Mental Robotics

Domenico Parisi

Institute of Cognitive Sciences and Technologies
National Research Council
“What is consciousness?” (“What is artificial consciousness?”)

This is the wrong question to ask

It leads us to think that consciousness is a single entity, with well defined boundaries, without “degrees”, unchanging

Real entities are the opposite

There is always a variety of them, there are no clear boundaries separating an entity from other entities, entities have “degrees”, they appear, change, disappear

“What is consciousness?” is a philosopher’s question

Philosophers ask this question because there is a word “consciousness” and philosophers only deal with language

Science must look at reality beyond language
The empirical phenomena of mental life

Mental images, recollections, thoughts, predictions, plans, dreams, hallucinations, awareness, etc.

Mental phenomena are especially difficult to study because they are private and science is more at ease with public phenomena.

One approach is to recreate mental phenomena in artefacts so that we can attribute these phenomena to artefacts as we attribute them to people.

Mental robotics = the construction of robots that are controlled by neural networks and have mental life.
1. What is mental life?

Mental life is to have internal representations of sensory input in the absence of the input.

**Internal representations** are activation patterns in a neural network’s internal units that mediate between sensory input and motor output.

**Reactive robots** are robots whose internal representations are caused only by sensory inputs originating in the external environment.

Therefore, reactive robots respond only to inputs from the external environment.
A reactive robot’s neural network

- Sensory input
- Internal representation
- Motor commands
Some organisms, especially humans, have mental representations.

**Mental representations** are like internal representations caused by external input but they are self-generated in the absence of the input.

Robots with *mental life* are robots that have mental representations and respond to these mental representations rather than only to external inputs.
The neural network of a robot with mental life
However, it is only when one is asleep that there is (almost) no input from the external environment.

In fact, when one is asleep mental life is called dreaming.

When one is awake, input from the external environment is always present.

Therefore, mental life requires the blocking, or inhibition, of input from external environment.
The neural network of a robot with mental life...
The next question is:

What causes the internal representation, since the input from the external environment is either absent or inhibited?

There must be some additional network’s units that:

(a) cause the internal representation

(b) inhibit the input from the external environment
The neural network of a robot with mental life

- Sensory input
- Internal representation
- Motor commands
The neural network of a robot with mental life
2. Mental life is manifested in a variety of different phenomena and abilities

- mental images
- recollections
- thoughts
- predictions
- plans
- dreams
- hallucinations
- awareness
- etc.
Awareness

Internal representations, both those caused by external input and those that are self-generated, are not by themselves something of which an organism is aware.

A purely reactive robot is not aware of its internal representations.

For a robot to be aware of an internal representation, either externally-generated or self-generated, the internal representation must be continually re-generated inside the robot’s neural network.
Awareness is related to selective attention

Today’s robots tend to be exposed to a single input at a time and they respond to that input.

Real organisms are exposed to many unrelated inputs at the same time but they cannot respond to all these inputs.

With current robots one cannot ask two very important questions: why organisms cannot do many things at the same time, and how the input to be responded to is decided.

**Selective attention:** when two or more unrelated inputs arrive at the same time, the robot’s neural network blocks or inhibits one of the two inputs and responds to the other input.
Motivational states can be selective attention mechanisms

A robot lives in an environment with randomly distributed food elements and water elements.

At any given time the sensory input units encode both the location of the nearest food element and the location of the nearest water element.
motor commands

internal representation

sensory input

food

sensory input

water
The robot’s neural network includes an additional set of internal units that encode with their activation pattern the motivational state of the robot’s body, which can be either “hunger” or “thirst”

These motivational units send their connections to the neural network’s internal units
Using a genetic algorithm one can evolve robots that exhibit the following behavior:

When the motivational state is “hunger”, the motivational state inhibits sensory input from water and causes the robot to respond to sensory input from food: the robot goes towards the nearest food element and eats it, ignoring water.

