

# 3D Simulation: the Key to A.I.

~~Keith A. Hoyes~~

Inca Research, Ltd.

keith@incaresearch.com -- <http://www.incaresearch.com>

**Summary.** The proposal is a radical one – that human cognition is significantly weaker than we presume and AI significantly closer than we dared hope. I believe that the human mind is largely made up of tricks and sleights of hand that enamor us with much pride; but our pedestal might not be quite so high or robust as we imagine. I will pursue the argument that human cognition is based largely on 3D simulation and as such is particularly vulnerable to co-option by future advances in animation software.

## 1 Introduction

*“A is A” – Ayn Rand*

Monsters Inc. was an entertaining film and like so many others of its genre, it allowed us, for a time, to enter a world that never really existed. To the computers that generated the images, the world doesn't exist either, it is just so many 1's and 0's. But those bits got transformed into a language we could all understand; a world we can feel, fear and predict. Our eyes similarly take a cryptic stream of bits and somehow too create a world we can feel and predict. If you close your eyes and imagine entering your kitchen to get soda, you must surely have created a 3D world to navigate. As you re-open your eyes, just how are those dancing 2D patterns you see converted into the 3D virtual realities in your mind? [7]

In the virtual world, when a princess kisses a frog it turns into a prince. The real world does not work that way. For general AI to solve real world problems, its thinking needs to be bound by real world behaviors. All significant phenomena in the real world exist in three dimensions, or can be expressed as such. The common language describing computers, bicycles and brains is that of their 3D material existences animated over time (A is A). Further, derivative concepts such as math, stock markets, software and emotion can similarly be bound. If a concept cannot be described in three dimensions over time, it is quite likely false. Like the frog above, it may exist only in some virtual domain [8].

The real world cannot violate the laws of physics, logic or axioms to enter a fantasy world – frogs to princes. But the virtual can. It can be bound or

unbound. But when bound to physics, it can accurately simulate reality. This has important consequences for AI.

Finally, the real world is bound by time. The virtual is not. It can run time backwards and forwards and at any speed. It can also accept time discontinuities, freezes and gaps. The virtual can predict events in reality before they have even happened! It can represent the now, the future or the past. It's when the real and the virtual are mixed, that the magic really begins.

## **2 Pillars of Intelligence**

### **2.1 Deep Blue**

To many, the victory of deep blue – a mere computer, over the smartest chess genius alive was both disconcerting and also raised hope of a new dawn for AI. But in the end it didn't really amount to much. Politicians still got elected, deep blue got de-commissioned and most computers still act more like glorified calculating machines than thinking people. The fact one such calculator beat the smartest guy in the world at chess was just a freak anomaly. Just how Deep Blue beat Kasparov I will explain. Though a much more important question, I should think, is – Just how could Kasparov, hope to beat a computer?

Deep Blue operated primarily on just one of the three pillars of intelligence – time travel. I'll explain. The important aspects of a chess game can be simulated quite perfectly in a computer. At any given instant of real time, the game will, obviously, be in its current real state. Deep blue took that state as its starting point. It made predictions, grading each outcome as far into the future as time and resources would permit. Its final move was thus calculated to have the greatest probability of success. And the rest, as they say, is history.

### **2.2 Virtual Reality**

Deep blue simulated a world very different from our own, But there are simulated environments in software labs, movie studios and military compounds all around the world that are very much more like our own. The fact that virtual worlds will soon become quite convincing and compelling is taken much for granted these days. It is only seen as a matter of time before the accuracy of simulations can completely fool the expectations of our senses. This will form the second pillar of intelligence.

### **2.3 The Humble Earthworm**

An earthworm doesn't know much about "pillars of intelligence," but it represents one all the same. It has no ability for virtual time travel, no machinery for creating virtual environments, but it can surely feel the real world around it; the resistance of soil; the drying heat of the sun; the injury from a bird

predator – and act accordingly. It might be argued that such a simple organism, which presents evidence of feeling, is in fact demonstrating basic reflexive responses to stimuli – like a thermostat. So when I say feel, what perhaps I should have said is the ability to discriminate within the flow of sensory inputs, that which it considers good from that which is bad. In other words, it has a sensorimotor system linking the real world to the virtual. **These then: virtual time travel; virtual reality and an information bridge to reality, I present, are the three pillars of intelligence**[6].

### 3 Consciousness

The moment a silicon eye can stare back into your peering eyes, unflinching and following your every move – then, you will believe in such a thing as an artificial soul. You may be wrong, but not entirely so. We intuitively know that a thing that can break through the veil of vision, to make sense out of that mess of light and shade; to really “see” a living, breathing human being, has crossed that Rubicon. It will actually be achieved through unconscious, mechanistic vision processing. But our intuition in this case will be right – 2D to 3D instantiation is indeed at the very heart of consciousness. (In animation, rendering is the process of converting 3D scenes to 2D bitmap images for human viewing. Instantiation, in this context, is the reverse – the conversion of 2D bitmaps back to 3D environments, this is a key concept in this paper).

