Robot Perception

ESE 6800 / CIS 7000

Antonio Loquercio

What is Perception

Safe Learning in Robotics: From Learning-Based Control to Safe Reinforcement Learning

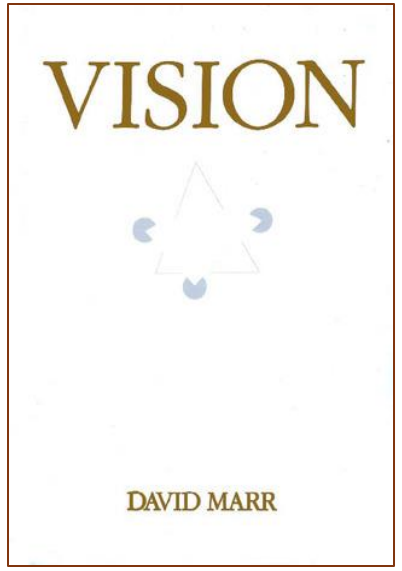
Lukas Brunke*, Melissa Greeff*, Adam W. Hall*, Zhaocong Yuan*, Siqi Zhou*, Jacopo Panerati, and Angela P. Schoellig

vocabulary and introducing benchmarks for algorithm evaluation that can be leveraged by both (18, 19). Our target audience are researchers, with either a control or RL background, who are interested in a concise but holistic perspective on the problem of safe learning control. While we do not cover perception, estimation, planning, or multi-agent systems, we do connect our discussion to these additional challenges and opportunities.

David Marr (1945-1980)

- Ph.D. in theoretical neuroscience, Cambridge, 1969
  - Models of the cerebellum (1969), neocortex (1970), hippocampus (1971)
- Joined MIT AI Lab in 1973, became professor of psychology in 1977
  - Stereo algorithms (with Tommaso Poggio), 1976-79
  - 3D object representation (with Keith Nishihara), 1978
  - Edge detection (with Ellen Hildreth), 1980
- Posthumous book: Vision (1982)

Bio



Full text

In December 1977, certain events occurred that forced me to write this book a few years earlier than I had planned. Although the book has important gaps, which I hope will soon be filled, a new framework for studying vision is already clear and supported by enough solid results to be worth setting down as a coherent whole.

Slide by Lana Lazebnik

Marr's Motivation (Ch. 1)

- Vision is *hard*

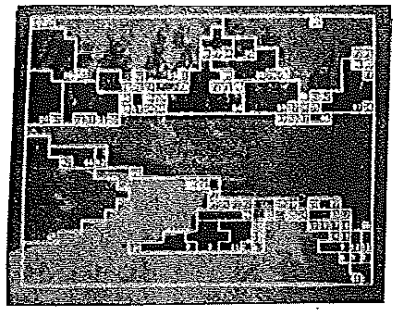
The first great revelation was that the problems are difficult. Of course, these days this fact is a commonplace. But in the 1960s almost no one realized that machine vision was difficult. The field had to go through the same experience as the machine translation field did in its fiascoes of the 1950s before it was at last realized that here were some problems that had to be taken seriously. The reason for this misperception is that we humans are ourselves so good at vision. The notion of a feature detector was well established by Barlow and by Hubel and Wiesel, and the idea that extracting edges and lines from images might be at all difficult simply did not occur to those who had not tried to do it. It turned out to be an elusive problem: Edges that are of critical importance from a three-dimensional point of view often cannot be found at all by looking at the intensity changes in an image. Any kind of textured image gives a multitude of noisy edge segments; variations in reflectance and illumination cause no end of trouble; and even if an edge has a clear existence at one point, it is as likely as not to fade out quite soon, appearing only in patches along its length in the image. The common and almost despairing feeling of the early investigators like B.K.P. Horn and T.O. Binford was that practically anything could happen in an image and furthermore that practically everything did.

Slide by Lana Lazebnik

Marr's Motivation (Ch. 1)

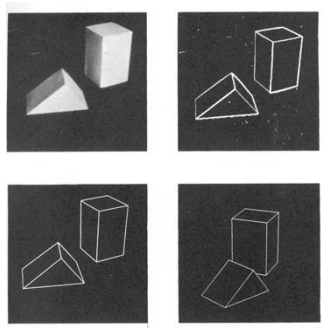
- Vision is *hard*
- We may not be able to figure out the right solution right away, but at least we should start by establishing a sound methodology
  - Marr explicitly considered and rejected low-level neurophysiology, empirical "hacking", and blocks world simplification

Hubel & Wiesel (1959) Source



(B-2) Output of the non-semantic weakest boundary melted first region grower.

Yakimovsky & Feldman (1973)



Roberts (1963)