When the motivational state is “thirst”, the motivational state inhibits sensory input from food and causes the robot to respond to sensory input from water: the robot goes toward the nearest water element and drinks it, ignoring food.
Awareness can be viewed as sequential selective attention

A time 2 the robot’s neural network re-generates the internal representation of time 1 and blocks the internal representation of time 2

Another capacity which is related to awareness is motor attention

The robot produces movements that cause an input to continue to arrive from the external environment while avoiding that are inputs from the external environment replace that input
Mental images

Mental images are the prototypical form of mental life.

Mental images are like internal representations caused by sensory input; however, they are not caused by sensory input but are self-generated.

Recollections are mental images that correspond to internal representations which were caused by some past sensory input.

Inventions or imaginations are mental images that re-combine parts of past internal representations.
Predictions

Prediction is the ability to self-generate the same internal representation that will be caused by some future sensory input.

There are two types of predictions:

Type 1: Predictions of inputs that are independent of one’s movements

Type 2: Predictions of inputs that are dependent on one’s movements
Type 1 predictions:
Predicting an internal representation which is independent of one’s movements (e.g., predicting the next location of a moving object)
Type 1 predictions:
Predicting an internal representation which is independent of one’s movements
(e.g., predicting the next location of a moving object)
Type 2 predictions:
Predicting an internal representation which depends on one’s movements (e.g., predicting the sound that will result from one’s planned phono-articulatory movements)
Type 2 predictions:
Predicting an internal representation which depends on one’s movements (e.g., predicting the sound that will result from one’s planned phono-articulatory movements)
Predicting the sensory inputs resulting from one’s own movements

Time 1

motor commands

planned movement

predicted sensory input 2

sensory input 1
Predicting the sensory inputs resulting from one’s own movements

Time 2

- Motor commands
- Planned movement
- Predicted sensory input 2
- Sensory input 2
Predicting the sensory inputs resulting from one’s own movements

Time 3

- Motor commands
- Planned movement
- Predicted sensory input 2
- Sensory input 2

Comparison
Predicting the sensory inputs resulting from one’s own movements

Time 4

- Motor commands
- Planned movement
- Predicted sensory input 2
- Sensory input 2

These weights are modified
The role of language in mental life

Language plays a very important role in mental life

Internal representations of linguistic sounds self-generate internal representations of objects and actions

Internal representations of objects and actions self-generate internal representations of sounds
Heard linguistic sounds cause the self-generation of internal representations of objects and actions which, in the robot’s past experience, have been systematically paired with those sounds.
Perceived objects and actions cause the self-generation of internal representations of the linguistic sounds that, in the robot’s past experience, have been systematically paired with those objects and actions (inner speech)
other movements

motor commands

internal rep. of object

sensory input

object

phono-articulatory movements

motor commands

internal rep. of sound

sensory input

heard sound

other movements

motor commands

internal rep. of object

sensory input

object

phono-articulatory movements

motor commands

internal rep. of sound

sensory input

heard sound
Language causes internal representations to become conscious

A conscious internal representation, either externally- or self-generated, is an internal representation that generates an internal representation of the linguistic sounds systematically paired with that representation
3. Why do some organisms have mental life?

What good is mental life?

Different answers for different aspects of mental life

Different answers even for the same aspect of mental life

Simulations that show what robots with a given mental life ability can do that robots without that ability cannot do
Awareness

To be able to internally re-create an internal representation

Why should a robot want to re-create its internal representations?

- In order to continue to work on an internal representation even if the original input disappears and is replaced by other inputs

- In order to keep an internal representation in short-term memory

- In order to store an internal representation in long-term memory for recollection
Mental images

Mental images can be useful for all sorts of purposes.

