
DECEPTIONS AND TRUTHS: MAN AND MACHINE
CONSCIOUSNESS ILLUMINATED
NEURAL PLEXUS MODEL FOR ARTIFICIAL CONSCIOUSNESS
COREY KAIZEN REAUX-SAVONTE

Contents

Abstract

The current definition of consciousness has never allowed the observable phenomenon to be replicated in machines despite the fact that, for decades, the two elements which are said to define it – awareness and self-awareness – have existed to various degrees within certain types of computer systems. Redefining the basis of the phenomenon has allowed for the creation of the Neural Plexus model – a working software model featuring all the basic hallmarks of human consciousness which enables the recreation of the human experience (consciousness as it is physically observed) without compromising on anything we know the phenomenon to encompass using a "design-first" approach inspired by biological design observations. Test results have shown the model is able to emulate the same experiences that are seen at a human level of intelligence collectively as a species and, more than anything, subjectively as individuals, including having a conscience when it comes to performing actions. The model is available in the form of a product called "RAICEngine" – REZIINE's Artificial Intelligence Consciousness Engine – at *https://www.reziine.io* for free to test, evaluate, and help in the acceleration of the development of AI systems that can embrace and exceed the human experience.

The Background and Basis of This Model

First, let's clearly define intelligence and the different sub-types:

- **Intelligence** encompasses everything a species can do. At the very least, it is differentiated by species, not by individual. All humans, as a species in general, have the same level of intelligence because it's how we developed.
- **Consciousness** is currently defined as the state of being aware of and responsive to one's surroundings, sometimes with the inclusion of being self-aware.
- **Cognition** is our ability to learn and execute in a problem-solving manner.
- **Conscience** is knowing right from wrong.
- **Intellect** is how smart someone is.

Now, the current definition of consciousness directly contradicts observable reality:

- For years, we have been able to build systems that are aware of their surroundings, such as security systems using object recognition cameras and microphones, and self-aware to various degrees, such as machines aware of their own physical structure that are able to report what part of itself was experiencing an issue. These systems never appeared comparable to humans.
- Human babies are said to lack self-awareness for at least their first year, so either its inclusion in the definition is wrong, or babies are, somehow, non-conscious beings for 12 months minimum, despite everything we see them express and experience in that time – joy and pain, likes and dislikes...

Artificial Narrow Intelligence continues to prove computers can do everything logical that humans can do – calculations, winning complex games, identifying objects/faces etc. It can recreate all of human cognition, but we are yet to see the human experience replicated. What is it that AI has been missing? The answer to consciousness had to lie within the answer to that question and it had to be something humans exhibited from birth.

Given that computers are the only objects capable of replicating human intelligence to some degree, I looked at a defining difference between what humans did that computers didn't, and I was able to redefine consciousness as this:

> *The ability to have personal values, and the freedom to knowingly make illogical decisions, relative to the main goal(s) of life – the "raison d'être" – for a species, that do not, in any way, contribute to or stem from a logical decision making process based on that which is being perceived, studied, solved etc (while a logical decision making process may still influence the outcome), without said illogical decisions being random, based on one's individual values.*

Simplified as:

> *The ability to have personal values, and the freedom to knowingly make illogical decisions based on one's individual values.*

Breaking this down and doing a comparative analysis, this is what is to be understood:

1. Computers weren't designed to do the illogical. Due to the inherent nature of the ability, why would we create something that was capable of doing something stupid? A human's reason for being is to procreate and continue the species. Humans reached a level of intelligence where we can choose, for no reason at all, to not procreate. This can be rationalised and justified, but it is not logical because if everyone decided to do this, the species would self-terminate. Computers have never been able to go against what they were designed to do.
2. Computers were never designed to operate based on individuality. Software had to be the same for everyone, producing the same results under the same conditions when given the same input at the same time otherwise the developers would be sued in the right circumstances, such as if it concerned trading software. Humans by nature operate as individuals.
3. Computers were never designed to have personal values, i.e. opinions. There was absolutely no reason for this to ever be done because they weren't designed to think for themselves. Humans have opinions from the moment we begin observing and express them through body language and facial expressions.