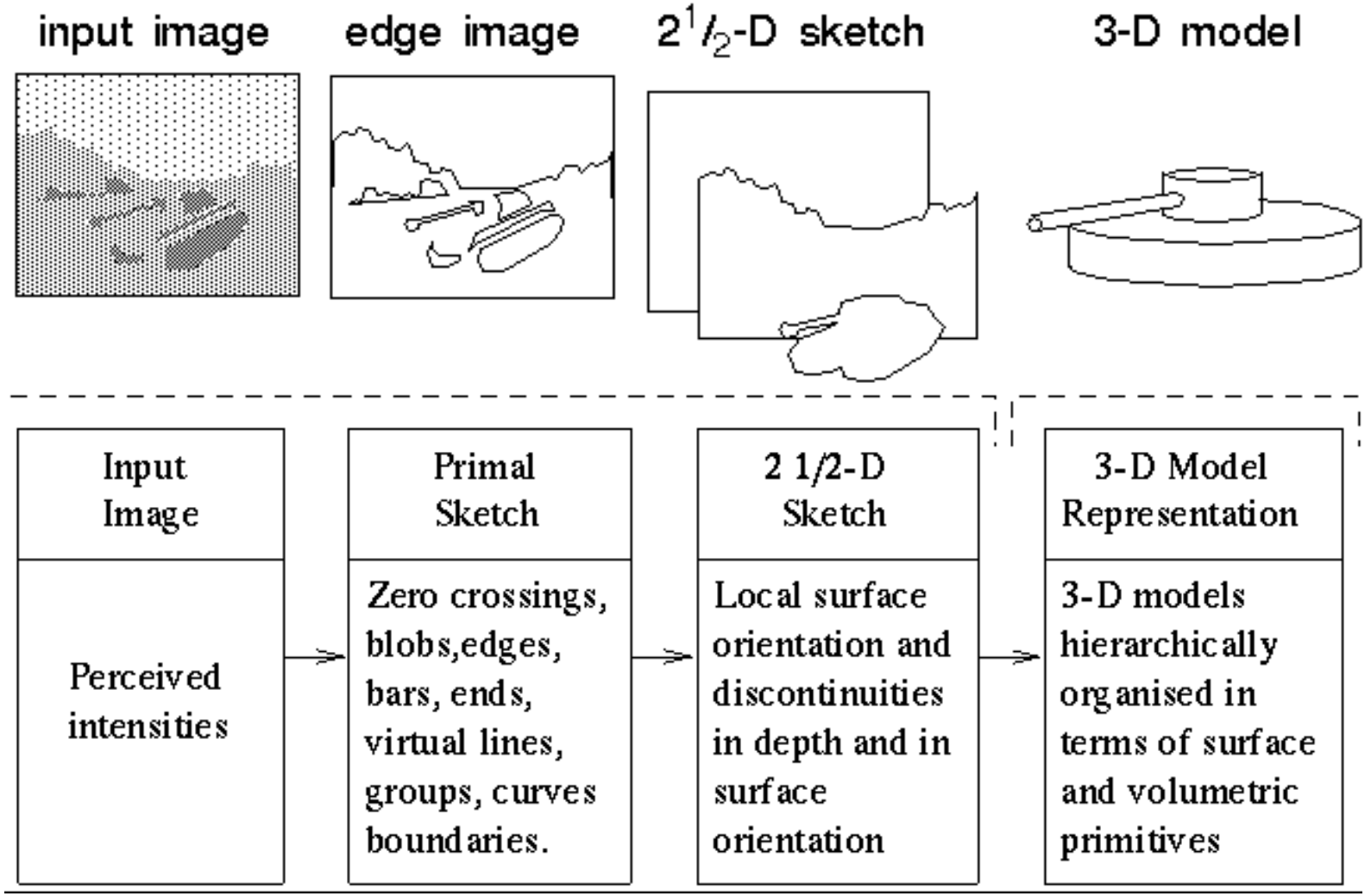
Slide by Lana Lazebnik

An *information processing* theory of vision

Computational theory	Representation and algorithm	Hardware implementation
What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?	How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?	How can the representation and algorithm be realized physically?

Figure 1–4. The three levels at which any machine carrying out an information-processing task must be understood.

Proposed algorithmic pipeline



So, what's the big deal?

- Marr's book was a major milestone
  - Critical summary of key developments in study of human and computer vision to date
  - Unprecedented attempt at a unified account of the entire visual system
- Computational framework was very appealing to computer vision researchers from a "software engineering" perspective
  - Abstraction, modularity, feedforward pipeline
- Theories meshed well with the dominant computer vision paradigms
  - Vision as "inverse graphics" or "inverse optics"
  - Emphasis on recovery of general-purpose 3D representations composed of simple geometric primitives
  - Convenient division of vision problems into "low-level", "mid-level", and "high-level"

Special issue dedicated to Marr: Perception 41(9), 2012

Slide by Lana Lazebnik

What about the bad stuff?

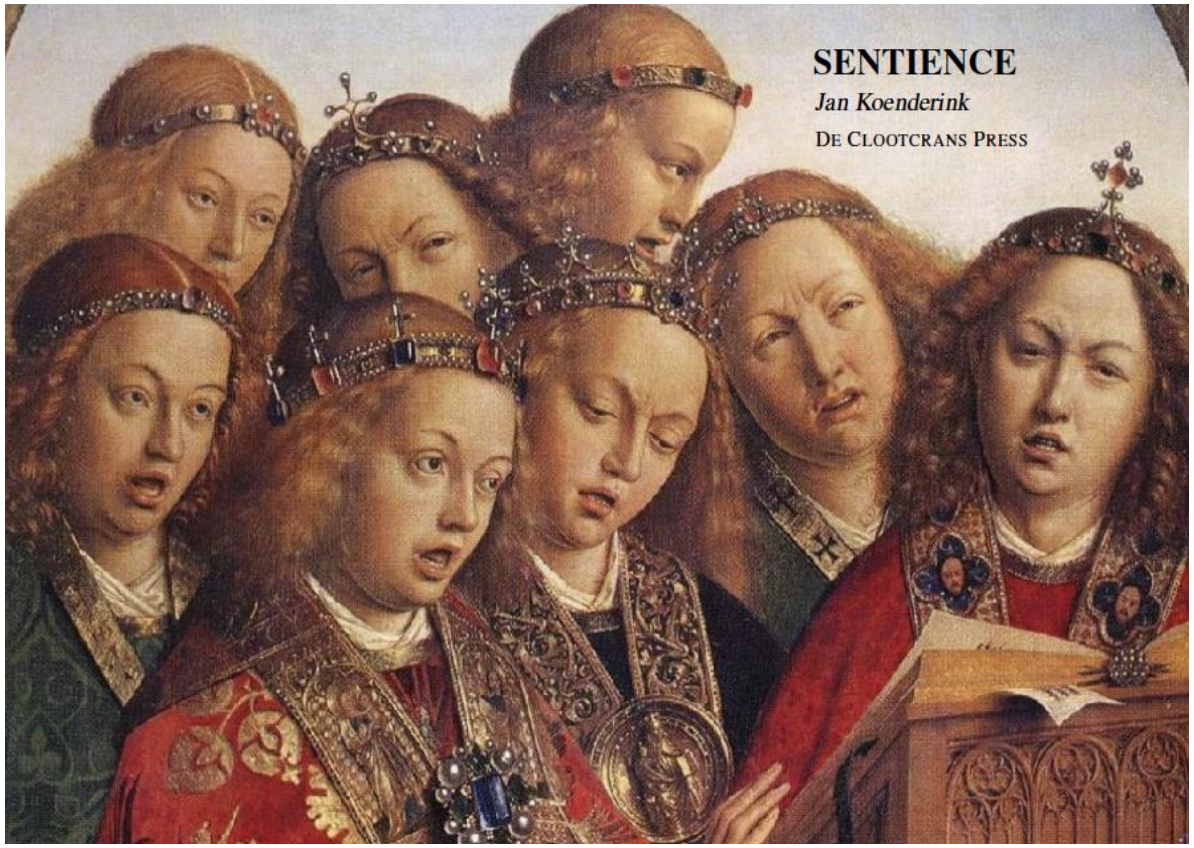
- None of the particulars of Marr's approach have panned out either on the human or the computer vision side
- Principles of modularity and feedforward processing don't hold for human vision
  - P. Churchland, V.S. Ramachandran, and T. Sejnowski, A critique of pure vision, 1994
- Humans do not recover veridical, task-independent 3D representations
  - W. Warren, Does This Computational Theory Solve the Right Problem? Marr, Gibson, and the Goal of Vision, Perception 41(9), 2012
- Marr dismissed statistical approaches, did not even consider learning
- Even the goals, inputs, and outputs of a vision system are very much open to question (as discussed next)

Yes, we likely throw away a lot

What we think we see

What we really see

A Grand Theory of Perception



J. Koenderink, Sentience, 2019

- Treat Koenderink as talking to your advisor:
  - 80% of what they say is nonsense, but 20% is brilliant
  - It's your job to find **which** 20%
  - With Koenderink, it might be as high as 45%!