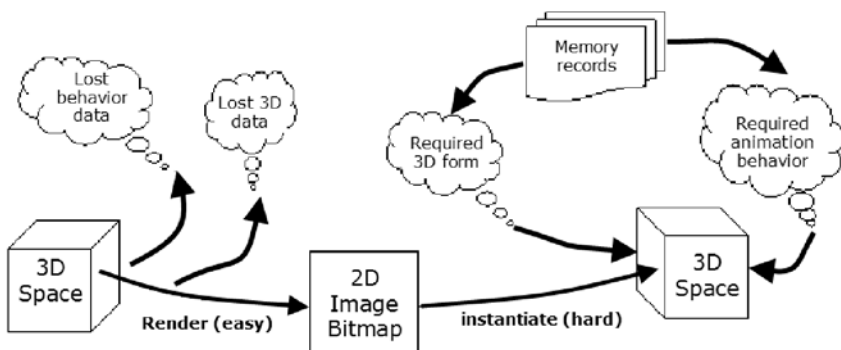


Fig. 1: Rendering and instantiation

Your basic human being is constructed from a virtual reality chamber connected to a biological, self assembling, nanotech robot with sensors. The chamber is self learning from exposure to the outside world and free will stems from a process of grading simulated predictions against pre-programmed genetic and culturally programmed schemas. **Without a simulated environment**

running behind our eyes, we would be totally blind. The stream of data can only ever represent a series of bit maps; there is no hidden information our eyes can see that a camera can't – there is less! The images are simply used as cues in the construction of a virtual environment. The contents of that environment are actually drawn from memory and the bitmaps simply maintain simulation alignment and paint texture over the model surfaces. The experience of consciousness is bound to that simulation.

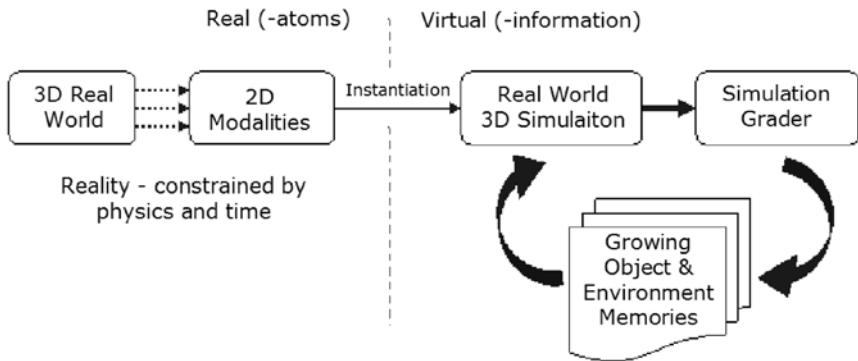


Fig. 2: Atoms and information

3.1 Feeling and Qualia

However fancy the arrangement, the human mind is still, only made from matter. There is no mysterious essence of feeling from magic atoms hidden away in the corners, and non-invisible real atoms do not feel! That leaves only one cause – information. When energy and matter are able to represent coherent information, such information can subsequently be graded, and that information interpreted as feeling. It is a computational trick; an illusion, necessary to control the behavior of biological creatures. Evolution uses the process of consciousness and the subjective “closed loop” belief in feelings, to guide behavior toward the survival and reproduction of genes. Although we may assert pleasure and pain to be computational illusions, we have no conscious control over the process; so telling ourselves pain is just information will not work[1].

For humans there is a first person relationship between the sense modalities and the affect within the mind. Every sensory receptor – whether from touch, sight, sound, smell or taste, will flow into memory somewhere, and maps directly to the first person perspective in a simulated environment. This nexus represents the “eye” or “I” of consciousness. Together with the muscles, the

sensorimotor system forges the creation of a simulated environment which is processed by filters to grade simulations according to genetic and cultural presets. These analyses produce the illusion of feeling and emotion. They are used to guide subsequent cognitive attention.

But this information can have no meaning until it is grounded, through instantiation, to virtual objects that have form and invariably a history timeline (behavior). What we actually perceive are virtual objects within our own minds, the senses are used to align these objects to external reality, to consolidate existing memories, or to train new memories if the objects are novel. Once the sensory flows are aligned to precedents, the scene can then become known. Not simply because of the informational connection to a form – the instantiation. But because the virtual object forms have known behavior precedent potentials and a “spatial” home within the mirror-world simulation.

This is the point at which subconsciousness can take hold, by taking these behavior precedent options and running trials “subconsciously,” away from the perceived scene – which may be linked or not to reality through the sensorimotor system. Subconscious simulations are fast, dynamic simulations, that seek out narrative with significant grading points. And this is where the issue of emotions and feeling states – “qualia” enter the picture. Genetically, the brain is programmed, and programmable, with a value hierarchy. Just as the eyes are formed in expectation of light, so the brain is formed with memory references in expectation of information against which to compare. We describe the subsequent gradings as our feelings and emotions. The most obvious example being the sexual beauty (form) and grace (behavior) of the opposite sex, emotional recognition that is innate.

The higher speed of subconsciousness is necessary to discover scene outcomes ahead of real world time. Such that actions aligned to goals can be discovered before it is too late – such as catching a ball. Emotional grading of simulations is generally more intense if they are currently aligned to reality through the senses. This motivates action in preference to reflection.

So, to summarize, the brain contains information describing object forms and behaviors. These memories are organized into a spatial hierarchy mimicking the external world. Some memories are created by the genes, but the bulk are forged into memory as the sensorimotor system interacts with reality. The contents of consciousness are the scene alignments currently in resonance within memory and reflecting back to the sensory cortices, such that the sensory envelope of the modalities can extend to embrace imaginary worlds. Subconscious processes are resonances not currently aligned to the sensory cortices, although fully capable of being emotionally graded, leading to non conscious feeling states and motivations. The brain needs subconscious processes (i.e. the simulation and grading of memory precedents) in order to discover choices upon which to align consciousness and/or physical actions.