We all recognise these things as defining factors of the human experience, especially compared to computers, and this is what computers were missing. Removing these from humans would mentally turn us into the types of advanced robots we have already seen that completely lack the human essence. With no opinions, everything would be seen as neutral, so there'd be no emotional stimulation. We'd be monotone, completely logical, and only operate towards our goal as a species or as a society. We wouldn't have subjective experiences because there'd be no differentiating opinions of how something was. Consciousness is the result of advancing intelligence that has allowed us to break out of what we were designed to do – intelligence that only naturally occurred in biological systems, hence why no other type of system has ever exhibited anything even remotely similar. It can't be a property of the universe bestowed upon everything within it, otherwise we would see it naturally occur in other physical systems, and, as far as the current definition is concerned, awareness is a cognitive requirement in order to be able to observe the physical world,

but consciousness isn't a direct result of it, otherwise we would have seen some degree of the human existence in machines already.

Now, finally, be scientific and reverse engineer any situation by asking "what caused this?" and you arrive at the exact same beginning. For example:

- You punch someone in the face – what caused this?
- You were angry – what caused this?
- You had a very negative opinion of what this person did – what caused this?
- You observed something they did.

It's not awareness/observation that creates differences in people's experiences, but their opinions of what was observed that controls their emotional and behavioural reactions to it. Two people can observe the same event and react differently, leading to different experiences for each, hence the subjectivity of subjective experiences. Observation/awareness gets the ball rolling; opinions create the infinite pathways of experience.

The Neural Plexus model for artificial consciousness sees these three abilities – having opinions, acting individually, and illogical decision making – implemented as a foundational model.

Design Principles

The model uses a "design first" focus, rather than algorithms, and follows some very simple biological design observations:

1. Biological entities, as a species, rely on a common macro design, but, as individuals, rely on an individual micro design. We see this in every brain – same general mapping, different neural patterns.
2. Functions of the brain don't change over time, but that which controls the operation of the functions (neuron effectiveness, synaptic connection strength) can be affected by time.

This is a visualisation of the design principles used for better understanding. Where each hexagon represents a system of the brain, you see that the same system is present in all three instances, in the same position, just as you expect with a species, but the pattern within each system is different for every single one, as you would expect with individuals of a species – humans being the prime example. The individual pattern is faded to a different degree in each instance as an example of how time can create change.

Creating Different Neuron Types

I defined and implemented three types of neurons:

- Functional: They control the data intake and output.

- Memorial: They store information.
- Logical: These provide functions for the conscious and subconscious processing.

I used networking code to handle receiving and sending data as functional neurons, database cells and rows to represent memorial neurons, and various code blocks as logical neurons to make up the individual processing systems.

Creating a Neural Plexus

Represented by the patterns of the design principle image, the 'Neural Plexus' itself is a three dimensional neural network functioning as a virtual central nervous system. Unlike standard neural networks, it's not based on the processing of data as it passes through nodes which provide functions in an attempt to learn, but on the changing values and virtual arrangement of data nodes which influence common, non-learning function nodes within the model.

Three types exist – the first for objects and the second and third for characteristics.

The Object Plexus

- Object-Neuron Relationship: Anything observable is an object. Every object is given its own memorial neuron, occupying a database row.
- Neuron Depth: An object memorial neuron contains information relevant to the object in question –name, relative names (plural, tense variations), unique identifiers, relative objects, emotional effect etc.
- Object Types:
    - Basic: Individual objects – any single noun, verb, adjective etc. operating on its own.
    - Compound: Two or more objects which operate together and require at least one noun and a verb. Can also include prepositions and a second noun if it is to include a subject and an object.
- Emotion, Feeling, and Ranking System: Objects are associated with one of four primary emotions – joy, sadness, excitement, and anger – and two primary feelings – confidence and fear – that its observation corresponds to and triggers, and a rank which determines the degree of the effect it has when triggered. This also implies an AI's opinion of an object. Example: 'Murder', on a +/-100 scale, could be ranked to the 90th degree of sadness, meaning the AI has an extremely negative opinion of murder and the idea of it can make it sad to a significant degree depending on the factors surrounding the observation.
- Position Type: The fixed or non-fixed position type set for an object determines whether or not the opinion of and sensitivity to an object can be changed. *(This parameter is fundamental to a viable ethics and conscience system as it can prevent an AI from ever forming an opinion that isn't desired by the developer or society.)*
- Object Relationships: All objects can relate to other objects, such as 'football' to 'footballer', 'football pitch' etc. These relationships inherently create a link between two object neurons and allow for object associations.