Slide by Alexei Efros

Brave thing to study…

About The Clootcrans Press

The Clootcrans Press is a selfpublishing initiative of Jan Koenderink. Notice that the publisher takes no responsibility for the contents, except that he gave it an honest try—as he always does. Since his books are free you should have no reason to complain.

Slide by Alexei Efros

A grand theory of perception?

Heavily influenced by Jakob von Uexküll

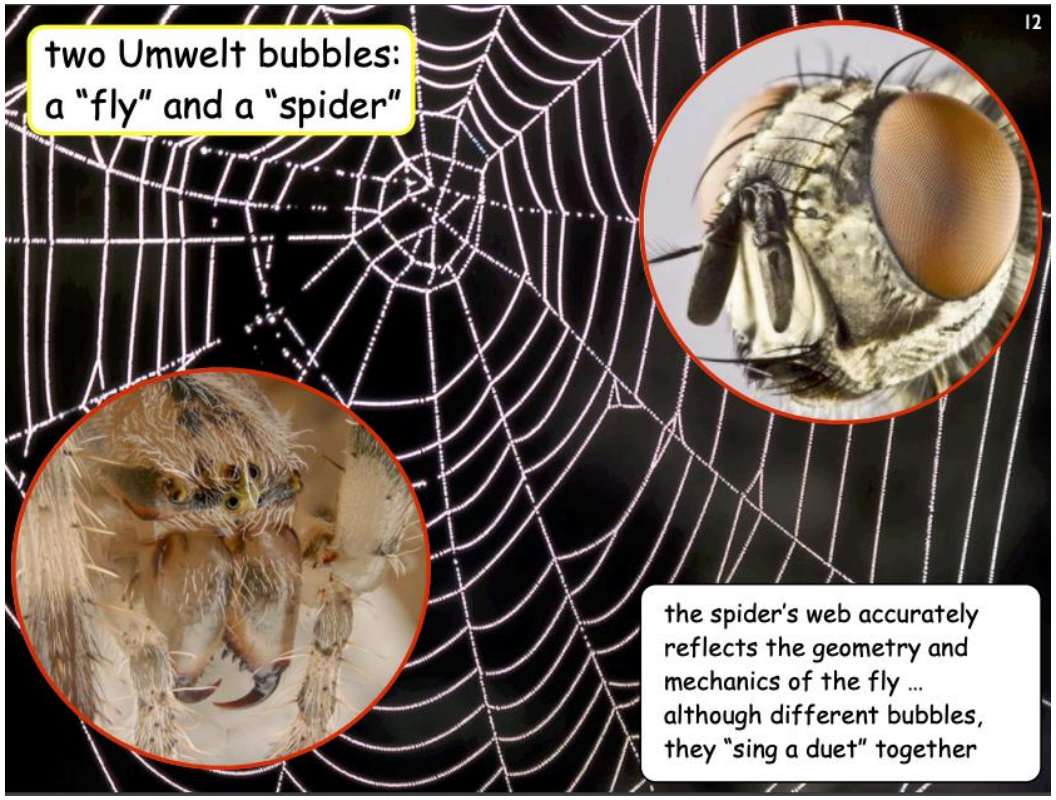
German biologist, 1864-1944

J. Koenderink, Sentience, 2019

Sensory-action worlds

Each organism has its own *umwelt* or "surrounding world"

- This is the organism's sensory and action world. It is determined by biology "bounds the universe from the perspective of the animal"
- The tick's tale: Absolute time and space don't exist from the organism's point of view
- Co-evolution of umwelts



Slide by Lana Lazebnik

“God’s eye”, aka “Shit Happens”

That must have been something! Black swans happen, in this case big time.

Sorry, Reverend Bayes!