The analog nature of human biology, which interfaces electrical, chemical and cellular processes beneath the computations of “mind,” leads to physical pathways, sensitized chemical boundaries, linking the computation of feeling

to the sensory “feeling” of feeling – to qualia. There is strong evidence to suggest that supplemental sensory regions exist beyond the traditional modalities, mapped instead to an existential feeling space deep inside the body, but in actuality merely extending across chemical boundaries within the brain. Such a mechanism would provide powerful feedback paths of the same “class of feeling sensation” as from the touch senses, but without the concomitant external body surface mapping. Instead, it is as if some “phantom limb” were at the body’s core. Thus the feeling effects of subconscious emotional script analysis will have physical manifestations. Evidence from pharmacology clearly points to the existence of chemical pathways affecting emotion and states of mind.

Like a “second sense,” emotion would impart an evolutionary advantage even before the emergence of higher cognition. Since primitive emotions can provide effective shortcuts to otherwise complex, or slow, effortful cognitive processes. It can often be seen directly in children before they learn to subordinate their emotions to their emerging wider scope cognition. This same effect occurs in the processing of language, providing short cuts to understanding. Much of our social language is predominantly emotional, often pre-empting and short circuiting rational thought; since the whole meaning really is meant to be just the emotion tags. It would often be considered quite disingenuous to even attempt a rational analysis. Spock comes to mind here!

Emotions are not only used by the brain to grade simulations, they can also be linked to objects to help predict their behaviors. **Animation within a simulated environment will involve causes, objects (actors) and effects.** The emotional states of objects (which include people and animals, as well as inanimate objects) originate from the context, the initial conditions and from the historic memory records. This empathic knowledge within the simulation is different from the first person emotional analysis of sub consciousness used to grade the scripts. It instead provides behavior cues to more accurately guide the simulation. For example, empathic knowledge of joy or anger in a character will significantly affect their expected behaviors and interactions. Even traditionally inanimate objects can be injected with empathic behavior attributes as evidenced in cartoons, such as an “angry car” or a “cheerful flower.”

## 4 General Intelligence

Intelligence, as a computational process, is a continuum rather than an end point. When bound to the real world, it is the ability to so deeply understand the nature of reality, of which it can provide increasingly accurate predictive power. This includes the ability to run predictions in reverse to build causal or historic relationship chains.

Humans use this predictive power as a means to interact purposefully with their environment – to aid survival and promote adherence to genetically

prescribed or socially engineered values and goals. But for Intelligence to exist at all, there are certain environmental pre-cursors:

1. A physical medium upon which it can bind the predictions: *reality*.
2. A representative medium in which it can model the predictions: *virtual*
3. A motive force: *energy*

And for intelligence to speculate on our reality it needs a means to:

1. Access that reality: *exposure*
2. Perceive that reality: *modalities*
3. Decipher that reality: *instantiation machinery*
4. Model that reality: *modeling machinery*
5. Grade the simulations: *emotional machinery*
6. Classify and store data: *memory machinery*

A presumption is that due to quantum effects at the very small level and chaos effects at the very large – prediction, and thus intelligence, will remain illusory. Added uncertainty arises with other biologically constructed animated beings. Constrained by physical law, yet animated by reflex, genetically programmed instinct, or from their internal cognitive processes. How can such complexity ever be intelligently predicted? Yet we ourselves appear able, at least to some extent, to overcome all of these effects.

At the atomic level, it is rarely necessary to predict particle animation with certainty, because all significant effects occur in the aggregate, where statistical probability can reliably model behavior. Also, predictions can be constrained to avoid chaotic events (so rather than walk a tightrope to get from *A* to *B* you take the foot bridge). With biology, statistical prediction still works well on macro events, but is limited in the details. So although intelligent prediction does appear to have some constraints, there are still very large areas where it can be relied upon. Within the oceans of chaos there is much dry land upon which to build a rational intellect.

It is further presumed that computers are deterministic and humans non-deterministic. Therefore, given an initial set of conditions, a computer can only ever follow a predetermined course. Whereas a human, with “free will,” can follow his own. For all intents and purposes both can be considered non-deterministic, though statistically predictable. The study of human twins illustrates how the same largely deterministic genetic inheritance can be affected by real world chaotic forces. Like internal brain chemistry guiding emotions; sensory data flow, unique first person perspectives and the resulting memory structures; differentiated emotional responses etc. Add all these variables and more together and you have a combinatorial explosion. Genuine AI will similarly benefit from many of these same forces. Even blind random inputs could be easily added if found beneficial.

Intelligence is non-judgmental and the pursuit of knowledge morally neutral. But any action affecting other conscious entities creates moral hazard.

Morality arises from the exigencies of biological survival within a social framework and is dominated by genetic and social programming biases, as show in Table 1.

<i>Positive emotional grading</i>		<i>Negative emotional grading</i>	
mating	pleasure	physical damage	pain
predicted food	anticipation	predicted damage	fear
fulfillment	satiation	lack of food	danger
profitable action	pride	social exclusion	shame

**Table 1:** Positive and negative emotional grading.

The primary genetically derived grading process leads to the basic positive moral status of survival (existence), feeding and mating. Secondary genetic and socially trained schemas lead to the moral grading of simulations involving cultural concepts such as cooperation, altruism, group patriotism, treachery, over consumption, monogamy etc.

**4.1 Human Intelligence**

The human mind is a particularly difficult thing to understand, but it is the best example we have of intelligence with intentionality. The brain appears to achieve this through a massive structure of neural networks which are able, over time, to effectively interpret sensory data in order to understand and predict the perceived environment – more usually our external world. Thus, research into evolving hardware and synthetic neural networks would appear a worthwhile endeavor[5].