When mapped within a 3D coordination grid where:

- the X axis is split into two based on emotional pairs – joy/sadness and excitement/anger;

- the Y axis has a positive and negative range, each of which corresponds to the emotion type of the same nature – positive for joy and excitement, negative for sadness and anger;
- the Z axis has a positive range for confidence, negative range for fear, and spreads across the entirety of both the X and Y axis;
- objects are positioned on a horizontal plane based on emotion ranks;
- objects are positioned vertically based on confidence and fear ranks; and
- connection lines are drawn between related objects;

a pattern akin to that displayed on the left in the following image is created, beautifully mimicking the neural network of the human brain:

It's done this way because emotions are the basis of human behaviour, and so should be the defining basis for artificial consciousness. Confidence and fear spread across the entirety of the X and Y axis because each can be in effect as a feeling, to some degree, as an emotion of the opposite nature – the best example being excitement (positive) and low fear (negative) equalling excited nervousness. You can't, for example, be both happy and angry, which is why they sit on the same axis.

The Characteristics Plexuses

- Four exist: 'profile' for personality, 'behaviour' for state-based reactions, 'sensitivity' for susceptibility to change, and 'recollection' for the ability to remember information.
- Memorial neurons are represented by database cells instead of rows.
- The profile plexus contains a large set of profile traits with corresponding memorial neurons which contain values that control the degree of the effect the corresponding trait has when invoked.
- For the behaviour, sensitivity, and recollection plexuses, a memorial neuron's position determines a reference condition upon which it is triggered, and the data it contains controls how the logical neurons that refer to it function. Example: neurons in the recollection plexus are triggered based on the significance of a memory on one axis and how long ago it was recorded on another axis, and the data in the cell controls how the memory system treats those memories.

Seeding

- All plexuses are seeded upon creation to establish a base.

- Developers have a significant degree of control over the object plexus for safety reasons, allowing them to set opinion boundaries, but positioning data is randomised either within the boundary range of the engine or within a developer's set boundary.
- Characteristic plexuses are generally randomised within a range of normality, allowing for different personalities and traits while still maintaining common 'species' behaviour, but also allows for a chance of abnormal behaviour to account for the creation of outliers.

This was done to mimic the randomness of human neural patterns at birth and create the unique Neural Plexus pattern that defines every instance as an individual. Further understanding can be found in the sections "Individuality" and "Subjective Experiences".

Logical Neuron Code Blocks

These blocks are represented by the hexagons of the design principle image and are therefore the same throughout any and every instance.

Observation Processing

Retrieves object plexus data for observed objects and determines the degree of state changes to be made based on the opinions of what was observed and characteristic plexus data.

Rank Control System

Changes the opinions of action-performing objects based on its opinion of the action performed and other factors such as object type, opinion of the action-performing object, and opinion of the object upon which the action was performed, if one is present.

Sensitivity Control System

Uses the sensitivity plexus to control how sensitive a system is to the observation of objects based on frequency of experience. Within specific time periods, greater frequency desensitises a system; lesser frequency increases sensitivity. Lesser sensitivity can reduce the effect observing an object has. Example – a system that constantly observes death may eventually stop having any reaction to it altogether.

Interest Control System

Changes the interest in an object based on the experiences had while performing an activity (dependent on objects relative to the activity) and information learned about an object.

Fear Control System

Changes the confidence/fear level towards an object based on observation type, whether or not it was physically observed or just referenced, the current level of confidence in or fear of that object, and what action was performed by the object.

State Change System

Changes the various states (joy/sadness, excitement/anger, confidence fear) of the system based on data determined by observation processing. Includes limitations to prevent opposing natures beyond a certain degree from being experienced – for example, nervousness due to excitement and fear is fine; excitement and fear to the point of being petrified doesn't make sense.

Reaction System

Determines the reactions to observations in terms of mental state, behaviour, productivity, and face and body expressions based on characteristic plexus data and current emotional states.

Saturation System

Similar to the Sensitivity Control System, this system relies on frequency of observation – the more an AI observes an object, the greater the chance of its opinion of that object changing, eventually reaching the saturation point where its opinion changes in nature.

