
    - includes a process of adding back those attributes to realize instances
    - Embodiment

(re)cognition
(abstraction)

instances

concept

a, a, a, *a*, ɑ, **a**, etc

concept of "a"

synthesis
(embodiment)
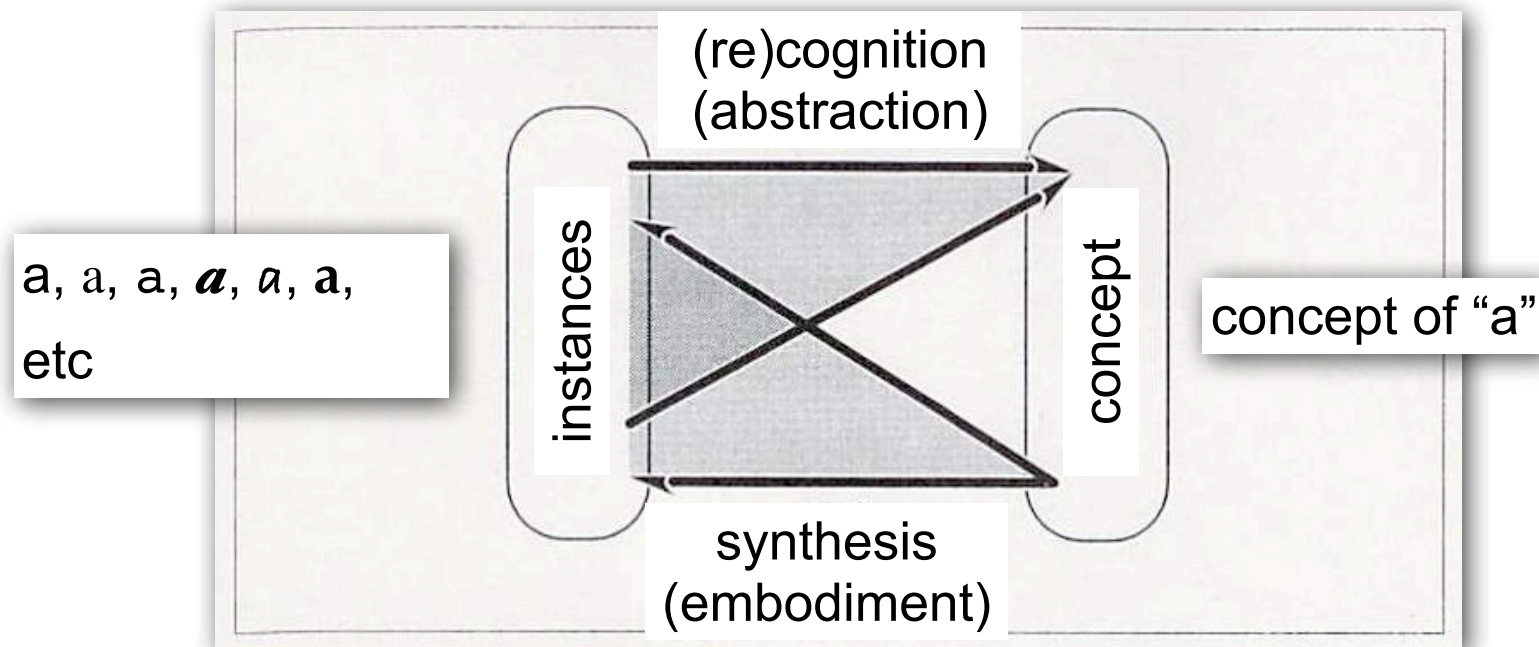
Logical and expressive

- Logical information and expressive information
  - Logical information
    - Interpretation does not depend on receivers, e.g. objective facts.
  - Expressive (KANSEI, 感性) information
    - Interpretation strongly depends on receivers, e.g. subjective impression.
    - Tastes differ (十人十色).



Is Tokyo the capital of Japan?

Which guy do you think is more handsome?

