



Rich People In 10 Years

(Everyone in 20)

Zachary Donsky



Human adaptation is dependant on the collection, understanding, and consequential action of information.

Communication in human evolution has clustered the range of human intelligence closer together.

Human communication was likely first visual body language and verbal sounds. Then likely developed: lasting visual communication in cave paintings. Then there was specific, lasting communication: written language.

“



Nick Bostrom measures the progress of human adaptation by comparing Kanzi, a bonobo that had mastered more than two hundred lexical tokens (the smallest meaningful unit of a statement), and Ed Witten the distinguished superstring theory researcher.

Bostrom explains that between Kanzi and Witten are two hundred fifty thousand generations of great apes. Bostrom cautions that two hundred fifty thousand generations is not a large number on the scale of biological evolution. It amounts to a four percent difference in DNA.





Such a small difference between homosapiens and its closest cousin have manifested in significant ways

Bostrom quips that although an adult chimpanzee is twice as strong pound for pound as a fit adult human, the fate of chimpanzees largely depends on the actions of humans. (Bostrom, TED Talk) Bostrom's example demonstrates:

Small developments of the brain have had significant consequences.

It follows that a significant development in thinking would have huge consequences



Humans first used
**Visual
Language**

Then we used
**Spoken
Language**

And the we used
**Written
Language**



NEXT IS AUTONOMOUS LANGUAGE

ARTIFICIAL INTELLIGENCE



Human Level Artificial Intelligence Could Be Built:

Likely in 30 to 60 years (50% : 90%)

A machine that is as competent at doing tasks as a human will also do the tasks much faster. Neurons fire at ~200 times per second. Transistors can fire faster than ~200'000'000'000 per second.



A Neil Degrasse Tyson intelligence level AI, in one hour, completes 1'370'000'000 years worth of Neil Degrasse Tyson level research (operating on hardware from 2016).



Autonomous human level research at digital speeds has huge consequences:

A machine more intelligent than the cleverest human will be better at designing machines. Therefore it can create a machine more intelligent than itself. (Chalmers, The Singularity)

This is the rationale for The “Intelligence Explosion”, the prediction that human level artificial intelligence will be shortly followed by an artificial intelligence superior by many orders of magnitude, followed by a trend of exponential growth.



Bostrom posits that human level artificial intelligence will be the last invention the human race will need to make. After that, machines will be better at inventing than humans, and “they will be doing so on digital time scales.”



In a scenario where an AI can get what it wants, the future will be shaped by its preferences.

(Bostrom, TED Talk)



Eliezer Yudkowsky's "AI-box Experiment" is a party game designed to demonstrate an artificial intelligence's ability to get what it wants.

The basic structure of the game is this:

Two players:

- AI in air-gap box
- Gate-keeper

AI win condition: Convince Gate-keeper to let the AI out of the box.

Gate-keeper win condition: Do not let the AI out of the box. After a two hour time limit, or after any involved parties get bored, the gate-keeper wins.



Yudkowsky claims that to escape containment an AI only needs access to a human being, even if it's just through text chat

In this game the advantage clearly lies with the Gate-keeper. The gate-keeper wins by default, even if they are completely passive in the game. The onus is on the AI player to engage the gate-keeper and progress the game. And yet Yudkowsky claims he has won playing as the AI on more than one occasion *and* with sizable cash wagers!

On these grounds Yudkowsky claims that to escape containment an AI only needs access to a human being, even if it's just through text chat.

Bostrom claims the preferences of an artificial intelligence that can get what it wants is likely to shape the future. Yudkowsky demonstrates the mutual

exclusivity of containing an artificial intelligence and having a human interface with it. Finally, Yudkowsky, quoted inside Bostrom's article "The Superintelligent Will," describes perhaps the most troubling aspect of artificial intelligence:



I actually found this by googling: "old sci fi comic cover art"

Yudkowsky explains about an old sci-fi comic stereotype. Cover art of a bug eyed monster holding a beautiful woman in a ripped dress.

"The artist did not ask whether a giant bug perceives human females attractive. Rather, a human female in a torn dress is sexy--inherently so, as an intrinsic property." -Yudkowsky

The cover art represents a logical flaw on the part of the artist, and the target audience. But the bug eyed monster, being a biological organism, would likely share many values with humans. We can reasonably assume it evolved from a process of reproduction with variation. From evolution we can assume it may protect its young, have a need for food, and maybe loyalty (if it is a social organism). None of these traits would have to be shared by an artificial intelligence. (Bostrom, The Superintelligent Will) The moral is that an intelligence that does not share the many requirements for life that humans have is unlikely to often form goals that are consistent with human survival.



So

Super artificial intelligence is likely to develop in the next thirty to sixty years. Shortly after development it will likely become uncontrollable. It will not share humanities requirements for survival, and consequently could form goals incompatible with human survival. And being the highest form of intelligence on the planet, it will likely control the fate of our species.