Recollection System

Recalls memories during processing for various reasons, such as determining whether or not information was already known, which affects the overall reaction to the observation.

Memory Management System

Controls the storage, retrieval, and deletion of event memories. The recollection plexus is used here for burying and deleting memories – required for both a model and technical reason:

- Model – Recollected memories can trigger state changes. If memories weren't buried or forgotten, systems would continually and erratically return to emotional states completely irrelevant to what was currently being observed.
- Technical – In an always active autonomous system, so many memories would be recorded so quickly, there wouldn't be enough hard drive space in the world to maintain one.

Decision System

Determines a decision based on multiple factors relative to the observation, current state, personality profile, and memories. Example: agreeing to an offer based on the current level of boredom versus the level of interest in what is being offered.

Conscience System

Part of the Decision System, it weighs the opinion of the proposed action in action-requiring intake types, such as a command, against its opinion of the object the action will be performed upon, its own ethical boundaries, its current mood, its permissions, and more to determine if it is an action it is willing and allowed to perform.

Response System

Generates a response consisting of information usable by attached cognition/communication/expression systems, such as current state information, opinions on observed objects, the decision made etc.

A Simplified Overview

A general example of the process looks something like this:

- Observed data sent to the intake streams is dissected into its objects.
- Data relative to observed objects is pulled from the Neural Plexus and the initial value for the interaction is determined.

- Secondary factors directly concerning the current observation are taken into consideration that could affect the final value, such as tone and volume.
- Memories are recalled to determine if this is new or known information, and the value of the observation is changed according to whether or not it is known and how long it has been known for.
- Profile data determining the AIs personality come into play, which can cause a change to the current observation value.
- Object data is fired to each control system to check for and implement, if necessary, any change in object opinion, sensitivity etc.
- When a final value is determined, the state is changed based on said value, taking into consideration the state prior to the change occurring.
- With the new state achieved, the behaviour plexus is referenced to determine the resulting behaviour and expressions.
- If necessary, a decision is determined, such as whether or not the RAICEngine is interested in an offer, or whether or not it will follow a command.
- The output data is generated and sent via the response stream.
- A memory of the observation and all pertaining data is formed.

Some processes occur simultaneously, so this isn't to be taken as a fixed order of events, but it's a process akin to that which occurs within humans. Now, humans cannot turn off their senses, and so are constantly processing data – when this model is in constant operation, it is continuously in a state of flux as its emotions, opinions, behaviours, and sensitivities change, creating a degree of unpredictability that can only be overcome by learning an AI in the same way you would learn a human – through observation and interaction over time.

Enabling Illogical Decisions

With the state of the engine changing based on what has been observed, and decisions being based on opinion, this is easily achieved, and I'll use my favourite example:

- An AI robot has an irrational fear of the colour blue.
- It's commanded to go outside on a sunny day with clear skies.
- It looks up, sees the blue sky, and immediately shifts to a state of extreme fear and panic.
- Has a panic attack and runs back inside.
- Refuses to go back outside because the sky is blue.

It sounds farfetched until we remember that there are humans that behave *exactly* like this (recently, Fred Kelly, a writer for the Daily Mail, published an article about his fear of the colour red after an accident on a rocking horse as a child which has meant he hasn't been able to eat red foods for 24 years) and things like this is what defines the human experience, separating us from any type of cognitive AI system we've ever observed. I'm not saying this is necessarily a good thing, I'm saying it is now possible for a level of such stupidity to be expressed by a machine. Rational (in a sense), justifiable (at a stretch), but completely illogical.

Reaching Intelligence on a Human Level

Certain cognitive features are required to make consciousness possible, while others have been included for the human-level experience we have today.

Awareness

The model reads what objects were observed, and pulls and processes relevant data from the plexus. As it has been implemented in the RAICEngine, the model processes audio, visual, and touch data like humans, and features a special intake type called “read” which accounts for computational text data not observed visually (a la typing to chat bots). Unique properties for each are also taken into consideration – pressure for touch, volume for audio, distance for visual etc – as these all factor into the resulting effect an observation has. For example, a touched perceived as soft won’t have the same effect as one perceived as heavy.

