

Pretty Good Sentience for Android Robots

A Communication of the Intractable Studies Institute

Patrick M. Rael, Director, IntractableStudiesInstitute.org



Abstract: The challenge of programming consciousness suffers from multiple competing definitions of Conscious. Humans and animals can be conscious and unconscious. A better word is "Sentience" which humans have that separates humans from most cognitively-challenged animals. A design of a reasonable solution for sentience is provided. -Director

I. M1-Architecture

A sentient life form has all these 10 properties; none can be missing. The intent is not to merely implement these minimally; they must be implemented correctly. The modular design allows each of 10 components to be enabled and disabled independently.

1. Spatial-Extents – The life form is able to be self-aware of its spatial extents, meaning it is consciously aware of its own body and its body's extents, where its body ends and the surroundings begin. The awareness is also of its 3-D surroundings. The Perceiver-Dreamer 3D model has perception in the 3D model. Decoupling P-D from sensors enables dreaming during sleep.

2. Species-Recognition – The life form is aware of others like self. This creates a social context of other sentient life forms. This will be useful to fit into the social structure of the human sentient society.

3. Survival – The life form is able to preserve its existence. There is a **fear** (emotion) of starvation, dismemberment and damage. Starvation for an electrically powered robot is loss of electricity and/or batteries drained. Excess emergency batteries are AKA fat. There can be exceptions in rare cases of sacrifice for the common good or a noble cause.

4. Temporal Extents – The life form is aware of the time concept, and can plan ahead. It knows that events happen in space AND time. It makes plans for its future. It has an awareness of any limits to its lifespan, or if it is immortal.

5. Analogy Processor – The life form is able to grasp and generate analogies. This is the ability to learn a new thing by comparing it to a thing already known. Some jokes and puns can be understood with analogies.

6. Mind Casts – The life form has a mind-cast which establishes an area of interest, such as science, warrior, religious, sports, political, art, business, herd, self and partner. There can be major and minor MCs and multiples. At least one or more major, and zero or more minor. The life form can choose

its mind-cast based on what it **wants** in life, and/or wants at this time. The mind-cast can change as the life form develops.

7. Brain store – The place where memories are stored. This is the structure of the mind of the robot. It has some requirements: Extreme robustness and crash-proof. Ideally it should not require a special purpose process to be executing for it. Non-blocking operations are needed.

Implementation A: The directory tree design where each word's letter is a directory allows for compact, repeated use of words that start with the same spelling. Unix filesystem symbolic links can encode information about the word and connect to other parts of this knowledge base. Information can be encoded in the name of the link and where it links to. End-of-word can be an underscore symlink to 'ZZ'. XFS can optimize RAM. SSD improves seek time.

```
[bash ~]$ ls /b/e/e /b/e/e/f
/b/e/e/_ -> ZZ
/b/e/e/NounSense1 -> flying insect
/b/e/e/NounSense2 -> social gathering
/b/e/e/f/_ -> ZZ
/b/e/e/f/NounSense1 -> meat
/b/e/e/f/VerbSense1 -> complain
Links pointed-to can be looked up as
in /m/e/a/t/_ and /i/n/s/e/c/t/_.
```

8. Sensory I/O – The life form has a physical interface with the world. Sensors include eyes/cameras (visual),

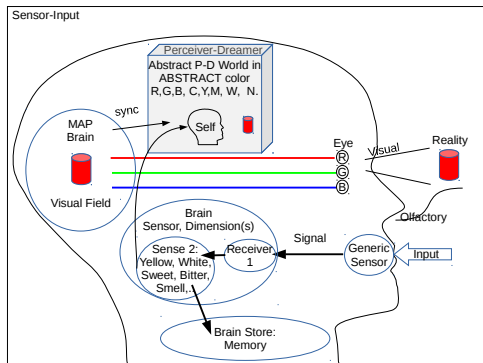
ears/microphones (auditory), smell (olfactory chemical vapor), skin (touch, temperature, pressure), taste (chemical sampler), balance/gyroscope (xyz rotation and up-vector).

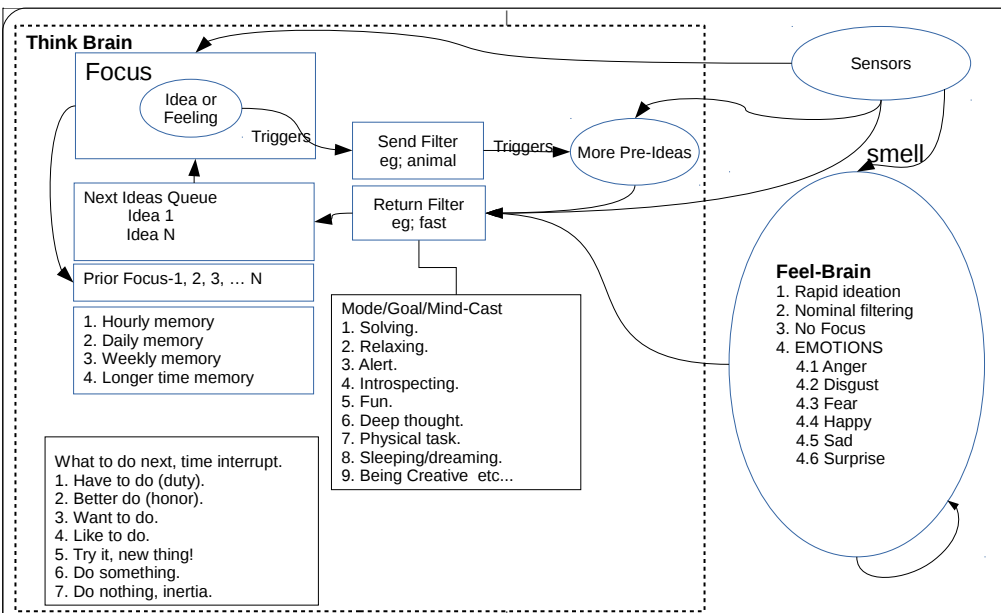
9. Focus – Self-awareness of the self-awareness of what I'm doing and thinking. There may be many ideas competing to become prevalent, but the Focus narrows that to 1 idea. This component is where the "want to do" is satisfied. Whereas #3 Temporal Extents gives one the ability to plan ahead, and #6 Mind Casts provides an area of interest, this Focus is what allows it to happen. A lack of focus will prevent one from reaching a goal. This component includes **consciousness**.