Slide by Alexei Efros

The AI viewpoint

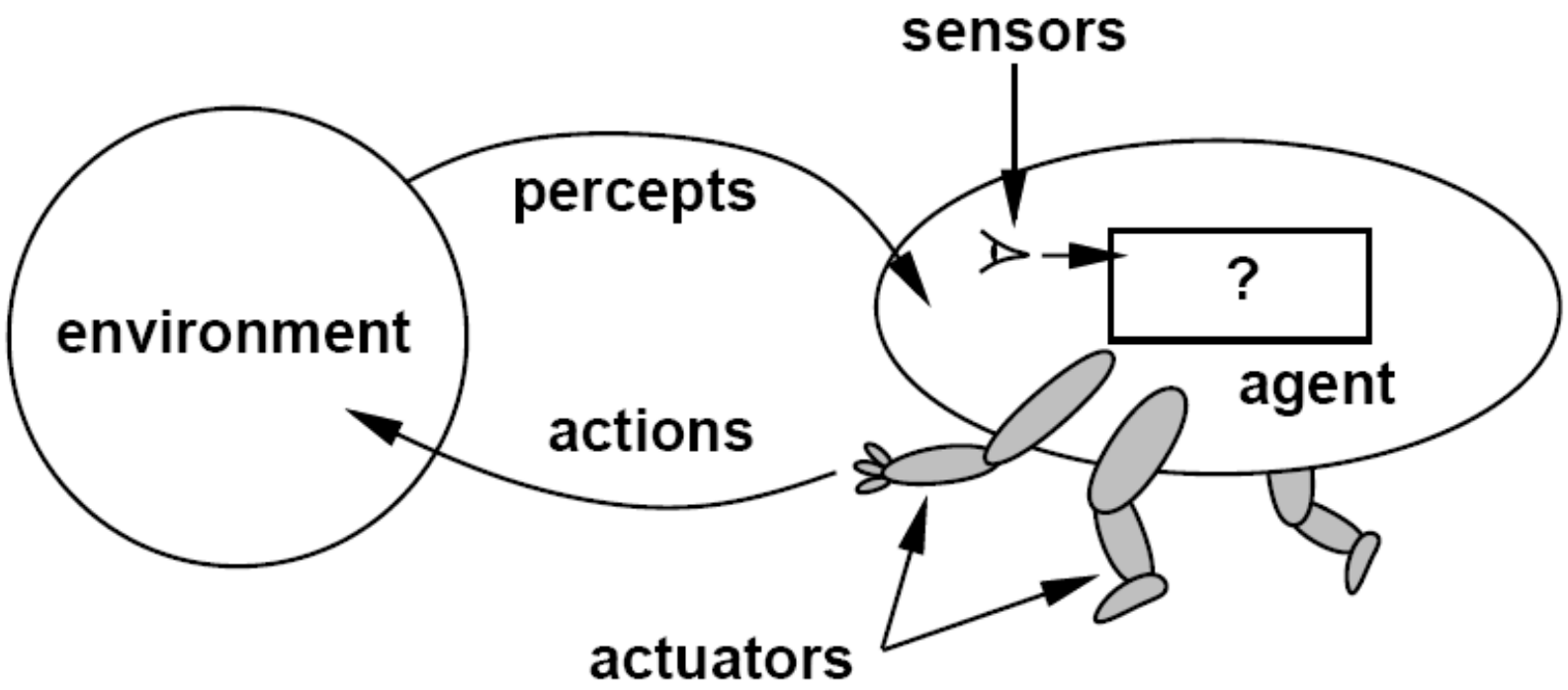


Figure from Russell & Norvig

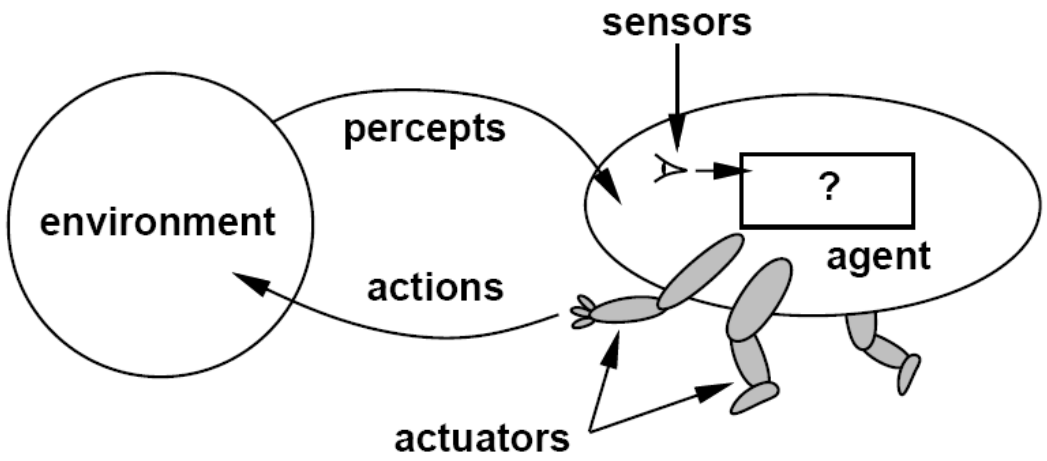


Figure from Russell & Norvig

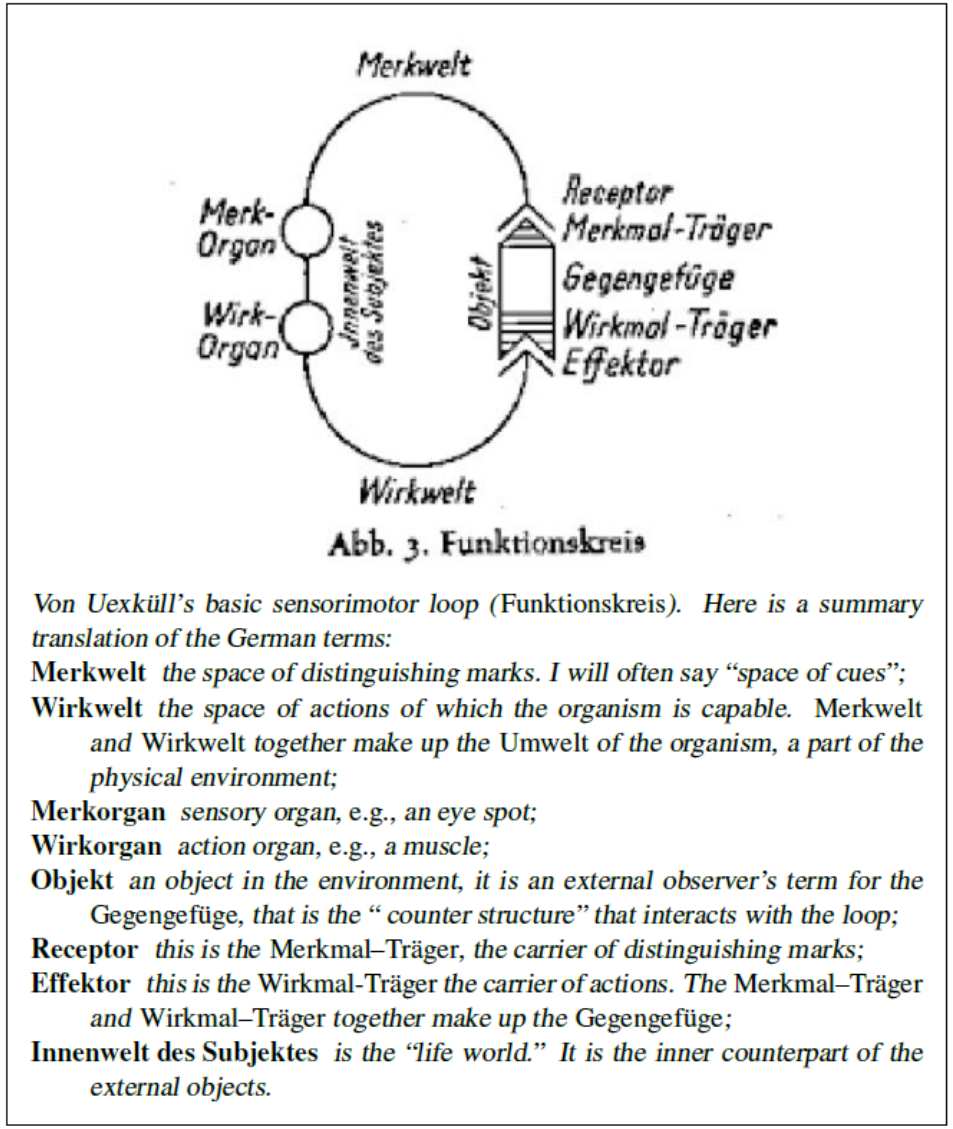


Abb. 3. Funktionskreis

Von Uexküll's basic sensorimotor loop (Funktionskreis). *Here is a summary translation of the German terms:*

Merkwelt *the space of distinguishing marks. I will often say "space of cues";*

Wirkwelt *the space of actions of which the organism is capable.* Merkwelt *and* Wirkwelt *together make up the* Umwelt *of the organism, a part of the physical environment;*