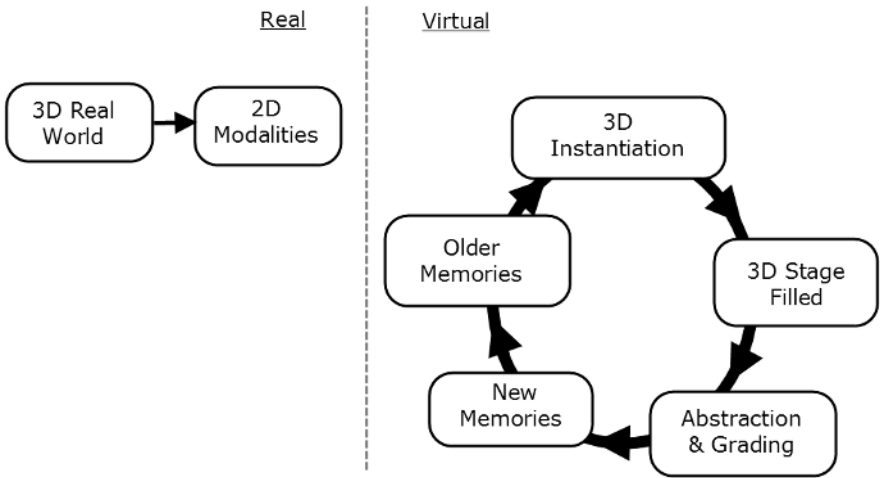
Biological intelligence evolved through natural selection. It developed inside a mobile mechanical body with rich sense modalities and programmed survival instincts to grade the information flow. It is protected during a nurture phase where a subconscious computational process can learn to extract meaning from the sensory modality flow and bind the internal simulation architecture to the physics and object behaviors of the real world. This subconscious simulation builds a personal feeling of familiarity with the outside world. Otherwise each moment would forever seem strange and new as if being met for the very first time. Intelligence then develops gradually through continued interactions with the environment being compared to script predictions. The level of intelligence reached is based on both the initial biological construction and from subsequent interactions with the environment – nature and nurture.

A human infant, exposed to the outside world, gradually learns to interpret the 2D visual images into 3D virtual objects. This process is significantly aided through muscular feedback, mobility and the other sensory modalities, together with genetically inspired dedicated machinery for this purpose. The



3D objects, once extracted, exist not in isolation, but within their virtual environments and as animated scripts. These will gradually build up structured and cross linked historic memory records, forming an increasingly accurate world model. Objects have textures and behaviors (animated shape morphs and/or motion scripts), together with empathic emotional hues.

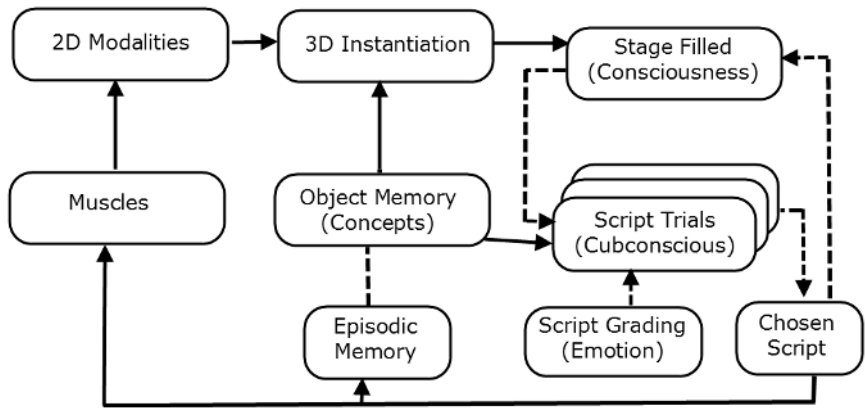
Intelligence, as such, begins to really kick in when a sufficiently detailed world model has formed and enough 3D object behaviors accumulated. The maturing mind can then focus more on the content than the 3D instantiation (sometimes referred to as binding – translating modality inputs to percepts [3]). An inner virtual world will come to map the external world, and the ability to notice and interpret anomalies between the two will increase, as will the ability to predict events from precedents.



**Fig. 3:** The human mind: learning

When a child awakes, her mind will resume the virtual model of her room and her waking eyes will orientate, texturize and track that model. She will experience a feeling of familiarity as her sensory flow matches the virtual model she holds in memory. As she moves, so the perspective of the model will too. In fact, a series of subconscious 3D script predictions will have pre-empted her motion even before she gets started. It will partly be those predictions that lead to her intentionality of action. As her eyes scan the visual scene, detailed 2D image data will paint accuracy into, and reinforce the authenticity of her virtual world. It is in this way that she is conscious she is in a room, and feels competent to negotiate reality.

An unconscious process runs memorized script behaviors ahead of real “modality” time to generate as many predictive script estimates as time or



**Fig. 4:** The Human Mind: Free Will

satiation permit. The best case script can be used to form new learned memories or to animate physical action by aligning the virtual simulation to the modality inputs, linking the virtual body animation to motor control in the real body.

There are two priorities to human cognition. The first, mentioned above, is reactive thought, which involves negotiating real world environments, objects and people in real time. Here, the subconscious simulators may operate at maximum speed and concomitant reduction in accuracy. The simulations are generally bound to the real world through the modalities. The second, reflective thought, involves thinking by processing memory records, with limited or no external sensory perception, but with far greater depth and precision.

