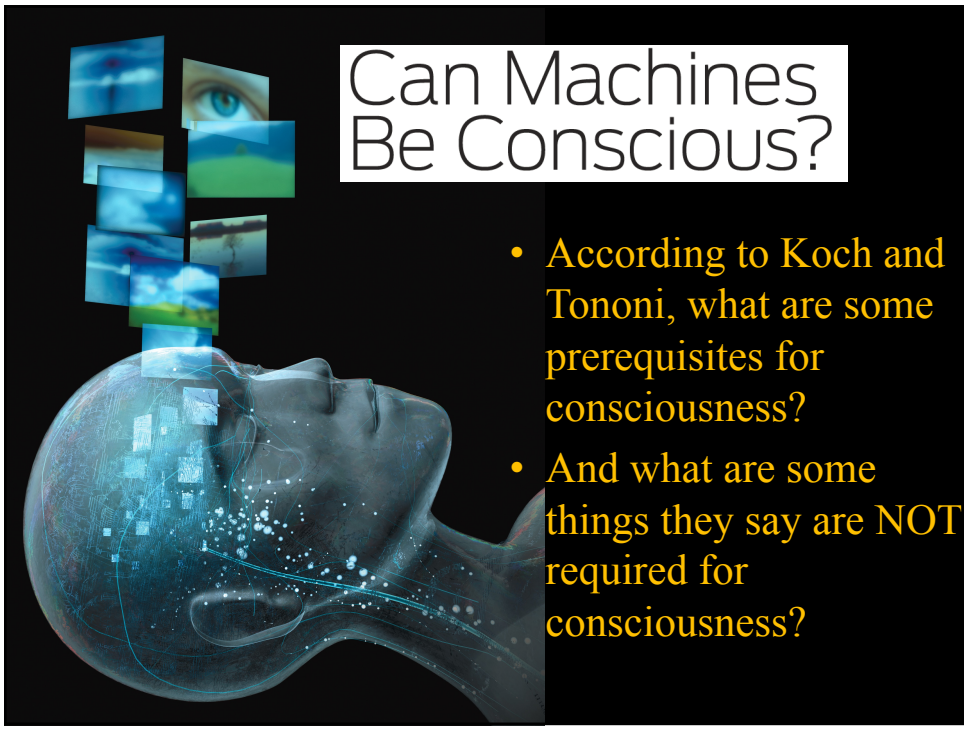
1

Introduction to Consciousness

in mammals and machines

1. Attention and consciousness: what are we talking about?
2. "Can machines be conscious?"
    1. Some functions NOT necessary for consciousness
    2. Meet the new Turing Test (same as the old)
    3. The importance of being integrated
3. Epiphenomenon or useful function?
4. Chinese Room Argument: is understanding (or consciousness in any form) reducible to an algorithm?

2



Information is:

how many possibilities get ruled out

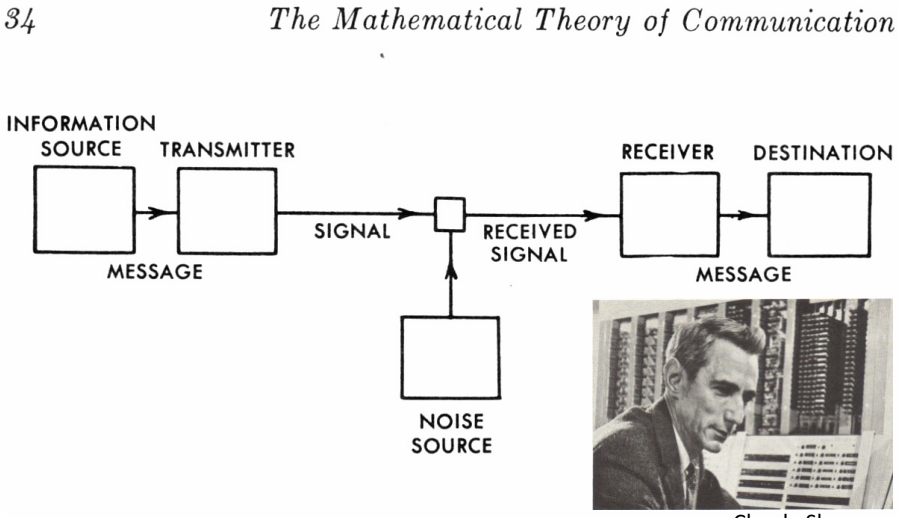
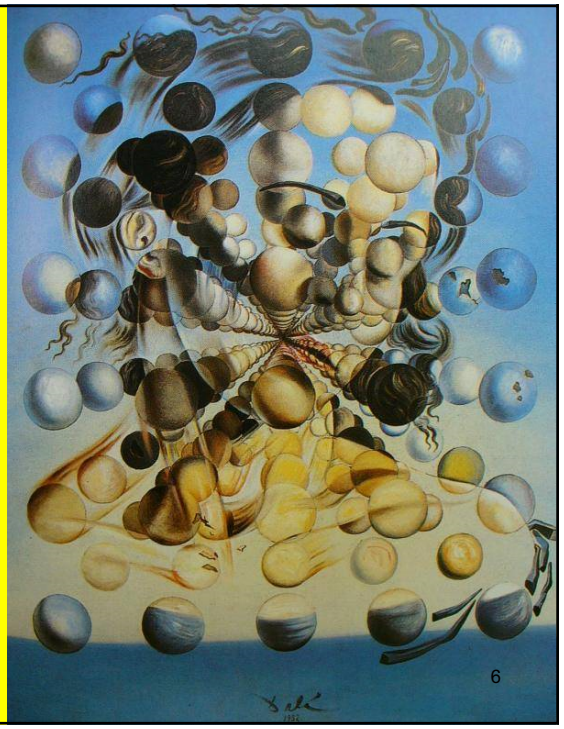