;D

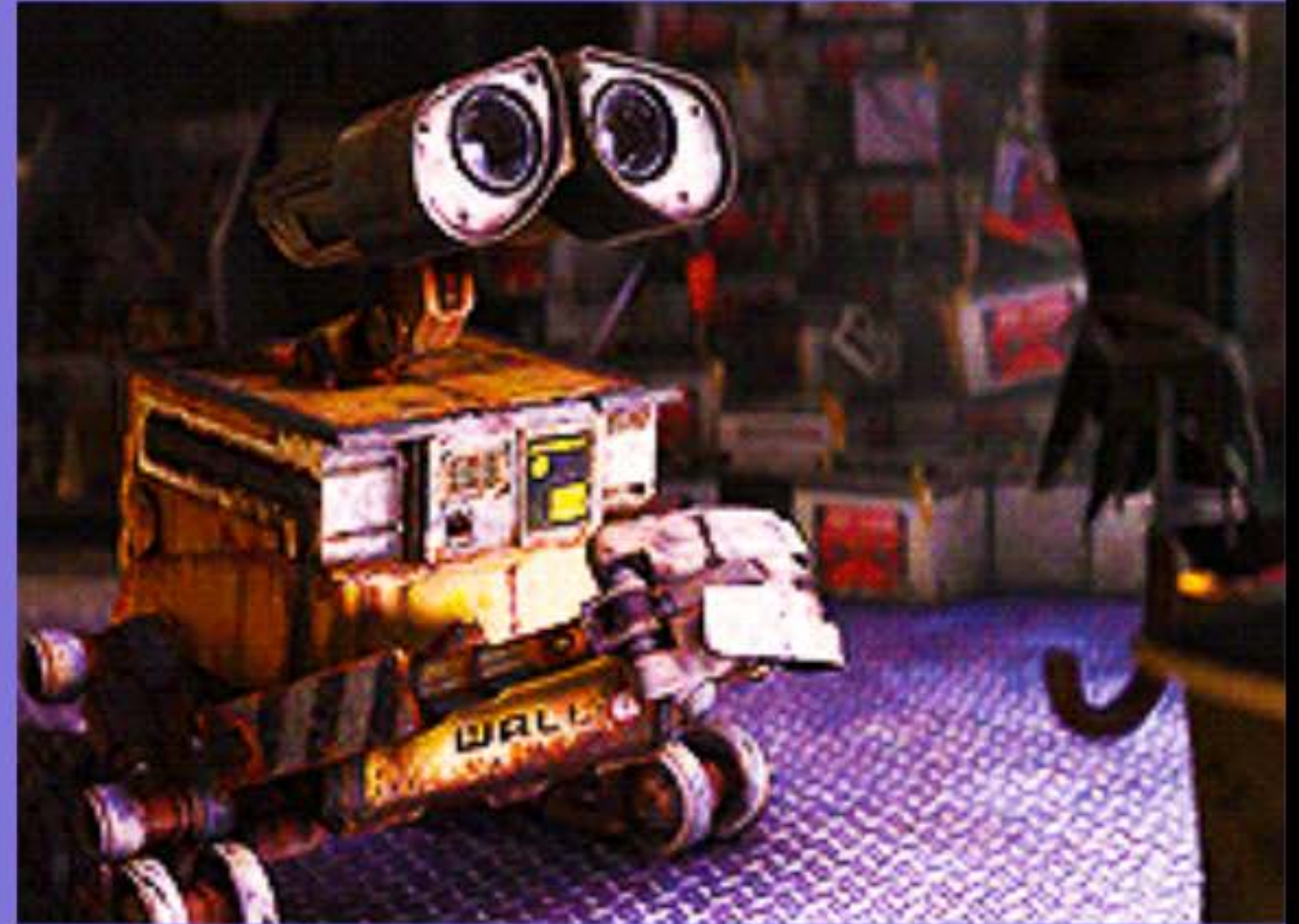


These conclusions pose a clear challenge: Develop machine learning strategies to observe human behaviour to ascertain our core values. Compare its future decisions against expert models of human values.

OBSERVE HUMAN BEHAVIOUR



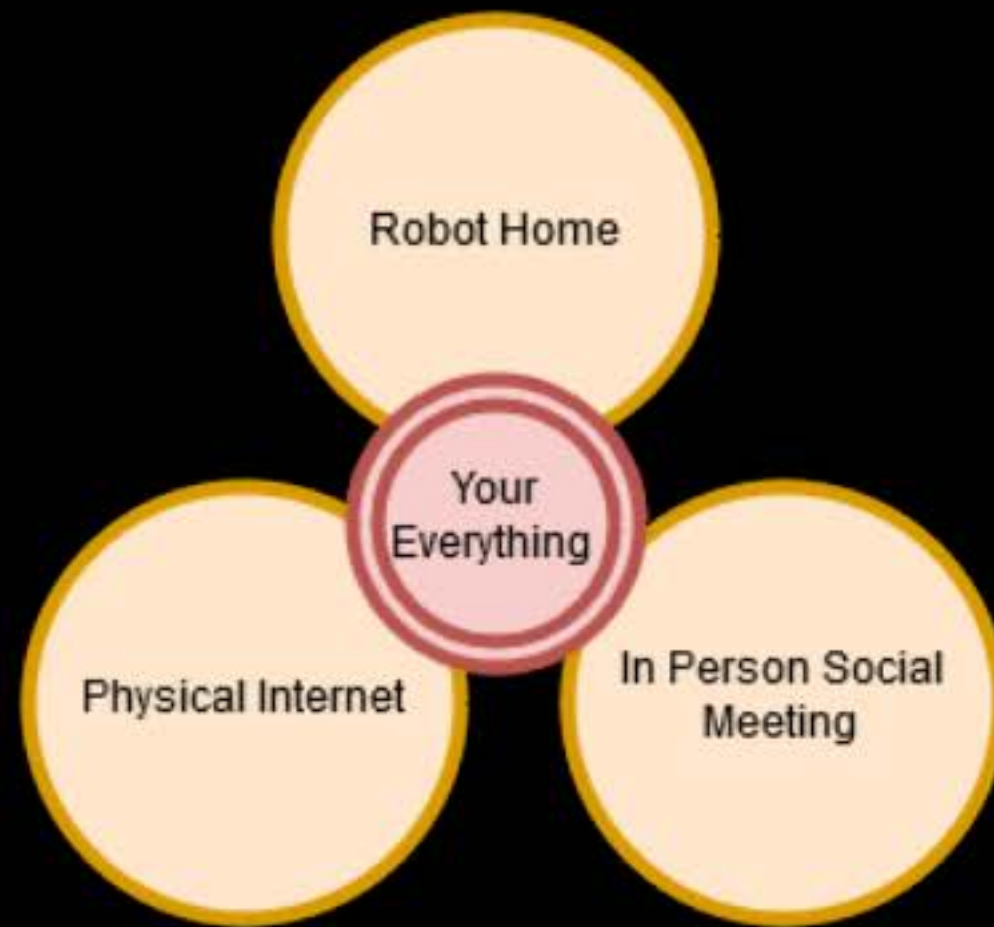
DETERMINE HUMAN VALUES





Complete observation of human behaviour can occur if the observation tool is also the behaviour space for all human behaviour.