The content of reflective thought is based on simulations built from learned objects and behaviors acting on historic episodic scripts. Virtual in nature, these simulations will be time discontinuous for easier layering, merging and comparison – in order to discover relationships and metaphor. Cost-benefit analysis and risk assessment are extensively used to guide, grade and judge this script discovery process. They are synonymous to human emotions. Compared to reactive cognition, these simulations are not driven by exigencies from the outside world.

Other factors influencing this process are genetically derived biases carrying heavy emotional content (like fear of snakes, desire for the opposite sex, etc.). Such imprints must surely have been written into memory by the genes and must also exist in the very same language as whatever instantiations the modalities cause. The fact that genetically derived instinctive triggers can be recognized and emotionally graded and responded to from untrained input, categorically implies a priori knowledge of that percept and of a common language for its recognition. For 2D visual input, where 2D images can so

easily disguise content, 3D instantiation is by far the most credible link. Thus genetically derived instinctive imprints must have a direct correlation to our modalities – particularly vision, with the most likely common language being 3D instantiation.

The process of human learning is thus predicated on exposure to the real world through the sense modalities. The mind gradually builds historic records of familiar environments, 3D objects and features, with increasing fidelity. Adding more objects and details as time goes by. The power of time shifting, time discontinuity and layering/blending in virtual simulations leads to rational prediction and intelligent cognition when aligned to goals. A side effect of this process is the seductive lure of unbinding the virtual models from real world physics and historically learned behaviors, and promoting instead, an internal world of fantasy. This process is further encouraged by the effects of biological feedback in the form of emotion. Human cognition is highly tuned to emotional cues within content, and uses them as short cuts to cognitive effort. Unbound simulations can thus be used to amplify emotion in a simulation. Presumably, attending to material survival have kept such processes in check.

## 5 3D Simulation and Language

Man successfully learned to express and then codify knowledge by symbolic notation. It could then be externalized and preserved through generations as a common resource to be shared and built upon. But language has a subsidiary relationship to reality. If you take a 3D cube to represent all space time, what lies inside that cube is reality. But the virtual extends both in and outside of that cube. Language, too, straddles both worlds like floating braids, weaving in and out of reality, embracing fairytales and hard science alike. As such, it may not be so reliable a foundation upon which to base AI.

Even when language tries to constrain itself to describe real world objects or behaviors, it is not always so easy to test whether the braid is really bound by reality. It is often ambiguous. There are other problems:

1. Language can break physical law and logic with impunity.
2. Language is interpreted differently by each conscious entity.
3. Language does not fully circumscribe or instantiate an event.
4. Language is time serial in nature, consciousness is parallel.

Nowadays, visual media too can subvert the authenticity of our simulations by invoking fake imagery, the way language has always been able to do. In any event, the best way to test the truth of any language is to bind it to reality through physical experiment. But can virtual 3D simulations be bound by real world physics to keep them in the “reality cube”? It is often said that a picture is worth a thousand words. Maybe a 3D model is worth a thousand

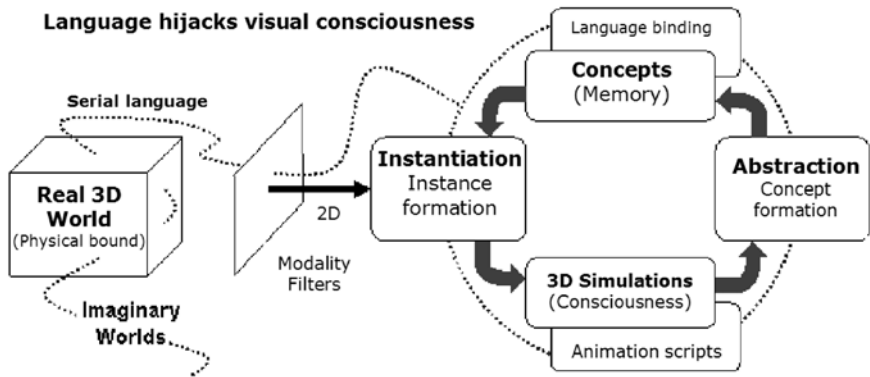


Fig. 5: Language hijacks visual consciousness

pictures. At one million words per model, 3D simulations might build a better basis for AI.

The syntactic structure of language often implies precision and completeness, but only by translating language into the form of a simulation can any ambiguity, or breaches of physical law or logic be discovered. Language processing is notorious for its blindness to common sense, which become glaringly obvious the moment a simulation is run.

Language is also used to impart emotion, either through the delivery, emphasis, or choice of words. Emotional analysis will form a key part of any 3D simulation. The best way to discover the meaning of a given piece of language, is to run a simulation around it, test its adherence to reality and grade its emotional procession in relation to values or goals.

Any language must exist within the context of a simulated world model; this will help determine boundaries. Nouns are drawn from object and environment memories, verbs from the spatial and temporal “behavior” memories. Thus language can build simulation scripts – or allegories. Script validity may be discovered by testing the simulation for violations of logic etc. But for much of language, real meaning is hidden within inference or metaphor (Therefore, the substitution of disparate objects but with matching behavior patterns or vice versa). These metaphorical script trials can similarly be interpreted based on context, logic and graded through emotional cost-benefit analyses.