Fig. 1. — Schematic diagram of a general communication system.

Claude Shannon
1916-2001

The Binding Problem

"Exactly how the parallel streams of sensory data are melded into perception, images, and ideas remains the Holy Grail of neuroscience."



Consciousness as Integrated Information: a Provisional Manifesto

GIULIO TONONI

Department of Psychiatry, University of Wisconsin, Madison, Wisconsin

Abstract. The integrated information theory (IIT) starts from phenomenology and makes use of thought experiments to claim that consciousness is integrated information. Specifically: (i) the quantity of consciousness corresponds to the amount of integrated information generated by a complex of elements; (ii) the quality of experience is specified by the set of informational relationships generated within that complex. Integrated information (Φ) is defined as the amount of information generated by a complex of elements, above and beyond the information generated by its parts. Qualia space (Q) is a space where each axis represents a possible state of the complex, each point is a probability distribution of its states, and arrows between points represent the informational relationships among its elements generated by causal mechanisms (connections). Together, the set of informational relationships within a complex constitute a shape in Q that completely and univocally specifies a particular experience. Several observations concerning the neural substrate of consciousness fall naturally into place within the IIT framework. Among them are the association of consciousness with certain neural systems rather than with others; the fact that neural processes underlying consciousness can influence or be influenced by neural processes that remain unconscious; the reduction of consciousness during dreamless sleep and generalized seizures; and the distinct role of different cortical architectures in affecting the quality of experience. Equating consciousness with integrated information carries several implications for our view of nature.

Reference: *Biol. Bull.* **215:** 216–242. (December 2008)

7

WIRED: The internet is integrated. Could it be conscious?

Koch: It's difficult to say right now. But consider this. The internet contains about 10 billion computers, with each computer itself having a couple of billion transistors in its CPU. So the internet has at least 10^19 transistors, compared to the roughly 1000 trillion (or quadrillion) synapses in the human brain. That's about 10,000 times more transistors than synapses. But is the internet more complex than the human brain? It depends on the degree of integration of the internet.

For instance, our brains are connected all the time. On the internet, computers are packet-switching. They're not connected permanently, but rapidly switch from one to another. But according to my version of panpsychism, it feels like something to be the internet — and if the internet were down, it wouldn't feel like anything anymore. And that is, in principle, not different from the way I feel when I'm in

From "A Neuroscientist's Radical Theory of How Networks Become Conscious 2013

Introduction to Consciousness

in mammals and machines

1. Attention and consciousness: what are we talking about?
2. "Can machines be conscious?"
    1. Some functions NOT necessary for consciousness
    2. Meet the new Turing Test (same as the old)
    3. The importance of being integrated
3. Epiphenomenon or useful function?
4. Chinese Room Argument: is understanding (or consciousness in any form) reducible to an algorithm?

10

Epiphenomenalism

Mental events
are caused by physical events in the brain,
but have no effects
upon any physical events.

11

"The consciousness of brutes would appear to be related to the mechanism of their body simply as a collateral product of its working, and to be completely without any power of modifying that working, as the steam-whistle which accompanies the work of a locomotive engine is without influence upon its machinery." (Thomas Huxley, quoted in James, 1890, p.135).

"The particulars of the distribution of consciousness, so far as we know them, point to its being efficacious ... it seems an organ, superadded to other organs which maintain the animal in the struggle for existence; and the presumption of course is that it helps him in some way in the struggle ..." (James, 1890).