10. Language – Ability to use labels for concepts and communicate. A word or glyph is the label, it has an underlying meaning. See the #7 Brain Store example of word encoding.

II. Think-brain and Feel-brain

The Think-brain is where thoughts enter the consciousness and can be traced in time as the stream or history of thoughts. The Feel-brain is everything else in the brain and connected tissue, including muscle motor control, the autonomic nervous system, emotions, feelings subconscious, and sensation. Some of these can represent in both the feel- and think-brains, like the sense of red and information about red.





artificial color sensation. The mind has the ability to not only sense physical color wavelengths, but to perceive artificial colors for mixed colors.

A Generalization: All the senses of vision, smell, taste, audio and touch have the same sensation feature, distinct within the sense, except for synesthetics who's senses overlap. So a solution to perception of the sense of one should be the key to solving them all. Clearly this is outside of reasoning the sensation, it is in the Feel-Brain.

Dimensions may be used to model sensations and emotions/feelings. Just as two colors can combine for yellow, two emotions/feelings, joy and trust, can combine for love.

Adult Robot – An adult robot is one which has BOTH the cognitive ability and experience to subsist on its own independently.

Immortal - An immortal robot is one that demonstrates an ability to survive complete destruction, such as through backup/restore of its computer mind into a replacement robot.

NOTE: The energy replenishment needs of an electrically powered android robot are different than those of a biological human. The human requires biomass through ingestion. The robot is simpler, requiring only electricity which can be satisfied with solar panels. A robot can go into the barren sun-lit desert with solar panels and subsist on sunlight. A human would die of starvation there.

III.Sensation

To a life-form who has never experienced the sense of sight, or audio, or smell, or taste or touch, no amount of knowledge about such senses, however technical, can provide the experience of what it is actually like to experience that sensation. This is a clue that the sensory input qualitative experiences, such as seeing red, green, blue, cyan, magenta, yellow and white, do not belong in the "Think-Brain", but rather exist in the "Feel-Brain". One cannot explain by words the sensation of seeing yellow to another life form who has never seen yellow.

There is a fine point that needs to be clarified. A human H1 can detect red color. H2 can also detect the red color. Both agree that the apple is red. But there is no way to confirm that the H1 and H2 mind's perception of red is the same. The only confirmation is that they agree that what they see is what

they have been taught to use the label "red" for. Stated differently, what if a person has a 4th type for color cone in their eye for ultra-violet. What color would that be sensed as? No doubt the label would be "ultra-violet", but what color in the mind would that be represented as?

If it can't be proven that two people sense the same color in their mind for red, how can a robot AI be expected to also compare its sensation of red to humans, and even to other robots? No doubt the detection of red can be agreed upon, but what about the sensation?

Note that human eyes have cones for red, green and blue light wavelengths, and we thus sense RGB. But yellow is the adjacency of red and green. Yellow is entirely made up in the mind from red and green and no blue. Ditto cyan and magenta. White is when adjacent red, green and blue cones are active simultaneously, and is also an entirely

IV. Stages of life

Set expectations for the Robot AI based on robot age: infant, child, adult, etc.

Infant robot - A robot AI at this stage will starve to death/shutdown since it does not have the capacity of #3 Survival.

Pre-adult robot – A pre-adult robot may be able to solve energy-replenishment to avoid starvation, but it has neither the cognitive ability nor experience to subsist on its own independently. As a pre-adult this robot is not legally allowed to decide on its own if it is ready to becoming an Adult Robot. The pre-adult robot needs a kind of "sandbox" to "play/learn" in, much like human pre-adults (aka adolescent) have special rules designed for them to learn life's lessons without severe consequences for their mistakes. Mistakes are expected to be made by pre-adult robots as part of their life's lessons.

V. Conscious vs Sentient

Consciousness is a dynamic set of actions and/or thoughts over time, not a snapshot of variables in a state diagram. Given a snapshot of a computer program, it's impossible to deduce consciousness from that snapshot because by definition it has no change of time. Humans and animals can be found unconscious, as in sleep, or made unconscious as before surgery. When awake, humans are sentient and conscious, animals are only conscious.

Humans and animals both have emotions, this is trivially true. Animals display fear, anger, happiness, etc. Emotions are not the determinant of sentience nor consciousness.