But how can a 3D simulation interpret concepts such as math, statistics or software? The temptation, of course, is to not bother interpreting to a simulation at all, because binary computational algorithms are already naturally suited to these domains. But that would be a mistake. An algorithm can solve a calculation millions of times faster and more accurately, but there will be no concomitant understanding of what happened. It is only when the numbers, graphs, or code are modeled, and analyzed in simulation with reference to

historic representations of reality, that meaning and understanding can occur. The simulators within the human brain are not well suited to modeling mathematical or repetitive iterative processes due to rapid informational decay and weak cognitive focus. Thus, we tend to use memorized shortcuts to help maintain momentum.

If the goal is to test for possible relationships from a set of numbers, they might enter a simulator as columns of varying height. The simulator could draw on its historic memories of common number series. Such as shoe sizes; imperial weights; removable storage media sizes; French coin denominations. Or from calculated series, like prime numbers or various other mathematical series. It is thus by the sorting, layering, scaling, merging and comparing of these graphic patterns that relationships or meaning can be found within the numbers behind them, and that subsequent meaning bound to existing memories and thus representations of the real world. The traditional brittleness of computers dealing with numbers and language in the context of AI, stems from the difficulty of blending the data into wider knowledge integrations, particularly through metaphor, where the substitution of disparate knowledge areas extends the reach and depth of understanding.

The proper place for math and language notation is as a mechanism for the coding and serialization of information, so it can be efficiently stored, transferred or retrieved from constrained informational channels. Within AI the best way to process such shorthand notation is to translate back to the 3D domain where it can be bound to the constraints of either real world physics, or at the very least a notional 3D space and have behaviors referenced to historic precedents.

Language gives the illusion of delivering more content than it really does, and it is this very imprecision and ambiguity that gives it such flexibility for social communication. But the devil is in the details and it is those missing details where the real action lies. Ayn Rand states a single word can imply a thousand instances, but an implication is not the same as the thing. To identify a chair or a molecule as a class might be efficient, but it is not precise until it is instantiated as a specific chair or molecule at a specific location. 3D simulation is the real fire in the mind, but to be fair, by adding symbolic language, it's like throwing gasoline on that fire – by adding turbo charged addressing and scripting to our 3D memory records. Language thus leverages our simulators hard, as if on steroids, igniting the firestorm of our wider human culture.

Language is used extensively in human cognition to economically build up simulations and to express their script procession in a serial communicable form. Further, It is almost certainly the coding mechanism used to classify objects for subsequent retrieval from memory and possibly even a predominant part of our episodic scripts. But serial language is simply insufficient means in dealing fully with the real challenges of AI; though it is certainly an essential element. Language is to the mind as a scene scripting language is to animation software. It describes and directs the animation flow.

Some examples: Fred was in the living room practicing his putting. What would happen if he practiced his driving? How could AI based on language alone understand this type of common sense content? Or even more importantly, solve the following tasks: design a mechanical human arm, a virus that can target cancer cells, or a three dimensional memory chip. Simulate a 256 bit RISC processor core?

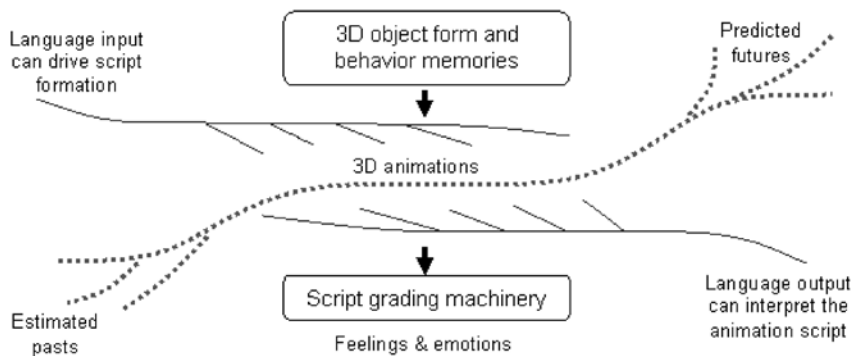


Fig. 6: Language and scripts.

## 6 Epistemology

To ensure survival, human consciousness has been dominated by guiding interactions in the real world to procure resources and mating opportunities. Mortality is the primary existential condition and leads to many biological prejudices, with physical and cognitive power rising gradually from childhood, peaking in adult life and then falling off again in old age. Humans have fairly rigid cognitive machinery which constrains their capacity for intelligence and they cannot easily alter their genetically pre-programmed emotion analyzer to favor mental effort over visceral pleasures.

If you take the image from an eye or camera, or you listen to speech, you create parallel information wave-fronts. These are meaningless without reference to a common reality – which for humans is existence. So how can it be that blind or deaf people can think? It is because they have constructed the same 3D world model from the remaining modalities; particularly touch and movement. For instance, a sighted person cannot see clear glass, yet he understands the concept of glass. If he is told a sheet of perfect invisible glass separates a room, though he may not be able to see it, he can conceptualize

its existence quite clearly and act accordingly. For a deaf person, the language tags directing simulations would be purely visual rather than audible in nature.

A predominant feature of human existence is physical animation. These abilities are likely heavily supported by specific trained neural networks rather than any intimate conscious control. Motion requires fast cybernetic feedback to handle momentum. To offload this work onto sub processes would leave consciousness more time to deal with higher goals, like a plane requires limited input to guide flight. So the human body can animate largely free of direct conscious control.

The human organism is but one half of the coin, the other is his environment. Moreover, Intelligence is but one aspect of a complex set of processes involved in biological existence. Any artificial intelligence in the true likeness of man will surely be quite an anomaly. For these variables are the source of all our biological motivations for survival and cognitive attention. The human organism uses exposure to the environment over time to facilitate the development of a realistic world model. Success in this endeavor aids survival. But human cognitive focus is largely dominated by biological imperatives. This drives much of our intentionality and subsequent physical activity, creating the curious human civilization in which we live.