12

William James' evolutionary argument against epiphenomenalism

- Conscious states correspond to our situation w.r.t. surviving and thriving (e.g. fire hurts), suggesting they are adaptations that evolved
- But to evolve by natural selection they must have effects on physical behavior

13

So what is consciousness for?

- Provides an "**executive summary**" of the current situation, useful for learning and planning
- For dealing with novel situations where unconscious specialist modules don't know what to do, or disagree with each other (cf. "advisors" in SemaFORR robot)

14

Bernard Baars

Broadcasting consciousness

The global workspace theory suggests that consciousness arises from highly coordinated, widespread activity in the brain

PERCEPTIONS

Sight, sound, taste and touch are first processed in small, localised areas of the brain

CONSCIOUS

When signals are broadcast to a wider network of neurons across much of the cortex - the global workspace - we become conscious of the sensation

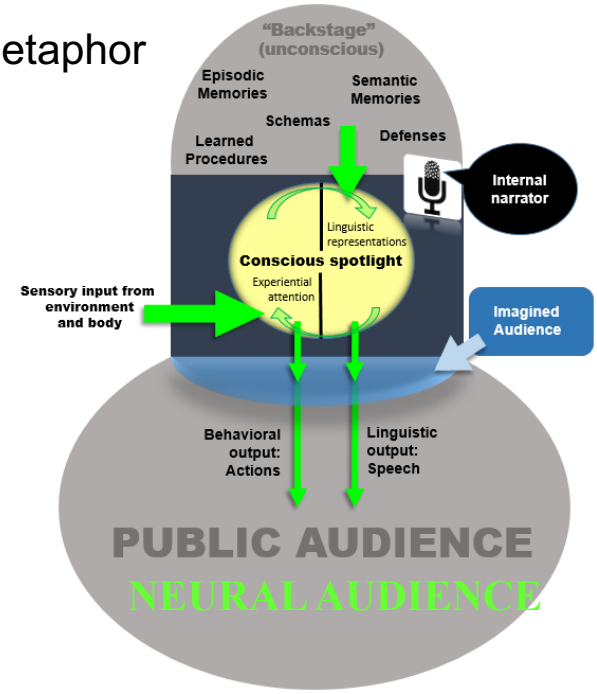
NON-CONSCIOUS

When signals remain localised, the associated sensations are not perceived consciously

Functional metaphor: BLACKBOARD used by a committee of experts

15

- Theatre metaphor



16

Introduction to Consciousness

in mammals and machines

1. Attention and consciousness: what are we talking about?
2. "Can machines be conscious?"
    1. Some functions NOT necessary for consciousness
    2. Meet the new Turing Test (same as the old)
    3. The importance of being integrated
3. Epiphenomenon or useful function?
4. Chinese Room Argument: is understanding (or consciousness in any form) reducible to an algorithm?

17

Searle's Chinese Room Argument against "the mind is the program"

- Put me in a room with Chinese character cards and rule book
- Receive cards through window and follow rules to output cards that answer questions on input cards
- I implement the program, but I do not understand Chinese
- Therefore, understanding Chinese is more than **just** a program or function

18

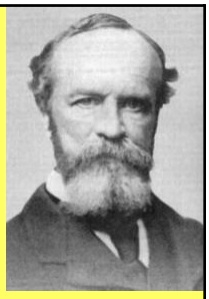
19

Attention!

1. Definitions and behavioral effects
2. Effects on neural firing rates:
   - Spatial attention
   - Attention to features
3. Directing attention:
   - Posterior parietal cortex
   - Frontal eye fields
   - Top-down and bottom-up attention

20

Attention chooses what enters consciousness



"Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others…"

—William James
Principles of Psychology 1890

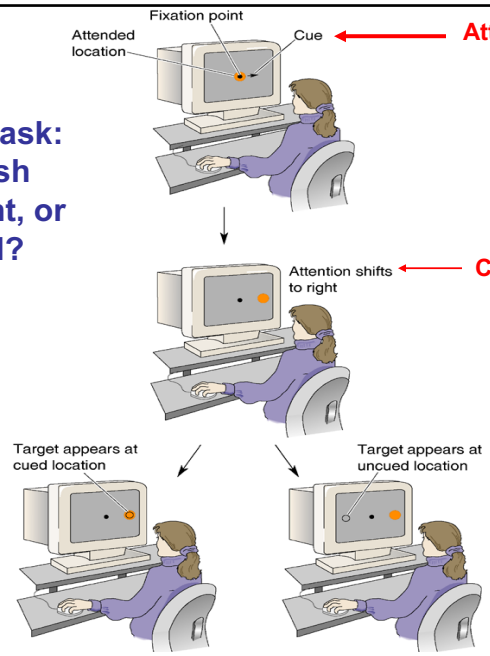
21

Attention

A state of selectively processing
some sources of perceptual information
while ignoring others
or
A process that selectively allocates a
limited capacity cognitive resource

22

Detection task: target flash on left, right, or not at all?