Merkorgan *sensory organ,* e.g., *an eye spot;*

Wirkorgan *action organ,* e.g., *a muscle;*

Objekt *an object in the environment, it is an external observer's term for the* Gegengefüge, *that is the " counter structure" that interacts with the loop;*

Receptor *this is the* Merkmal–Träger, *the carrier of distinguishing marks;*

Effektor *this is the* Wirkmal-Träger *the carrier of actions. The* Merkmal–Träger *and* Wirkmal–Träger *together make up the* Gegengefüge;

Innenwelt des Subjektes *is the "life world." It is the inner counterpart of the external objects.*

Figures from von Uexküll's *Theoretische Biologie*, 1920

Explanation from Sentience, Koenderink

Reflex agent

- Consider how the world IS
- Choose action based only on current percept
- Do not consider the future consequences of actions

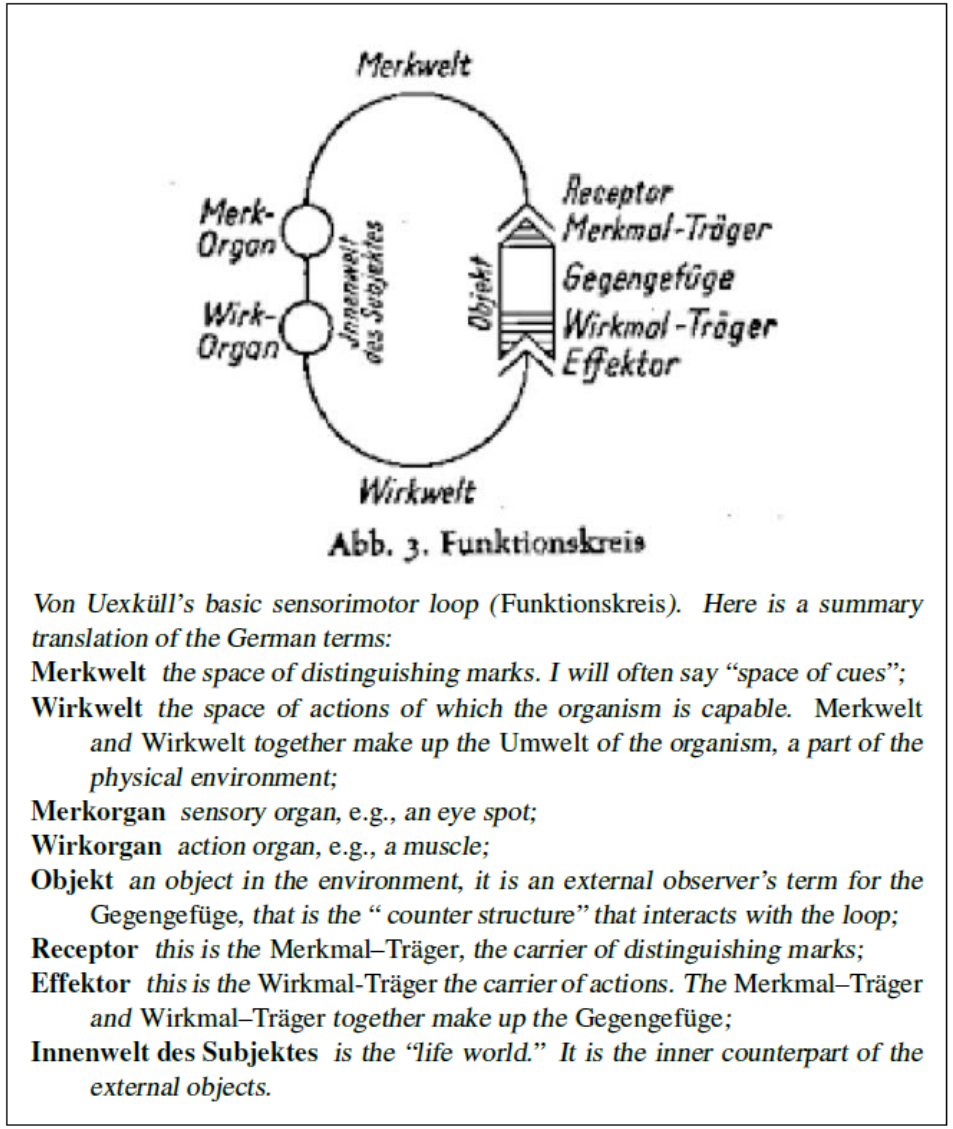


Abb. 3. Funktionskreis

Von Uexküll's basic sensorimotor loop (Funktionskreis). *Here is a summary translation of the German terms:*

Merkwelt *the space of distinguishing marks. I will often say "space of cues";*

Wirkwelt *the space of actions of which the organism is capable.* Merkwelt *and* Wirkwelt *together make up the* Umwelt *of the organism, a part of the physical environment;*

Merkorgan *sensory organ,* e.g., *an eye spot;*

Wirkorgan *action organ,* e.g., *a muscle;*

Objekt *an object in the environment, it is an external observer's term for the* Gegengefüge, *that is the " counter structure" that interacts with the loop;*

Receptor *this is the* Merkmal–Träger, *the carrier of distinguishing marks;*

Effektor *this is the* Wirkmal-Träger *the carrier of actions. The* Merkmal–Träger *and* Wirkmal–Träger *together make up the* Gegengefüge;

Innenwelt des Subjektes *is the "life world." It is the inner counterpart of the external objects.*