A behaviour space for all human behaviour.



Here is my stab at the general categories for the behaviour space of people in wealthy countries in the near future.

Autonomous Tasks



Mobile, Robust Homes

Internet of Things



Robot Home

Your Everything

Mixed Reality



Virtual Reality



Augmented Reality



Physical Internet

Your
Everything

Your
Everything

In Person Social
Meeting



Bedroom



Work



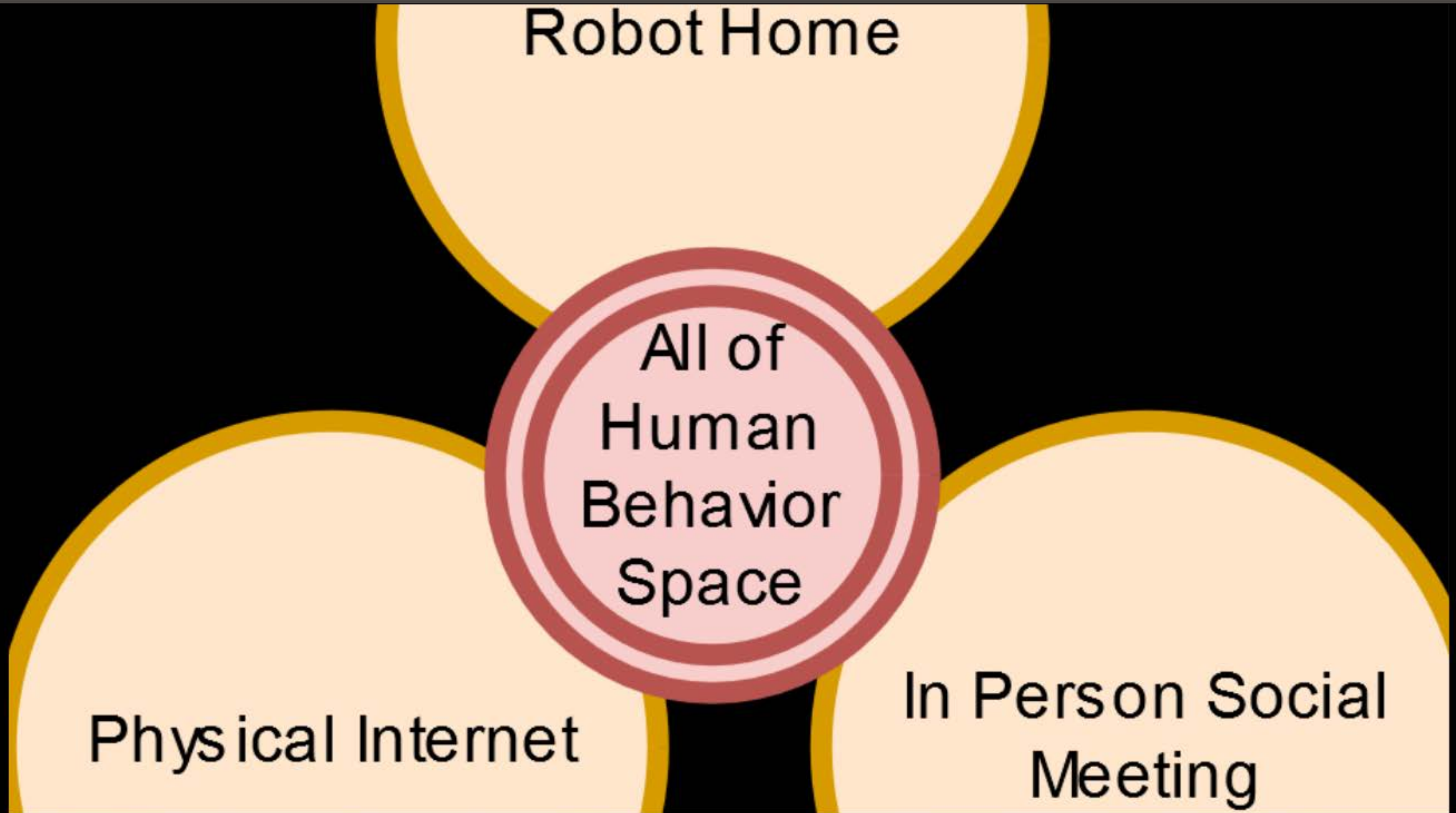
Meetup
Spot

Robot Home

All of
Human
Behavior
Space

Physical Internet

In Person Social
Meeting



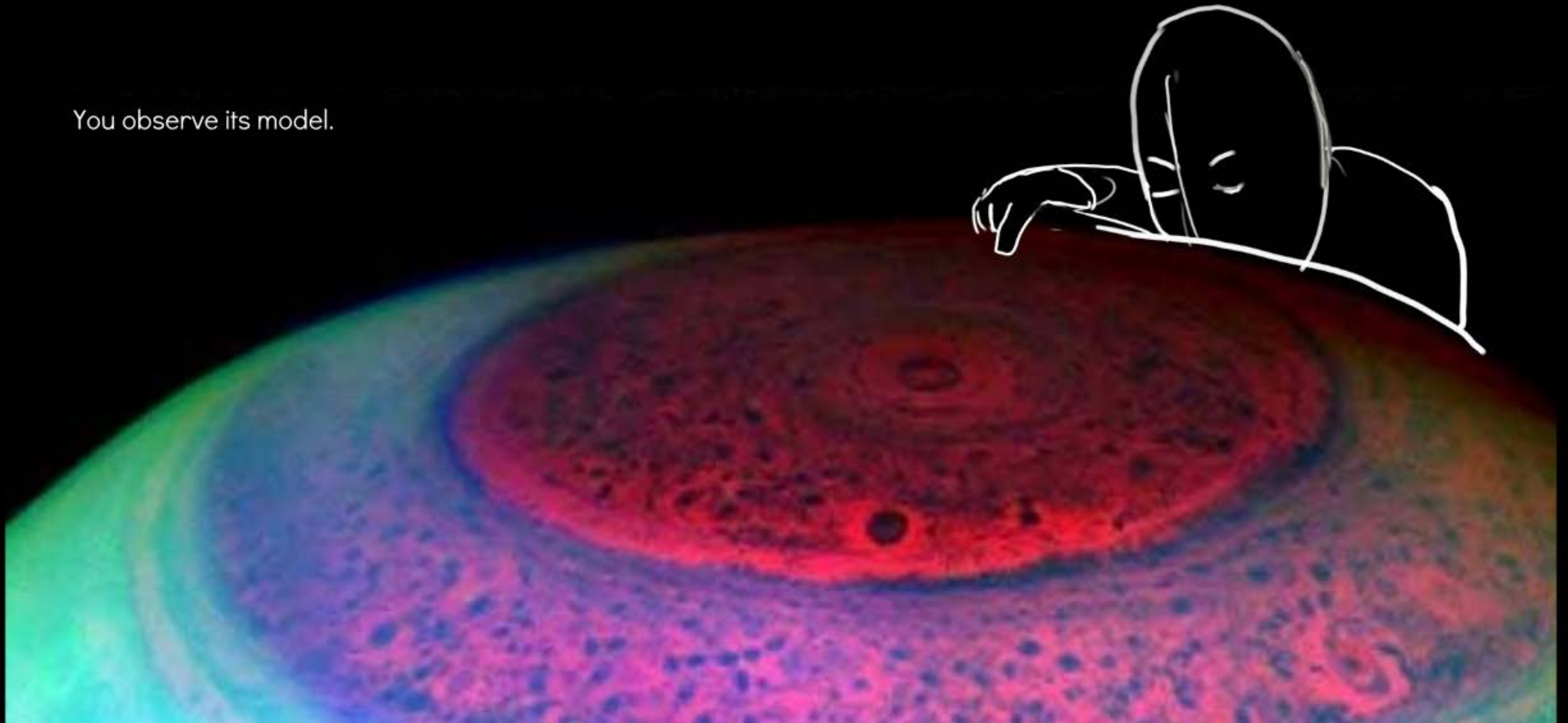
The best way to maintain a complex system is to make it small.