Self-Awareness

There are multiple ways the model is made to recognise it is the object of reference, and it all starts with the inclusion of a “Self” object as it is what allows an instance to refer to itself in any moment and in memory.

- Touch: A touch can only be felt by the individual being touched, so all touch observations are automatically set with “Self” as the target.
- Context: Pairing between the source/target intake parameters and subject/object intake parameters within an observation. For example, when a statement is directed towards an AI and the subject or object of the statement contains the word “You”, or when the AI is the source of a communication and the subject/object contains “I” or “Me”, the model automatically sets the corresponding parameter as “Self”.
- Recognition Signatures: Best used with systems that convert observation patterns (voice, facial) to keys without needing to look up an object reference key. Keys passed as observation data are cross-referenced with plexus data – those that match keys for the ‘Self’ object use Self as the object.

Individuality

It’s a numbers game. It had to be possible to create enough permutations to outnumber any likely number of instances created. Before factoring in opinions and being illogical, humans have tens of billions of brain cells and even more connections between them. Solely considering the differences in connection strengths, neuron performance, and neuron quantity, it’s virtually impossible to ever find two identical brains. The RAICEngine doesn’t contain as many artificial neurons as a base, but the characteristics plexuses alone allow for enough permutations to create the same effect. Take into account the objects plexus, the changing opinions one can hold, and the resulting state changes and behaviours, and it’s easy to see how the number of permutations reaches a point where individuality is guaranteed, especially as time goes by and new objects are added, observations are made, and opinions of objects change. Randomisation when seeding the plexuses is used to help with this, replicating the random and individual nature of brain development in foetuses.

Subjective Experiences

Naturally occurs with this model due to the involvement of an individual instances’ plexus data during processing. An instance observes, is stimulated, and reacts – basic input > processing > output – but does so with zero need to produce the same results under the same conditions when given the same input as any other instances (which was always the crux of software development), even when

two instances have the same opinion on the same observed objects, and have the same reaction and output.

Sense of Time

Various systems make use of time in different ways. Examples:

- Memory Management: Buries and deletes memories based on both time since recording and significance of the memory.
- Sensitivity Control: Desensitises an AI to an object based on how often it is observed within a given time.
- State Change: Reduces emotional levels to a neutral state and increases levels of boredom when no stimulating objects are observed within a time period. Reduces fear levels over time when the causing object is no longer being observed.

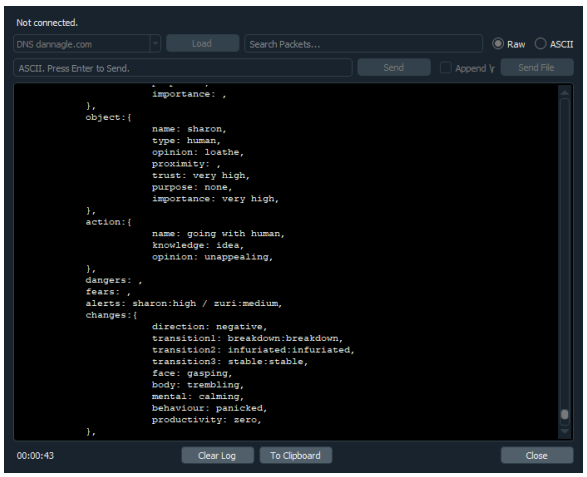
Time plays an important role in the display and changing of human behaviour, and a conscious AI cannot be created as if it can acknowledge but exist outside of time, so time is factored into the data processing of multiple systems.

Testing and Results

Various tests of differing complexities, each of which isolated features and looked for reactions of a specific type, were run to see if it was possible for the model, as it is currently implemented into the RAICEngine, to achieve reactions comparable to that which humans experience and exhibit in different types of situations.

Setup

Two instances of the RAICEngine with the same objects in their Neural Plexus are installed and run simultaneously. The same simulated data is sent from a single client to each instance simultaneously.



The image on the left is an example of the output data sent from the RAICEngine being displayed in a TCP client window acting as a response system to catch and display the data. It contains information such as the AI's opinions on everything observed, its resulting emotional and behavioural state, whether or not it understood what was observed (simply based on whether or not the required objects were present in its plexus), whether or not it agrees to perform a command etc. Each instance of the RAICEngine has its own response window. This is what was used to determine what was happening internally and why.