If we come to the question of our objective in building machine intelligence, we might ask – is it to replicate as closely as possible the human condition? Or will other goals be better aligned to our technology and desires? The Human means to knowledge occurs over many decades through full-on reality immersion with subsequent repetitive trial and error learning cycles. Such methods, even if practical, might be too slow a strategy to developing useful AI. One might presume that AI will have been achieved once the Turing test is successfully passed. However true this may be, it might not actually be the wisest of strategy for current research. The reason being, the test presumes anthropomorphic qualities in a machine are necessarily indicative of the most advanced state of consciousness to be sought; where concepts such as social inclusion and biological proclivities are pre-eminent. To put it bluntly, knowing how to eat a banana or understand a joke, admirable though they may be, might not be quite as important as an ability to accurately model a specific protein fold, and predict resulting regions of subtle chemical reactivity!

## 7 Instantiation: the Heart of Consciousness

Possibly the greatest software challenge for AI will be the instantiation engine. It must reverse a 2D bitmap render of vision (or indeed from any modality input) to recognize the environment and objects from internal memory correlates (concepts) to recreate the virtual 3D scene. There are really just a few common classes of environment sets – countryside, office, kitchen, work

bench, shop, theatre, plane etc. If any environment match can be found, a fully instantiated scene framework will be ready to go, leaving only image scale, detail and perspective to be resolved.

A few pound lump of clay can instantiate a greater variety of forms than the entire number of atoms in the universe. But only a tiny subset of those forms will have any meaning attached and be associated with any behaviors – cat, fridge, airplane etc. The human mind is able to, with only a few pounds of meat, instantiate form and behavior from novel 2D vision scenes at the rate of about one object per second. Considering how many 3D pattern matches must be made against our library of known objects, this is quite an achievement. In most circumstances, significant mystery can remain within a scene (bitmap areas without instantiation), so long as the major items are decoded out; such as environments, significant life forms or emotionally charged objects.

Possibly, with unlimited time and processing power, artificial instantiation could be achieved through 3D scene estimates, rendered down to 2D and then compared with the bitmap input. Corrective feedback cycles could iteratively discover the light sources (from radiosity and shadow effects) and camera perspective (from room edge key points or with lock-in provided from a single object discovery). But it should be possible to design faster search algorithms than such brute force trials, perhaps by comparing pre-rendered trial object “icons” to the 2D scene. Or in reverse, by extracting edge patterns from the 2D image, normalizing scale and tossing those into a search path through memory to catch shape and/or surface pattern matches.

The challenge is to design a 3D object description language that can be interrogated rapidly and one based on fuzzy search criteria. You cannot use a polling search metaphor against a million images, each of a thousand orientations; you have to use an “interrupt” or “vector” search metaphor. Human vision is based on the identification of features rather than exact form, thus a violin twisted around a pole can still be recognized; or a clock printed on a crumpled table cloth. The challenges of high speed instantiation make the decisions where to focus attention; on the motion of a cat or to follow the eyes of a human, seem almost trivial by comparison.

Just as a human is built upon autonomous biological layers, cognition has its own autonomous layers. For instance instantiation, morphing and tweening (the construction of in-between time frames during simulation). When we script a human actor entering a room, the motion tweens do not need to be consciously re-calculated; their construction is either automatically generated or already stored in memory as an animated motion tween. Only the environment, context and emotional attitude need to be scripted in order to direct simulations.

Rendering is the translation of 3D scenes to 2D bitmaps. Instantiation is the reverse, the creation of 3D scenes from 2D bitmaps. Using a neuron array metaphor, where a projected image triggers firing along an axon. If those neurons each have say 256 axons (connections) propagating out, within that tangle there is spatially encoded all possible orientations and translations of