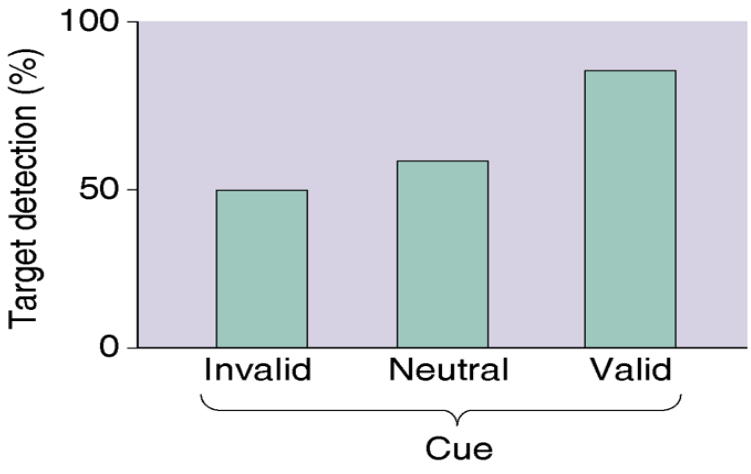


Attentional cue: 80% valid

Covert attention shift: no eye movement

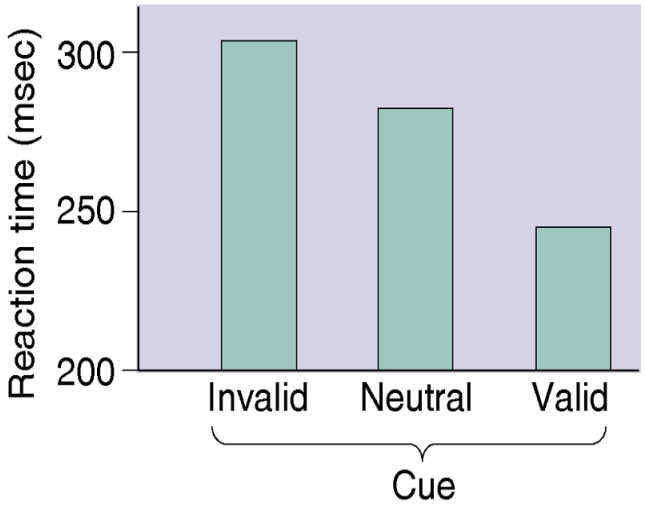
23

Attention enhances detection



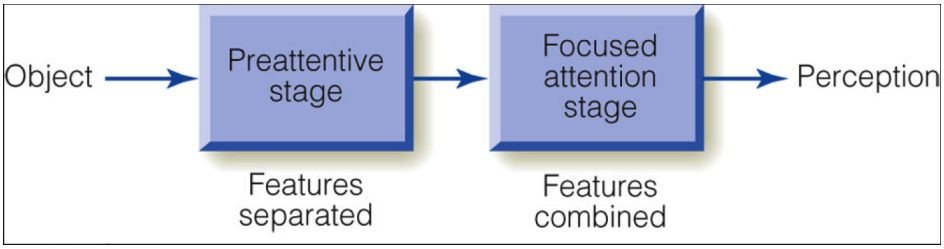
24

Attention increases reaction speed



25

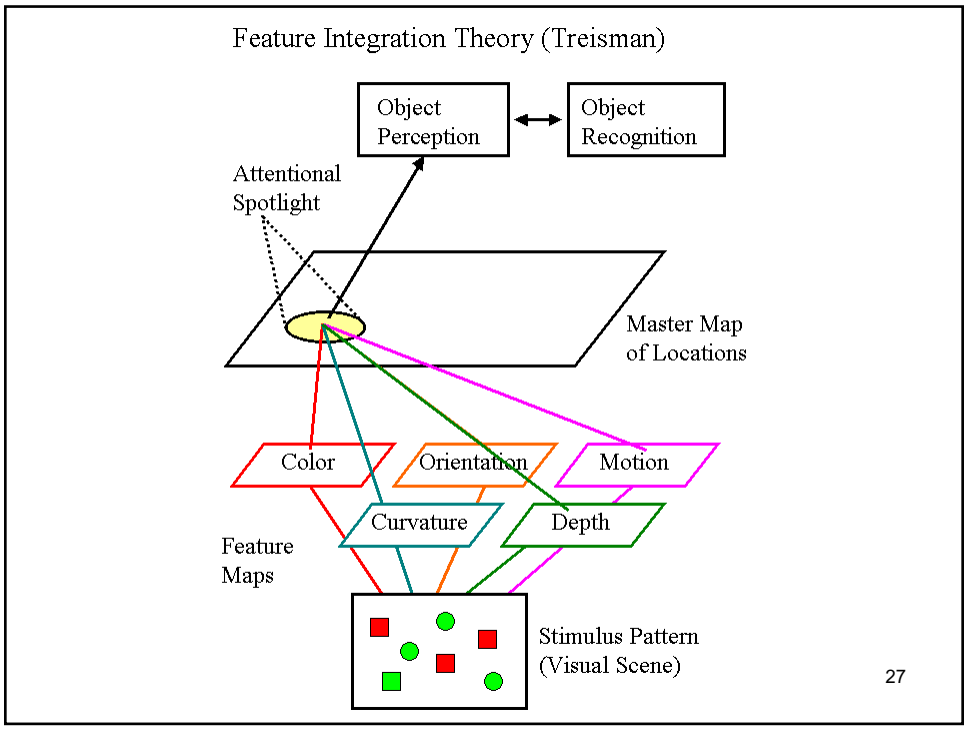
Treisman's Feature Integration Theory (Lab 9 visual search)

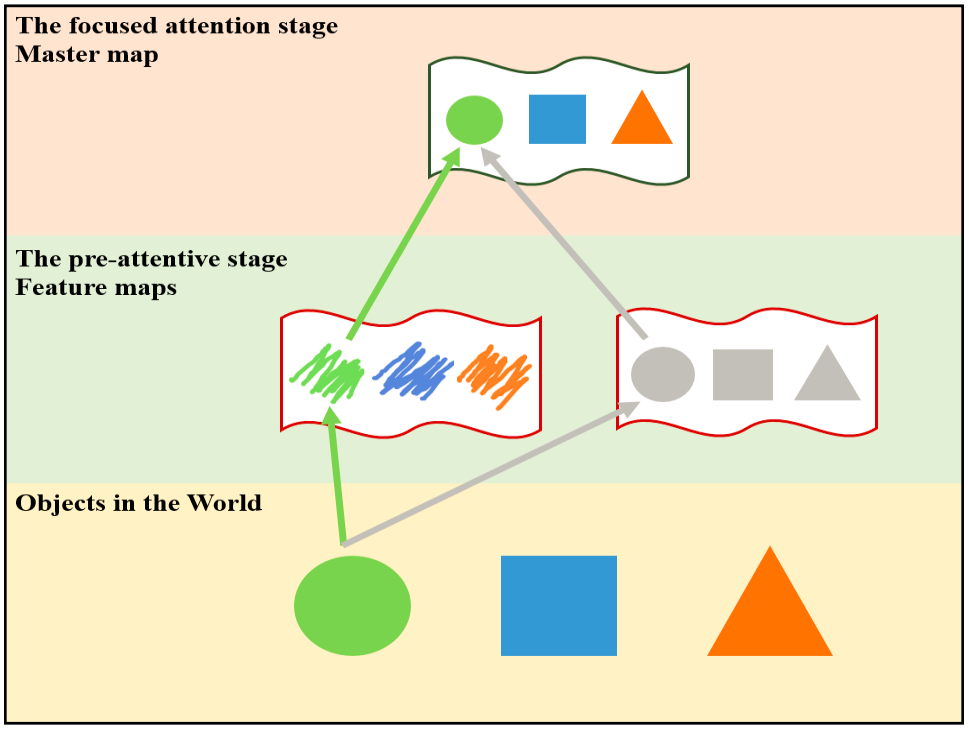


Decomposed
Analyzed

Synthesized
Integrated

26


Feature Integration Theory (Treisman)
Object Perception
Object Recognition
Attentional Spotlight
Master Map of Locations
Color
Orientation
Motion
Curvature
Depth
Feature Maps
Stimulus Pattern (Visual Scene)
27


The focused attention stage
Master map
The pre-attentive stage
Feature maps
Objects in the World

"Illusory conjunctions" when attention is taxed

29

Bottom-up

- Stimulus-driven: salience, automatic
- Feed-forward neural projections
- Faster

- e.g.
  Loud noise, flash of light

POPOUT!

Top-down

- Goal-driven: voluntary, effortful
- Feed-back neural projections
- Slower

- e.g.
  Looking for keys

CONJUNCTION SEARCH

30