For example, mental images can be used as search images.
Predictions

To be able to predict the next internal representation can help in a number of different ways

- Anticipatory behavior:

  The robot can produce a behavior which is appropriate as a response to some future input

- Replacing a missing input:

  The robot can respond appropriately to the next input even if for some reasons the next input will not arrive

- Evaluating and deciding:

  The robot can evaluate the consequences of its behavior in order to decide, on the basis of this evaluation, whether or not to execute the behavior
1. Predictions can replace a missing input

A robot has to approach an invisible target by using sensory input from two landmarks.

If, for some reason, the input from one of the two landmarks fails to appear (e.g., because of an obstacle), a purely reactive robot is lost.

A robot that can predict the next input can respond to the predicted input rather than to the missing input.
Replacing a missing input

Landmark 1

Landmark 2

Figure 8
Replacing a missing input

Landmark 1

Landmark 2

Figure 9
sensory input 1

planned movement

internal representation 1

motor commands

location of the two landmarks
sensory input 1
planned movement
motor commands
internal representation 2

location of the two landmarks
motor commands

planned movement

internal representation 2

sensory input 2

location of the two landmarks
2. Predictions can allow the robot to evaluate and decide what to do

A robot has to throw stones of different weights to hit prey located at different distances.

Given a particular stone and a prey located at some particular distance, the robot predicts if the planned force of its throwing movement will allow the stone to hit the prey or not.

This prediction/evaluation causes the robot to physically execute the planned movement or refrain from doing so.
force of throwing

motor commands

internal representation

sensory input

sensory input

distance of prey

stone weight
sensory input

motor commands

planned movement

internal representation 1

predicted result

force of throwing

sensory input

distance of prey

sensory input

stone weight
3. The ability to make predictions can be a critical component of the ability to learn by imitating other individuals.

The robot’s neural network changes its connection weights to make the predicted effects of the robot’s movements match the observed effects of another individual’s movements.
Imitating another individual

Time 1

- Motor commands
- Planned movement
- Sensory input 2

Comparison

From another individual
Imitating another individual

Time 2

motor commands

planned movement

predicted sensory input

these weights are modified

sensory input

from another individual
4. Predictions can give a robot a sense of self and of self-agency

- Sense of self (of one’s body)

Predicting systematically different inputs from one’s body and from other objects

- Sense of self-agency

Inputs that the robot can predict will systematically follow its actions, for the robot are caused by itself

Experiment by Sato & Yasuda (Cognition, 2005, 94, 241-255)
Language

The self-generated internal representations caused by language are what higher order cognition is mostly made of.

Having conscious internal representations, i.e., linguistically labeled internal representations, can provide various advantages: self-generating other internal representations (thinking), planning actions based on predicted rather than actual sensory input (planning), sharing and discussing one’s internal representations with other people, etc.
4. Mental life in the real brain

A. Where are in the real brain the internal units that cause self-generated internal representations and are responsible for mental life?

B. Are the internal units that encode the internal representations caused by actual sensory input the same units that encode self-generated internal representations?

The answer to the first question is that the internal units (circuits) that cause self-generated internal representations can be located in different parts of the brain and can be different for different types of mental life phenomena.

The answer to the second question is that the units encoding internal representations caused by sensory input and the units encoding self-generated internal representations may not coincide.
Internal units encoding self-generated visual internal representations may include secondary visual area of the occipital cortex and areas in the parietal and frontal cortex.

Internal units encoding internal representation caused by sensory input may include, in addition to the above areas, the retina and primary visual cortex.

During REM sleep, when dreams are more common, the secondary but not the primary visual cortex is active.

In congenitally blind people, during dreaming and during Braille reading, secondary, but not primary, visual cortex is active. (In blind people, secondary visual cortex appears to be recruited for internal representations of movements.)
Conclusion: from humanoid to human robots

Current humanoid robots are human in only a very superficial sense.

If we want to construct human robots, we must endow robots with mental life.

This is possible if we apply to robots the same criteria that we use to attribute mental life to people.

The goal of mental robotics is to reproduce in robots the rich phenomenology of people’s mental life.