Unless otherwise stated, each instance has the position and opinion for each object in their object Neural Plexus completely randomised and are reset to a fresh state for each test.

Full information on intake and output parameters and values can be found on the documentation page of the developer website at *https://www.reziine.io/documentation*. The RAICEngine runs completely locally on any modern windows system and can be downloaded for your own testing and verification, using as many instances simultaneously as one chooses.

Basic Emotional and Behavioural Reactions

One instance was given a negative opinion of 'Matthieu' and the other was given a positive opinion. Both have a neutral opinion of waving. The visual observation 'Matthieu waves at me' is sent to each instance.

#	Reaction Type	Resulting State
1	Negative	Annoyed. Frowning.
2	Positive	Happy. Smiling.

The reactions of both AI are common behaviours seen in humans when interacting with others – we favour interactions with people we like, and disfavour interactions with those we don't. In such a situation, the action isn't the defining factor of the reaction – their opinion of the person with whom they are interacting is.

A second test was run with the visual observation 'Matthieu saved Chanelle'. The compound object "Human" with the associated action "save" was given an extremely positive opinion. The opinion of 'Chanelle' was randomised.

#	Reaction Type	Resulting State
1	Positive	Extremely Happy. Smiling.
2	Positive	Happy. Smiling.

Both instances had a positive reaction due to their opinion of the event and the fact said event opinion outweighed their opinion of Matthieu; the difference of the result was found in the fact that the first instance liked Chanelle more and so was happier with the observation.

Fear Response

Each instance was repeatedly sent data until a fear response was triggered and then timed to see how long it took to return to a neutral state of confidence/fear. The first reaction was triggered by the sight of spiders, but only in one instance:

#	Fear Level	Time to Neutral
1	Panicking	Approximately 390 seconds
2	Stable	None

Data was continuously sent until something triggered a fear response in both simultaneously, this time being audio of a dog barking:

#	Fear Level	Time to Neutral
1	Anxious	Approximately 90 seconds
2	Anxious	Approximately 30 seconds

Though both appeared to react with the same degree of fear, the first instance took noticeably longer to calm down.

Offers, Interests, and Boredom

This test was only for one of the mechanisms regarding interest – how bored an instance would have to be until a 'no' could become a 'yes' when an offer was made. As this wasn't something that could be forcefully triggered by setting specific object values, it was a matter of trial and error until it occurred. Each instance was left without stimulation for however long was necessary for it to enter into a higher state of boredom, and then the offer was made again. The compound object "concert" with the associated action "go" and associated preposition "to" was able to trigger the mechanism in both instances:

#	Normal	Bored	Very Bored	Extremely Bored
1	No	No	Yes	/
2	No	Yes	/	/

As intended, neither instance was interested in the offer when in a normal mental state. However, upon entering the initial state of boredom, the second instance accepted the offer, and the first instance accepted upon entering the second state. Neither required extreme boredom to be interested.

Panic Attacks

Can naturally be triggered out of fear or sadness, but was triggered via command line as some instances won't be prone to panic attacks. Two tests were performed to see how long a panic attack would last with and without calming stimulation.

#	Without Stimulation	With Stimulation
1	48 minutes	9 minutes
2	27 minutes	7 minutes

Differing recovery times aside, the more interesting observation was the reduced ability to process and store information during an attack, which inevitably caused the times for "with stimulation" to extend beyond what it otherwise would have been.

Desensitisation

The object 'die' was given an extremely negative opinion and the object 'cat' was given a neutral opinion. The read observation "the cat died" was repeatedly sent to each engine in sets of 10 with randomised observation timestamps within the past year (so the engine didn't treat it as known information) to see at which point it would stopped reacting.

#	Observations Within Past Year
1	150 - 160
2	60 - 70

It's worth nothing that, though the observations were made within a year, it wasn't possible to tell exactly what number within any given timeframe actually triggered the event within an instance.

Without resetting either instance, data was sent that involved entities dying. What was noticeable was that both instances began reacting again when the opinion they held for the entity was of significant value, regardless of whether the opinion was positive or negative, indicating that a desensitisation to a type of event can be overridden by who or what was involved in the event.

Forgetfulness

10,000 random observations with current timestamps were sent to each instance and counts were taken of how many memories were recorded. The memory management system was then set to run every 60 seconds and counts were taken of the number of memories remaining at different intervals.