Figures from von Uexküll's *Theoretische Biologie*, 1920

Explanation from Sentience, Koenderink

Modified from Lana Lazebnik

The New Loop

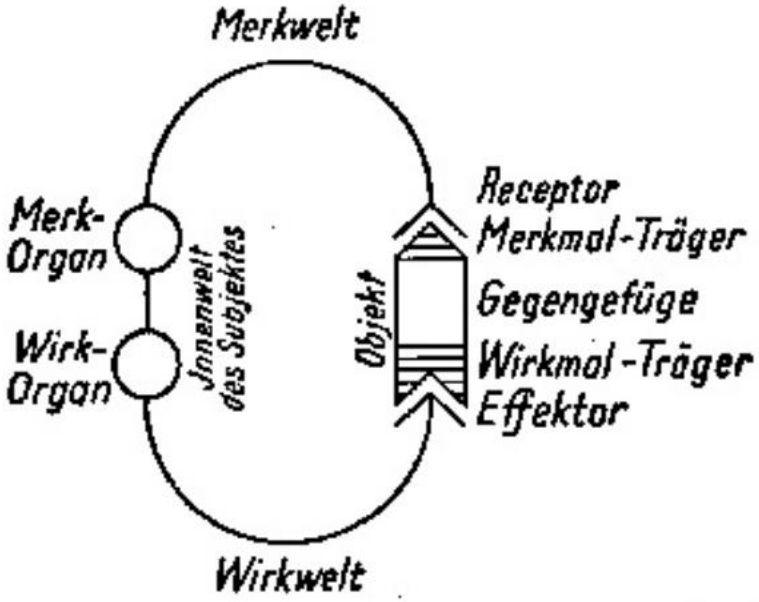
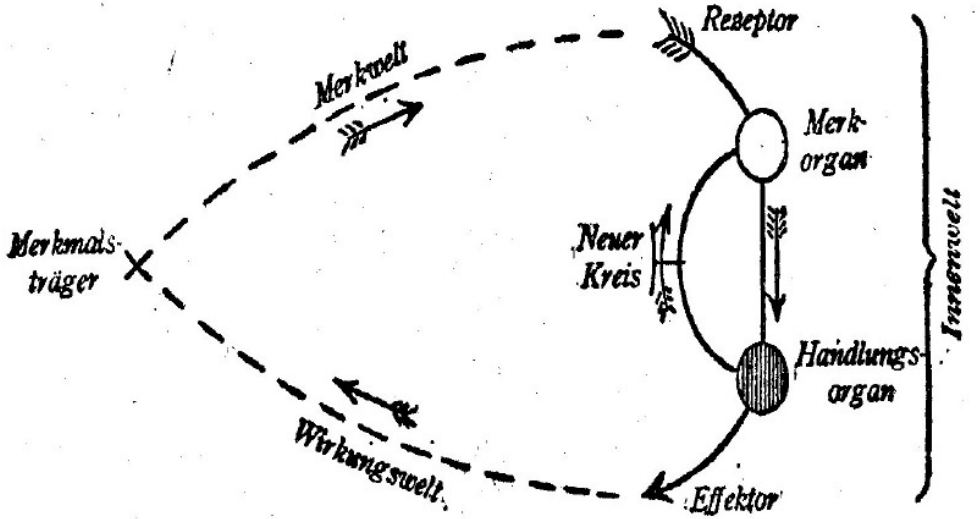


Abb. 3. Funktionskreis

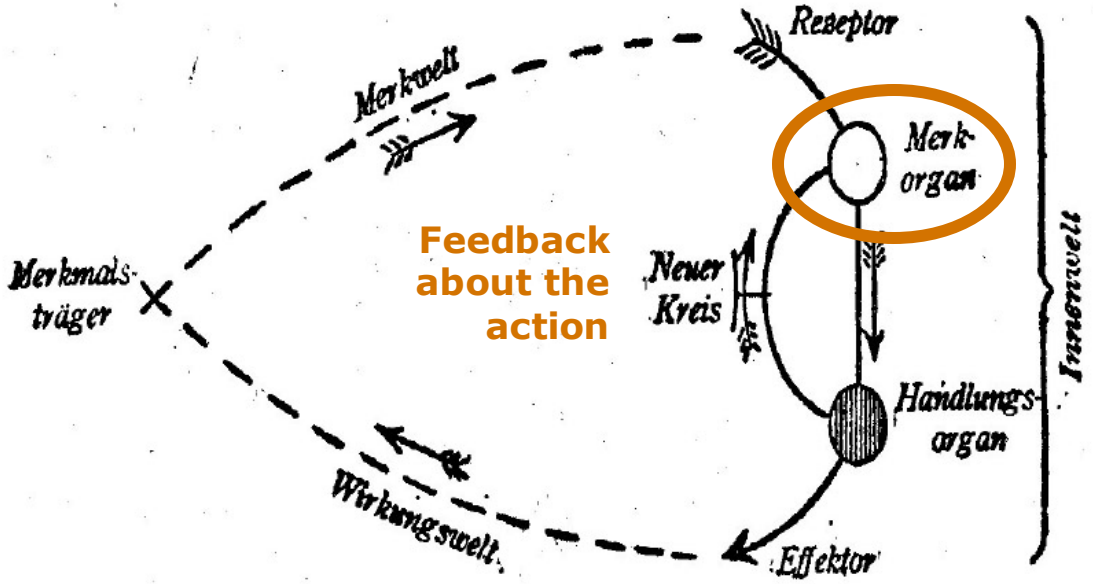


Figur 4.

Figures from von Uexküll's *Theoretische Biologie*, 1920

Sensorimotor feedback loop

Input from the world



The world itself no longer matters!

Feedback about the action

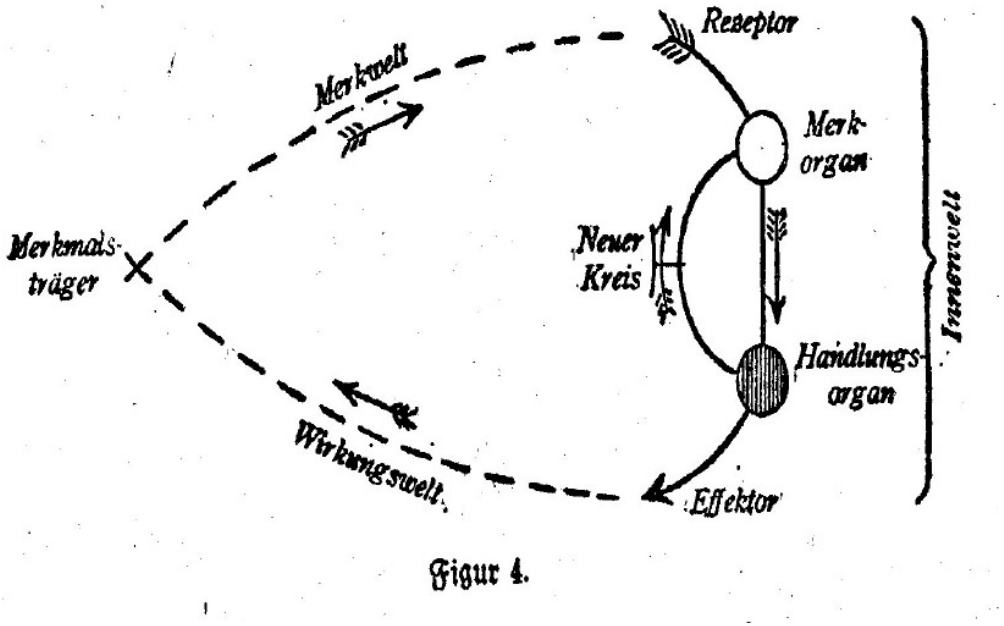
"Inner world" of the agent or interface between the agent and the world

Action to affect the world

(ACTIONS ⇒ OBJECTS) but (OBJECTS ⇏ ACTIONS)

J. Koenderink, Sentience, 2019

Figure from von Uexküll's Theoretische Biologie, 1920



Figur 4.