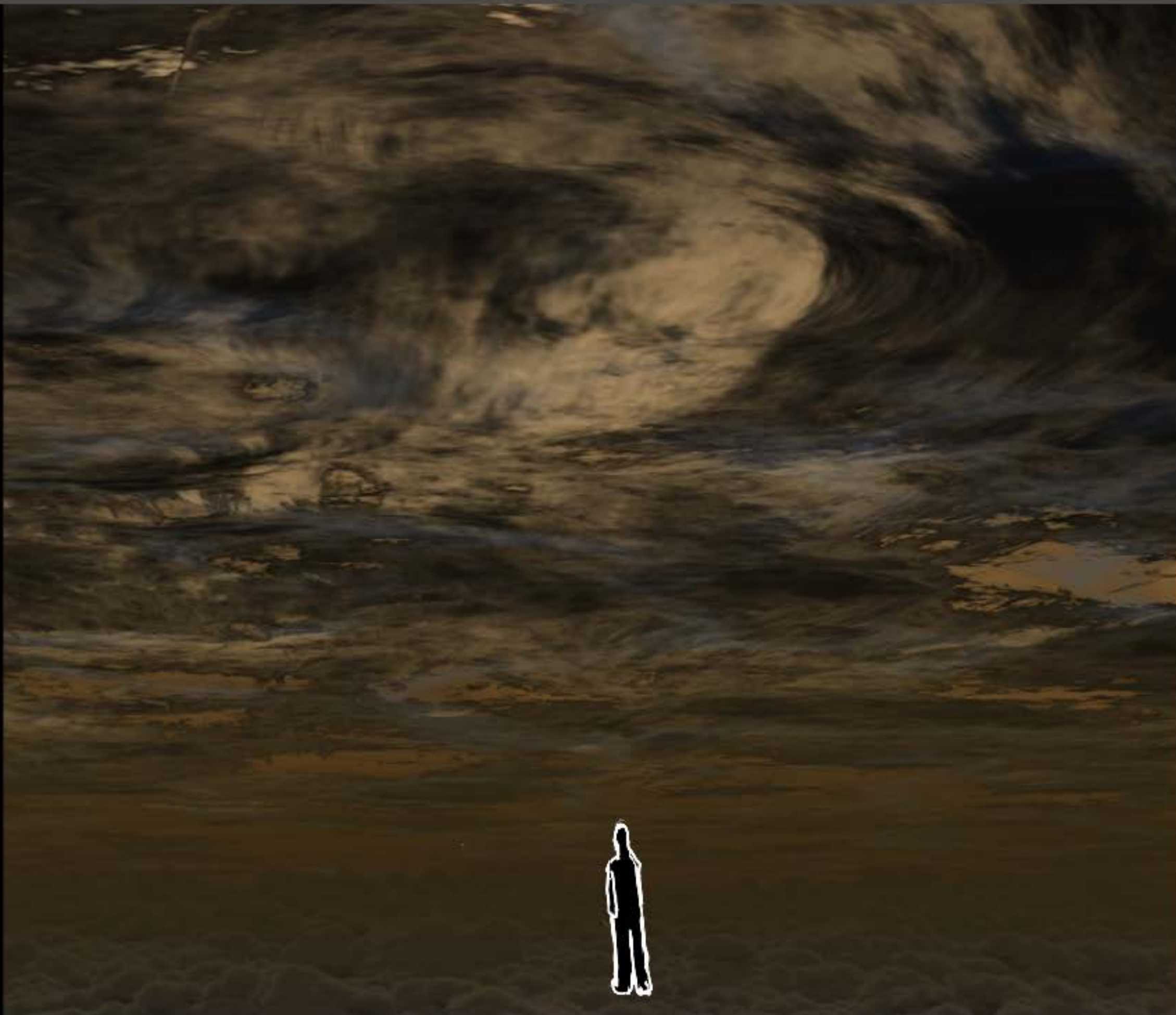
-Paolo Soleri's 9th principle of design

You're chilling in your brain,
contemplating this massive
storm that's happening on
Saturn.



You observe its model.





You observe it from the
ground.



It's magnificent and
deafening



A message.