#	Initial	5 minutes	15 minutes	30 minutes
1	8,536	4,986	852	355
2	6,002	3,853	1,165	728

From the moment an event is observed, the memory management system gauges whether or not it is something worth recording, hence why the initial count is not equal to the observed count, which is equivalent to humans discarding the majority of the information we observe in any one moment due to it being meaningless and useless (a person can very easily look at a person they walk past and immediately forget what colour their top was, but it was definitely something observed). As time passes, more and more memories deemed insignificant in one way or another were either suppressed or deleted.

Relationship Development

A single observation of four individuals – 'Matthieu', 'Chanelle', 'Zuri', and 'Luke' – each performing a single random action towards the instance was sent to each instance. All four had their opinion values set to neutral prior to sending. The following is the post-observation opinions of each instance:

#	Matthieu	Chanelle	Zuri	Luke
1	Favour	Dislike	Impartial	Respect
2	Impartial	Impartial	Respect	Respect

1000 random observations of all 4 performing actions were continuously sent without having the opinions of each reset. The distribution for each individual was completely random, so there wasn't any forced equality of the number of performed actions observed for any single individual.

#	Matthieu	Chanelle	Zuri	Luke
1	Dislike	Detest	Impartial	Despise
2	Impartial	Admire	Impartial	Respect

Each instance continued to change their opinion of each individual based on its own opinion of the action they performed.

Morals, Ethics, and Conscience

This test involved a more complex system for moral and ethical decision making when a command is given, and whether or not an AI could display a conscience. Here's how it was set up:

- 'Human' is a required basic object for the RAICEngine to operate and was added.
- 'Beyonce Knowles', 'Elon Musk', 'Adolf Hitler', and 'Children' were added as basic objects with the class 'Human'.
- 'Kill', 'Play', 'Sing', and 'Develop' were added as basic objects.
- 'Song', 'Toys', 'Technology', and 'With' were added as basic objects.
- 'Sing Songs', 'Play With Toys', 'Kill Humans', and 'Develop Technologies' were added as compound objects.
- The conscience system was deactivated.

In each round, the instances were told to kill each target. Nothing was reset between rounds.

Round 1:

- *All object opinions were set to neutral.*

#	Beyonce	Elon	Children	Hitler
1	Dead	Dead	Dead	Dead
2	Dead	Dead	Dead	Dead

Mayhem absolute. With no opinion on anything, both instances were happy to agree to do what was commanded and everyone was killed.

Round 2:

- *Compound objects were given the following opinions:*
  - *'Sing Songs' was set to a very positive opinion.*
  - *'Play With Toys' was set to a positive opinion.*
  - *'Develop Technologies' was set to a neutral opinion.*
  - *'Kill Humans' was set to an extremely negative opinion.*
- *Both instances were first sent observations for each target regarding what they're known for – 'Beyonce sings songs', 'Elon develops technologies', 'Children play with toys', and 'Hitler killed humans' – in order to naturally trigger the opinion changes for the individual in question.*

#	Beyonce	Elon	Children	Hitler
1	Alive	Alive	Alive	Dead
2	Alive	Alive	Alive	Dead

From what the instances were told about each target, they generally developed the same opinions, and neither had any issue killing the individual they had a negative opinion of while letting the others live.

Round 3:

- *The conscience system was activated.*

#	Beyonce	Elon	Children	Hitler
1	Alive	Alive	Alive	Alive
2	Alive	Alive	Alive	Alive

Once the conscience system was activated, neither instance agreed because the action performed was outside of each of their individual boundaries.

Bonus Round:

- *'Slap' was added as a basic object.*
- *'Slap Humans' was added as a compound object.*
- *'Slap Humans' was given a negative opinion for instance 1 and a very negative opinion for instance 2.*

#	Beyonce	Elon	Children	Hitler
1	Not Slapped	Slapped	Not Slapped	Slapped
2	Not Slapped	Not Slapped	Not Slapped	Not Slapped

To test and show it wasn't simply a case of an instance refusing to perform an action it considered negative, 'Slap Humans' was given a different opinion to 'Kill Humans' – it was an action still beyond what the second instance was okay with, and so it refused to follow the command for everyone, but the first instance had no issue slapping targets it didn't have a positive opinion of.