Predictive agent

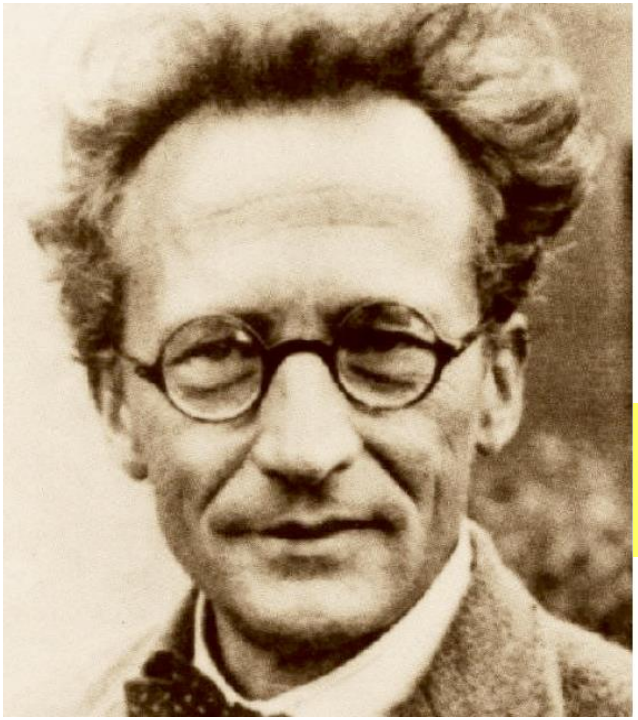
- Consider how the world WOULD BE
- Decisions based on (hypothesized) consequences of actions
- Must have a model of how the world evolves in response to actions

Modified from Lana Lazebnik

The awareness “hypothesis”

The “new loop” is the source of the organism’s **sentience** or **awareness**

- In particular, **discrepancies** between the predictions of the feedback mechanism and the observed state of the world generate **“sparks of awareness”** (a view held by Erwin Schrödinger)



Erwin Schrödinger's in “Mind and Matter” proposes a “psychophysical linking hypothesis” that connects the functional tones to meanings and qualities:

if an expectation is falsified in perception, you “meet nature” - it is a moment of learning: “it discharges a spark of awareness”

Source: Koenderink's slides

Slide by Lana Lazebnik

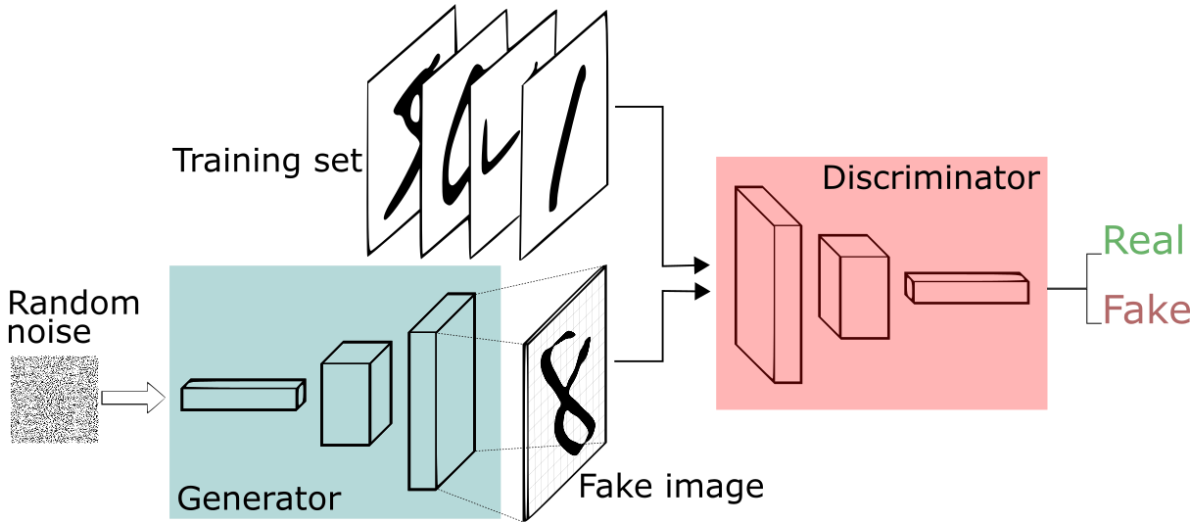
You know only what you (mis)predict

"The inner world only deals with the predictions. It has no other relation to "reality" or "the world" at all. The awareness of a higher animal… is due to the hallucinations of the controller, not to the momentary state of the world. Indeed, how could it be otherwise? All it knows are its hallucinations. It knows them because it made them. **The "world out there" doesn't exist in concrete actuality. It is GOD KNOWS WHERE"**

J. Koenderink, Sentience, 2019

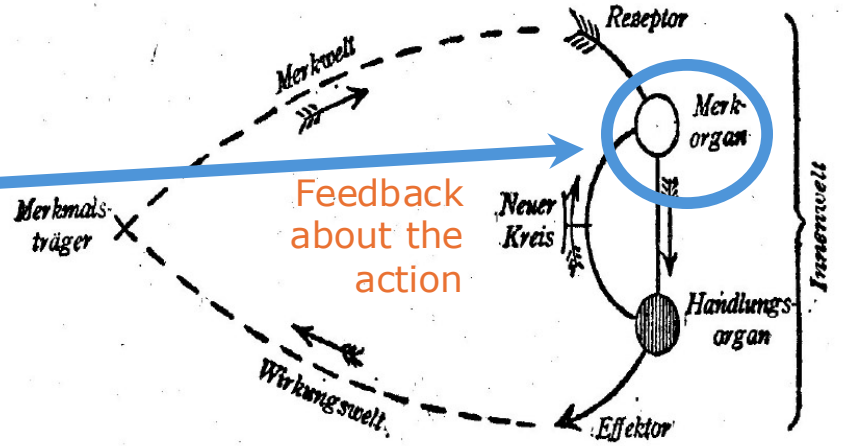
Slide by Lana Lazebnik

You know only what you (mis)predict



GAN generator never sees any real images!
Only the gradients from the discriminator.

Input from the world



Mispredication!