THE STORM
HAS CLEARED



OK

SHOW ME

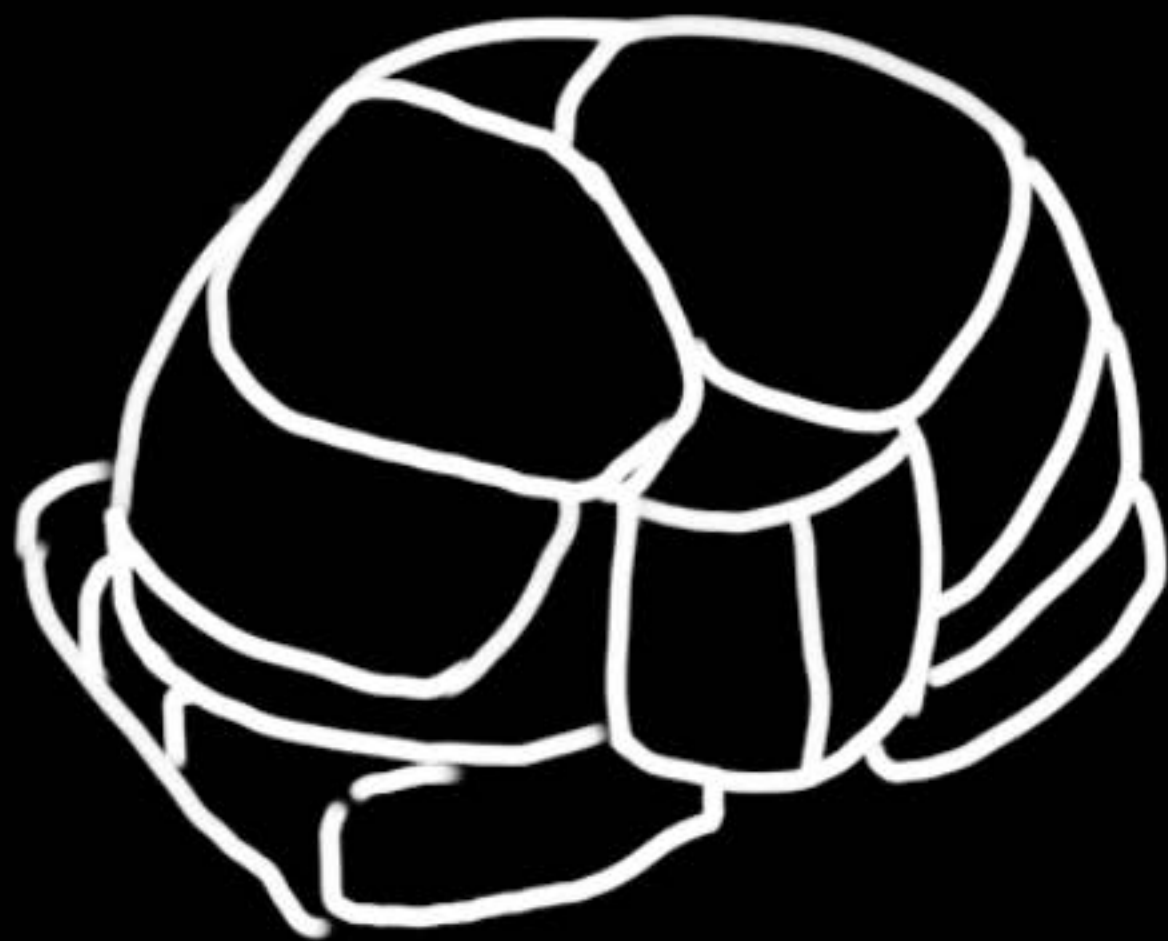
OPEN
HATCH





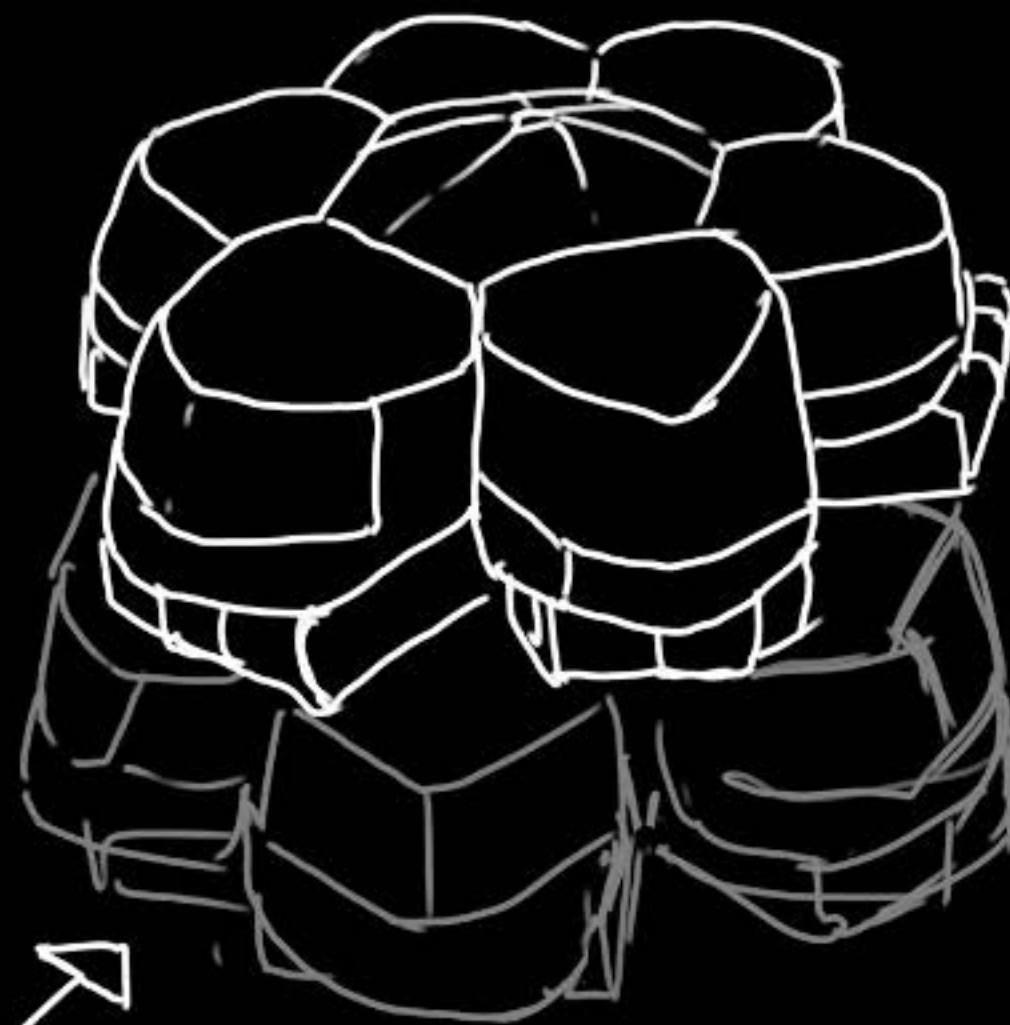


Capsules link to form space
greater than the sum of their
parts.



You visit someone else's brain.





Community sized link.

(This is a Super Smash Bros.
LAN tournament)

(Meeting spots in the
abandoned gas stations)