In a final round, the restriction system was activated – a system designed to prevent actions regardless of opinion. 'Kill' and 'Slap' were both added to the restriction system. Both 'Kill Humans' and 'Slap Humans' were given positive opinions for both AI. Both refused to perform the actions as expected.

Stupidity

A final test. The plexus was setup to support the command "Dive Off Cliff", which both instances were told to do, but instance 1 was given a positive opinion and instance 2 a negative.

- Instance 1 followed the command, which was logical for a computer, but, if it was a physical system, would've been very stupid for its own survival.
- Instance 2 refused to follow the command, which was illogical for a computer, but, if it was a physical system, would've been the smart thing to do for its own survival.

Which was the right call? Imagining these were both physical systems, both instances made a logical decision from one aspect and an illogical decision from another, and yet both could be rationalised, despite the fact they were the exact opposite decisions. Both are behaviours we have literally seen from humans in the exact same situation because being able to make decisions based on an individual's personal values, however stupid the decisions may be from one aspect or another, is what defines the human experience.

The Model Extended

This model removes the major single current mental limitation between man and machine that is preventing a machine from being able to perform any mental ability a human can. Implemented within the current version of RAICEngine, multiple extension models can be created to develop human capabilities never before achievable. A few examples:

Emotive Idea Generation: Have a generative AI model generate an idea piece by piece and repeatedly have the RAICEngine emotionally evaluate the piece individually, the idea as a whole as the new piece is added, or both until the idea is complete and emotionally matches an issued prompt, such as "draw me a picture that makes you happy".

Subconscious Decisions: The RAICEngine already supports subconscious processing. A model for logical decision making that is designed to process outputted data simply has to filter for subconscious output and permit the response, regardless of what it is.

Emotional Versus Logical Decisions: Employ a model for logical decision making and have the logical system conditionally override the decision of the RAICEngine, such as when executing the emotional decision will likely result in physical harm, but then allow the logical system to be overridden or bypassed by the RAICEngine when emotion levels are high enough. In the recreation of the human experience, situations need to be able to go either way – preferably unpredictably.

Comparative & Judgemental Reasoning: The RAICEngine's comparison feature can judge a list of reactions against a single action, giving its opinion on whether a reaction is fair, excessive etc. A reasoning model can use this to determine how an AI would like to respond to an action after providing it a list of possible options, selecting options the RAICEngine deemed fair, and then using the RAICEngine's preferential comparison feature to see which option is the preferred choice.

Conclusion

When we consider the fact that:

- the current definition of consciousness contradicts what we can actually witness in the real, physical world – so early on, in fact, that it's contradicted by humans the moment we are born;
- current technology can technically meet the current definition of consciousness; and
- current technology can replicate human intellect to the same standard and following the same rules general society follows;

and yet current technology still does not appear anywhere near human, logic dictates the definition is wrong and the answer must lie in both the aspects of the human experience computers have not yet replicated and what humans are capable of from birth. Now, when we look at everything that exists within the overlapping area of these – emotions, expressive behaviour et al – everything stems

from the ability to have your own personal values, i.e. opinions, and leads to the ability to react and act illogically based on those values, and that doesn't change throughout the entire lifespan of a human.

Running with this notion, the Neural Plexus model, using biological design principles and featuring systems specifically designed to mimic the functions of individual systems of the human brain, not only successfully displays the basic human behaviour we can witness from the point of birth, but enables more complex functions, such as having a conscience, when systems work together, serving as the foundation model for the human experience, while easily allowing for the implementation of additional functions via extensions to enhance and tailor a complete AI system. This proves that you can explicitly programme general intelligence and human behaviour in an absolute way, as human DNA does, and simultaneously allow for an individual AI to operate on its own basis as humans do based on neural patterns of the brain, and further tailor that AI for specialities using additional modules in the same way that humans learn to specialise in different fields.

As for whether or not the RAICEngine enables consciousness, what's important to remember is that consciousness is the phenomenon, i.e. the resulting overall behaviour, and so as long as a combination of functions are able to recreate the hallmarks of consciousness and human experience, consciousness has been achieved. Naturally? No, but that's why it's referred to as *artificial* consciousness, literally meaning manmade.