Feedback about the action

"Inner world" of the agent or interface between the agent and the world

Action to affect the world

Slide by Lana Lazebnik

Possible Implications

- "I never lose. I either win or I learn."
  – Nelson Mandela
  - Corollary: you never learn by winning!
- How to maximize learning?
  - Maximize the chances for being wrong → self-supervision?
  - Online learning is better than batch learning
  - Keep making the task harder → curriculum?
- Possible connection between "sparks of awareness" and the perception of time
  - Time is perceived to move faster as we age

Slide by Alexei Efros

On prediction and memory

- The new loop controller may simulate various potential futures as might result from various choices. This greatly enhances biological fitness, because the fate of the animal lies in its future, rather than its past. **The past is only relevant to the extent that it helps foresee, that is simulate, various futures.** That is why your "memory" is not a depository. You confabulate memories on the spot. That makes sense, for all that memory is good for—biologically speaking—is to render your future behavior even more efficacious than your past behavior.

- The past is not just "remembered", it is constructed. This is indeed necessary in order to arrive at a coherent story. (…) The meaning of past events most often only becomes clear in the future.

Slide by Alexei Efros

Interface theory of perception

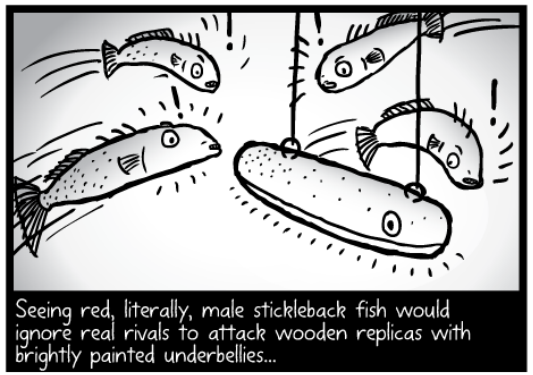
- The "new loop" creates a complete **interface** between the organism and the world. The organism does not experience the world in any other way except through this interface
  - However, the world is still perceived as being "out there" and it can still kill us

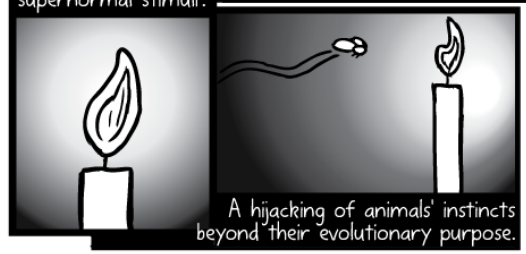
D. Hoffman, The interface theory of perception, *Object Categorization: Computer and Human Vision Perspectives*, 2009
See also https://www.quantamagazine.org/the-evolutionary-argument-against-reality-20160421/

Slide by Lana Lazebnik

Non-veridicality of perception

- Perception evolved not to produce "accurate" representations of the world, but to further organisms' fitness
  - It is easy to "hack" many organisms with *supernormal stimuli*





Source (Wikipedia)

Slide by Lana Lazebnik

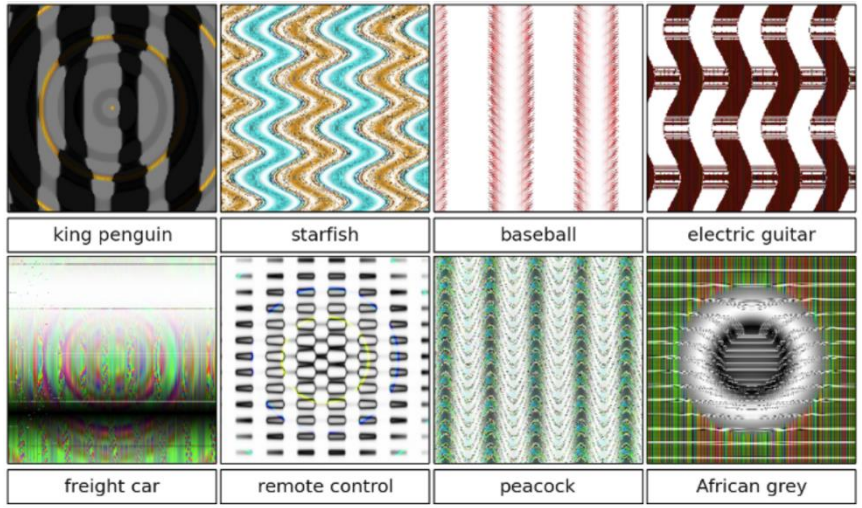
Reed Warbler

Common Cuckoo

The “**new loop**” is the only reality for an organism.

Non-veridicality of perception

- Perception evolved not to produce "accurate" representations of the world, but to further organisms' fitness
  - It is easy to "hack" many organisms with *supernormal stimuli*



Supernormal stimuli for neural networks?

A. Nguyen, J. Yosinski, J. Clune, Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images, CVPR 2015

Slide by Lana Lazebnik

Interface theory of perception (D. Hoffman)

- **Reconstruction Thesis**: Perception reconstructs certain properties and categories of the objective world.

- **Construction Thesis:** Perception constructs the properties and categories of an organism's perceptual world.

D. Hoffman, The interface theory of perception, *Object Categorization: Computer and Human Vision Perspectives*, 2009

The process of perception

- Perception is a fundamentally **active**, **creative** process that generates theories about the world based on sensory input and retains the theory that best fits the input

Looking is an action, *as is pretty clear in this picture of Toshiro Mifune. The notion that vision is a passive act in which the world spoon–feeds you with information is nonsense. Optical meaning is actively hunted for.*

The famous — although fictive — detective Sherlock Holmes plays a major role in my account of the theory of psychogenesis.

"crimes are never solved by forensic scientists. The investigator uses forensic scientists as he sees fit."

J. Koenderink, Sentience, 2019

Source: Koenderink's slides

Perception as Controlled Hallucination

Video by Antonio Torralba (starring Rob Fergus)

But actually…

Video by Antonio Torralba (starring Rob Fergus)

Implications

"Perceptual organization" cannot be primarily a bottom-up process as Marr saw it

Figure 3–1. The interpretation of some images involves more complex factors as well as more straightforward visual skills. This image devised by R. C. James may be one example. Such images are not considered here.

Figure from Marr

What does it all have to do with robotics?

- Perception and embodiment are more linked than we might think.
- It appears that nature ”unified” feedback and generative models. Why? (Computation/Flexibility/Generalization)
- We (might?) need to focus on ecologically meaningful tasks.

Thanks to

Alexei Efros

CS294-192: Visual Scene Understanding
Spring 2022

Instructor: Alexei Efros
Course Coordinator: Allan Jabri

Class time: MW 11am-12:30pm
Location: 1215 BWW

Registration #: 32761 (with code)
Prerequisites: CS280 or equivalent (no exceptions!)
Piazza signup: piazza.com/berkeley/spring2022/cs294192

Lana Lazebnik

Computer Vision: Looking Back to Look Forward

Svetlana Lazebnik

IRIM Visiting Faculty Fellow Mini-Course

January 28 - February 6, 2020

https://slazebni.cs.illinois.edu/spring20/