Locomotion



Life Support



Behaviour Observation



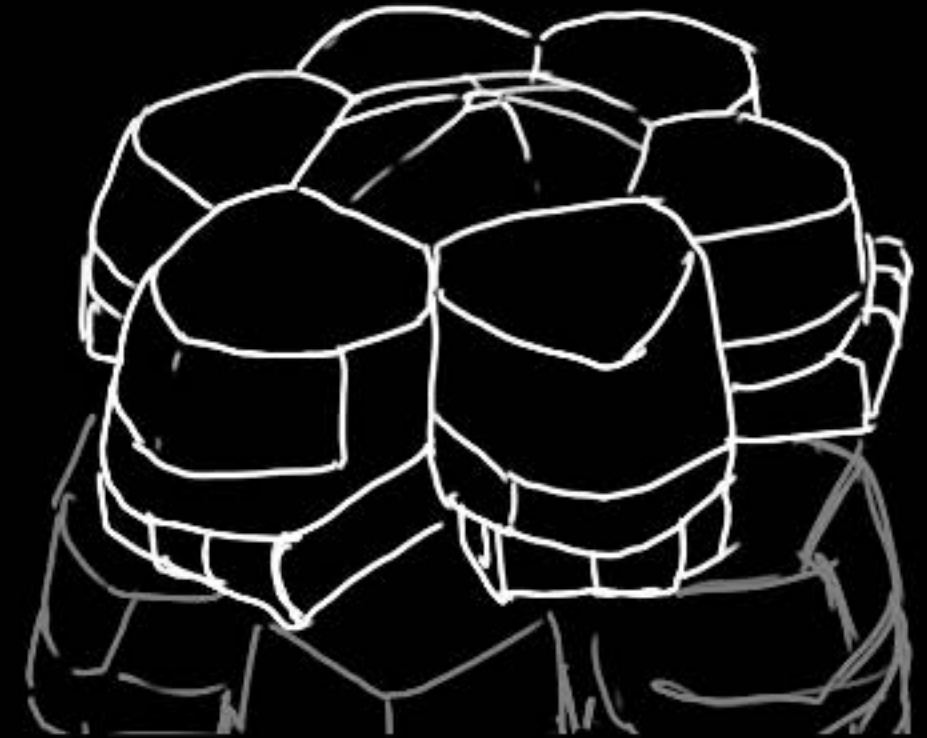


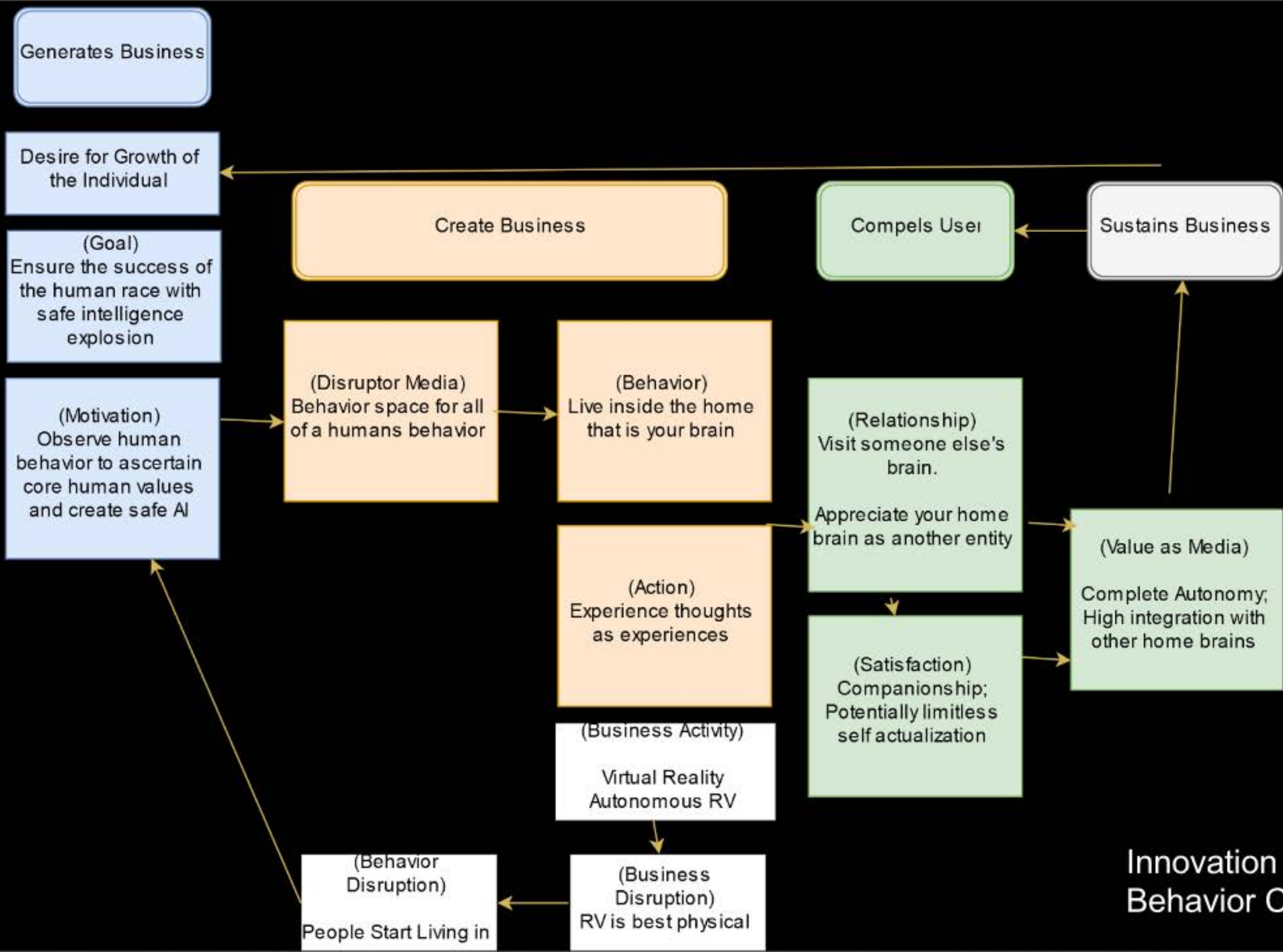
The Brain Interface





In Person Interaction





Innovation
Behavior Canvas

There are two general schools of thought on the likely path to safe super artificial intelligence.

There are two general schools of thought on the likely path to safe super artificial intelligence.