Logical and expressive

- Logical information and expressive information
  - Factors (bases) to describe expressive information
    - Facial expressions (as example)
      - 6 factors of surprise, fear, dislike, anger, happiness, and sorrow
      - A still debatable problem in psychology
    - Theory of mind [D. Premack et. al.'78]
      - The ability to attribute different mental states to oneself and others and to understand that others have different mental states than one's own.
        - Different individuals have different minds.
        - Those who don't have theory of mind have difficulty in understanding this fact.
      - One of the theories that explains the cause of autism (自閉症) [S. Baron-Cohen'91]
        - Difficulty in reading the mind of others and understanding that everybody has one's own mind.
        - Difficulty in reading the facial expressions.
        - Abnormality in information processing in the "old" brain.

高等哺乳類脳 higher mammal brain

下等哺乳類脳 lower mammal brain

爬虫類脳 reptile brain
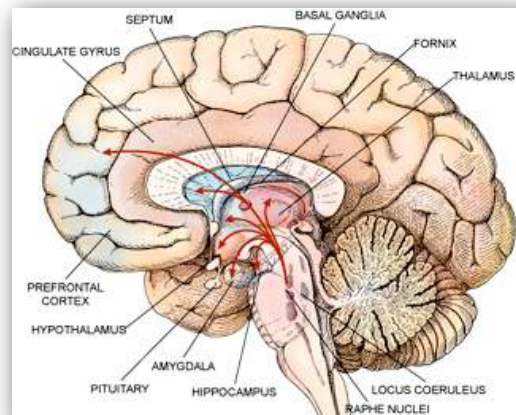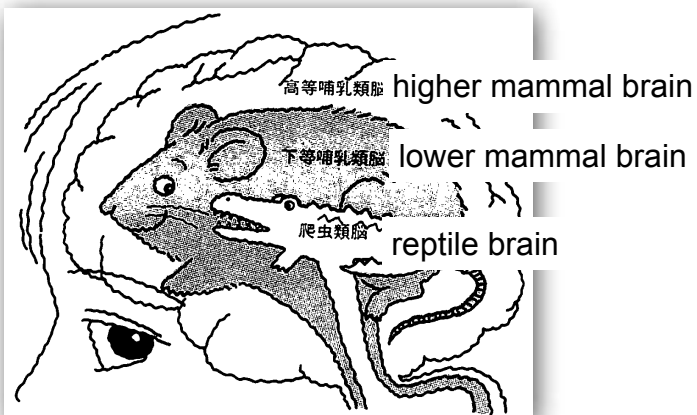
a, a, a, *a*, a, **a**, etc

Forms of info. in human communication

- Context dependency of information
  - "The lobster at no.18 is furious and about to burst into explosion."

The guest at table 18, who ordered a lobster, is very angry because the dish is not served yet.

相手が話す
文章の意味のほかに
も、相手が伝えようとす
る真意をくみとらな
ければねえ。

相手がいうこと
を理解するには、単
語や文をわかるだけでは
だめなのだ。

if only i could read your mind

"Can you pass me the salt?"
"Yes, I can."

Forms of info. in human communication

Today's menu

- The term of "information" used in human communication.
  - Two kinds of definition of information (C. Shannon vs. this lecture)
  - Data and information - intention of a sender and interpretation of a receiver -
- Various forms of information in human communication
  - Classification of media information
  - Context dependency of information
- Information and knowledge
  - From data to information
    - Knowledge-based cognitive processing
  - Unconscious processing
    - Your brain creates your world but you cannot be aware of the brain's processing.
  - Various forms of information and conversion between them
    - Recognition and synthesis: abstraction and embodiment
  - Logical information and expressive (感性, KANSEI) information
    - Behaviors and information processing of autistics

Title of each lecture

- Theme-1
  - Multimedia information and humans
  - Multimedia information and interaction between humans and machines
  - Multimedia information used in expressive and emotional processing
  - A wonder of sense - synesthesia -
- Theme-2
  - Speech communication technology - articulatory & acoustic phonetics -
  - Speech communication technology - speech analysis -
  - Speech communication technology - speech recognition -
  - Speech communication technology - speech synthesis -
- Theme-3
  - A new framework for "human-like" speech machine #1
  - A new framework for "human-like" speech machine #2
  - A new framework for "human-like" speech machine #3
  - A new framework for "human-like" speech machine #4

Recommended books