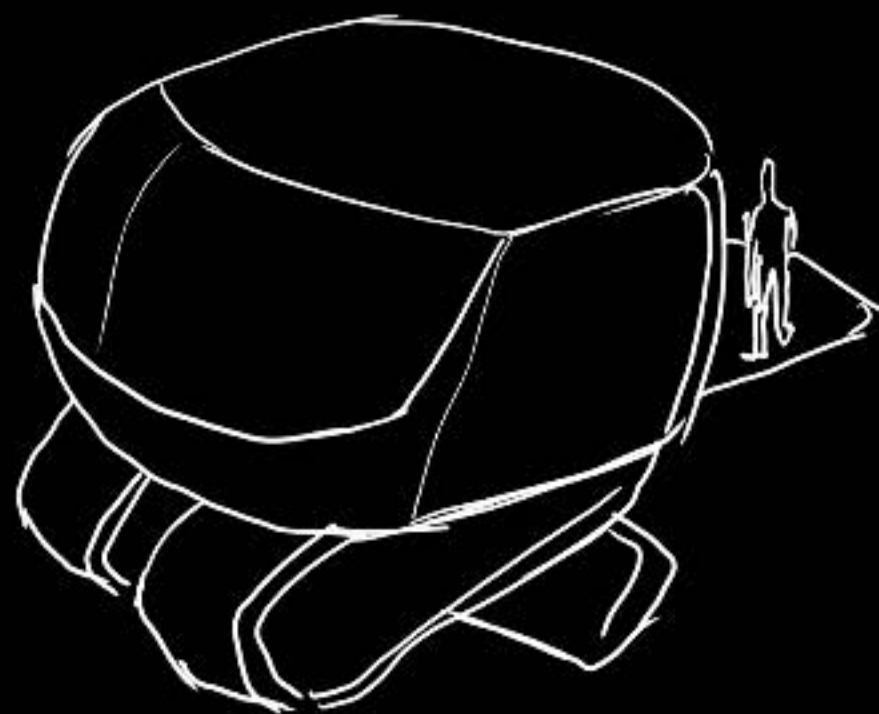
Observe human behaviour and ascertain core human values. Then compare future autonomous decisions against the value model.

Here are two common schools of thought on
“the likely path to safe super artificial
intelligence.”

Observe human behaviour and ascertain core
human values. Then compare future
autonomous decisions against the value model.

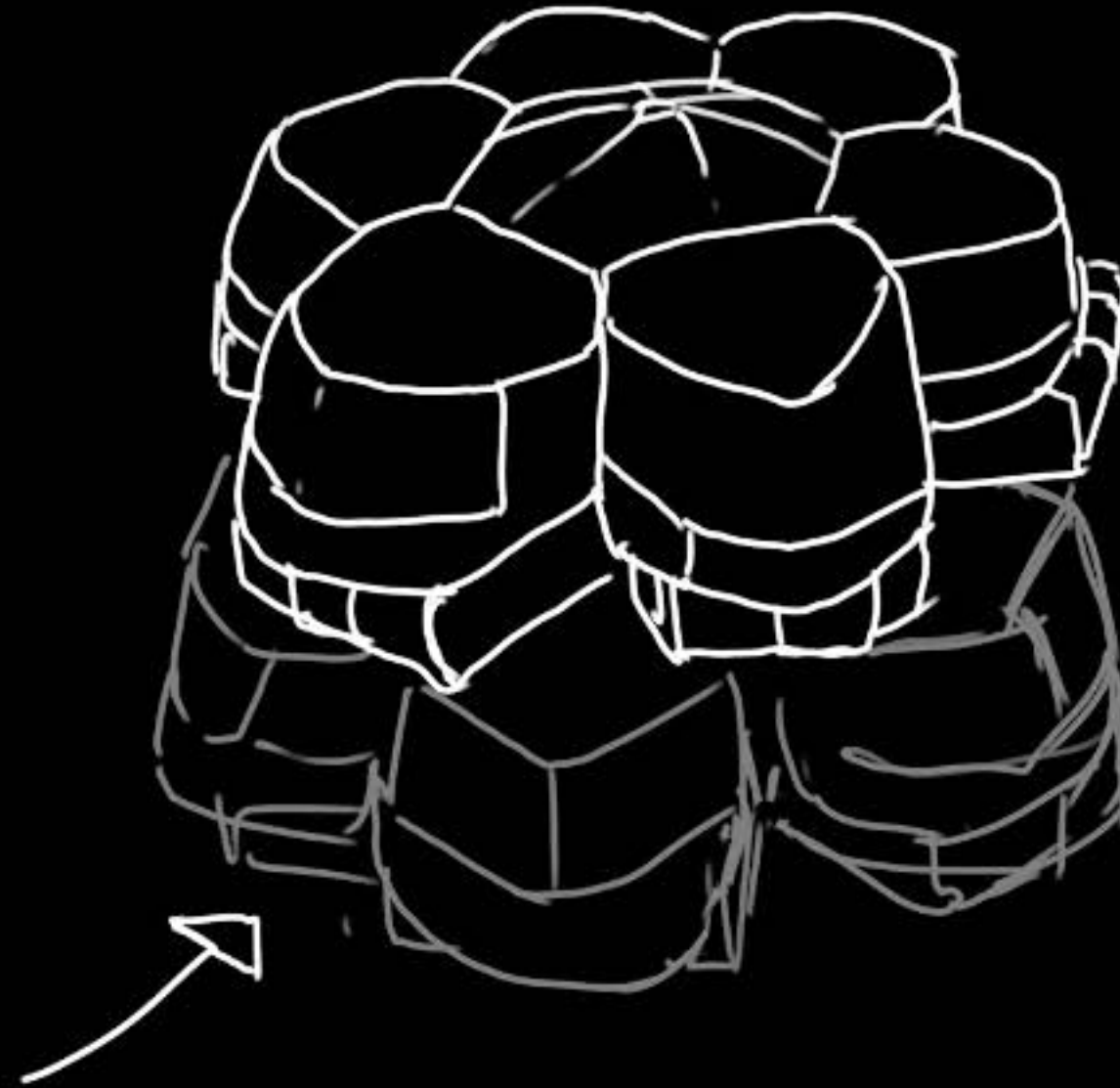
Interface with artificial intelligence through a
neural lace attached to the neo-cortex. Become
the limbic system of an artificial intelligence,
making us co-dependant.

I think living inside a robot home is a comfortable middle ground between these two solutions.



PS:

After watching "Mobile is Eating the World" I think Machine Learning answering questions for us (removing choice), may be sufficient for observing human behaviour. But I still think the Brain House is cool.



Works Cited

David R. Lawrence , César Palacios-González ,
and John Harris. "Artificial Intelligence."
*Cambridge Quarterly of Healthcare Ethics Camb
Q Healthc Ethics* 25.02 (2016): 250-61. Web.

"What Happens When Our Computers Get
Smarter than We Are?" *Nick Bostrom: What
Happens When Our Computers Get Smarter than
We Are? / TED Talk / TED.com*. Web. 30 Nov. 2016.

Yudkowsky, Eliezer. "The Simple Truth."
Yudkowsky.net. N.p., n.d. Web. 28 Oct. 2016.

Yudkowsky, Eliezer. "AI-box Experiment."
RationalWiki.org. N.p., n.d. Web. 27 Oct. 2016.

Yudkowsky, Eliezer. "Bayes' Theorem."
Yudkowsky.net. N.p., n.d. Web. 28 Oct. 2016.

O'leary, Daniel E. "Artificial Intelligence and Big
Data." *IEEE Intelligent Systems* 28.2 (2013): 96-99.
Web.

*Ray Kurzweil — Immortality By 2045 / Global
Future 2045 Congress 2013*. Perf. Ray Kurzweil.
N.p., 18 Jan. 2015. Web. 22 Oct. 2016.

Bostrom, Nick. "The Superintelligent Will:
Motivation and Instrumental Rationality in
Advanced Artificial Agents." *Minds and Machines*
22.2 (2012): 71-85. Web.

Works Cited

Eliezer Yudkowsky on Intelligence Explosion. Perf.
Eliezer Yudkowsky. *Youtube.com*. N.p., 21 Feb.
2012. Web. 26 Oct. 2016.

Chalmers, David J. "The Singularity: A
Philosophical Analysis" *From Time Travel to
Superintelligence Science Fiction and Philosophy*
(2016): 171-224. Web.