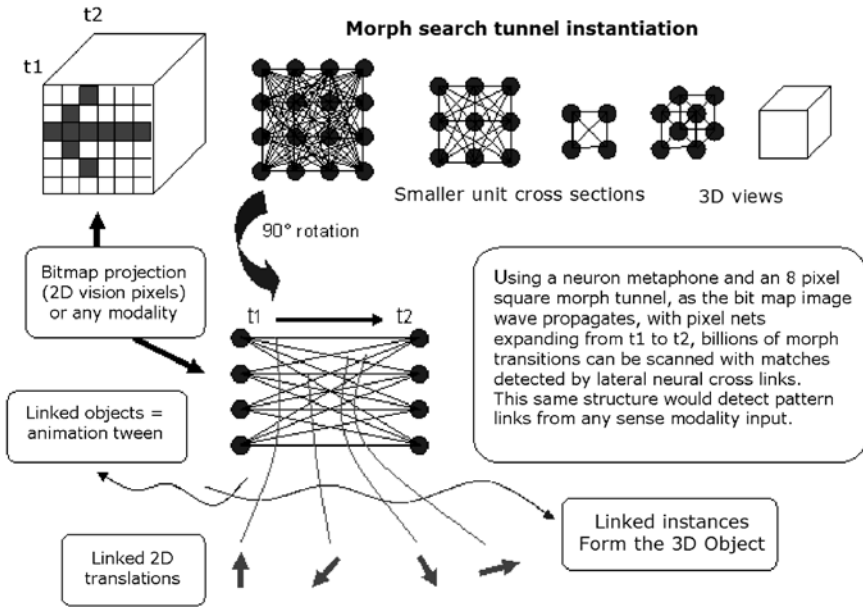


Fig. 7: Morph search tunnel instantiation

any 3D object. The decoding out of that data could be achieved from the propagating input wave function through time. For example, if each of the elements on two opposing faces are connected to every element on the opposing face. Therefore each input pixel has 256 vectors spreading out. If it took one hour for the signal of a firing neuron to travel along the axons between the surfaces, and you divided that time period up into small enough units, at any instant in time, a set of those vectors from the expanding pixel wave fronts will be optimally aligned to a specific translation of the projected object. Were those vectors known (trained), and linked together, the full 3D translation could in theory be described by those lateral connection sets. If those connection channels were two way, the objects could either be instantiated (identified) from input modality patterns, or in reverse, be used to trigger the same visual imagery (memorized experience) but directly from the linked network patterns, themselves connected to similar and associated modality patterns of visual and oral language tags, or even taste, smell and touch attributes.

Modality flows, whether from sight, sound, touch, taste or smell manifest in the brain as parallel analog data channels of specific and appropriate frequency, phase and dynamic (amplitude) ranges. The same principle of instantiation applies equally to all these analog sensory data sets; with receiving neural arrays, optimally tuned to the character of each input class. For example, the sound of a word or event, as with vision, will enter the neural

array as a parallel 2 dimensional analog data wave-front of frequency and phase channels or “aural pixels,” extending into the neural array as a third dimension through time. Cross connections linking spatial patterns will again identify those with the closest correlation to existing memory traces. In this way, as for vision, if only part of a word is heard, in any tone or accent, or even masked by other sounds, there will be sufficient signature correlation to make reasonable probabilistic guesses for subsequent wider context simulation trials and grading. These data signatures, being now instantiated, are thus linked to the universal environment map of objects and environments. Otherwise, the inputs would merely remain unidentified sounds bearing only fleeting similarities to known aural traces.

Instantiation processing from sensory modalities is automatic and unconscious; there is little mental effort involved, and further, not only are the 3D objects instantiated, but also are any associated animation tweens (object behaviors). Just as bitmaps link to 3D objects, so those 3D objects link to form animated behaviors, either as internal memorized tweens or newly constructed object motion or morph tweens.

Take a mouse object at time  $t_1$  and a teaspoon at  $t_2$ , place them in the same spatial location and connect their surfaces together with orthogonal vector lines. Divide those lines into equal “time” segments and render a perspective to create frames for the movie script. This process is known as “morph tweening,” and will represent one of the core visual translation tools necessary for AI to both interpret modality flow and to create new and novel content. During any visual thought process, creating smooth in-between renders between distant or disparate objects in time and/or space will be crucial.

Even apart from AI, the commercial spin-offs from an instantiation engine will be enormous. To start with, consider the possible re-animation of all historic language documents and visual 2D media, to create a cornucopia of rich, new, flexible animatable content.

## 8 In a Nutshell

Biological organisms interact with reality to survive. Sensory and motor systems evolved and so eventually did a computation engine in between. In humans, these computations create a simulated environment through exposure to the real world, converting existential matter into virtual representations. (Therefore external reality made from atoms, have digital/informational correlates). An internal map and a repertoire of environments, objects and behaviors develop through a process of exposure, perception, instantiation and memory formation.

All sense modalities converge to memory space as pure information, which is the very loci of consciousness. This information is instantiated, simulated and then graded to guide behavior and generate data we experience as feelings. Subconscious processes unlock the time and reality constraints enforced

by the external world via the modalities and allow object behaviors to flow freely, constrained only by “prior art” and processing resources. These script variations subsequently feed simulations back into the area of consciousness – like Aristotle’s “Cartesian Theatre.” The first person conscious observer occurs at the information interface between the rendered virtual simulations and those “rendered” by the modalities of the outside world. The subconscious processes place consciousness, and thus our perception of reality, into a known place, and a time event horizon of the present placed between predicted futures and a remembered past[2].

The subconscious uses cues from the external world, or recent episodic memory scripts, to seed script diversity as simulations are intimately dissected and transposed. Virtual time travel and time discontinuities are aggressively used to construct metaphor, meaning and relevance out of the resulting script compositions. This “meaning” is discovered using genetically and socially programmed emotional filters which grade the scripts according to factors such as survival, security and cost benefit analyses, prioritizing social and resource capital, such that for every act, a human will know to the best of his cognitive ability, what is most in his interests at that time. It is the breadth of this process of subconscious wide scope accounting with ever increasing circles of virtual time expansion within simulation scripts, coupled with “emotional” cost benefit analyses that defines the depth of a man’s intellect[14].

Subconscious processing uses the short cuts of context and precedent to speed script discovery, and when the rules of simulation are grounded in history and reality, the subject can use the simulations as the basis of learning and for future plans in dealing with real world situations, without then needing to physically act them out. Because the simulations are unbound by time, they can often beat external reality and thus anticipate real world events.

Once the optimum simulation script has emerged, real world human animation can be guided through one-step-ahead simulation linked to modality feedback. Trained neural cybernetic scripts would greatly enhance the speed, accuracy and grace of these animations, such that the individual control of limbs and body momentum are left to subsidiary pre-trained largely automatic processes. Human action subsequently follows with intentionality declared to be free will.

With subconscious activity constantly trawling memory records and modality stimuli, free will is simply the ability, at any given time, to flip life’s animated momentum to be aligned with alternative virtual script offerings, even a destructive one if proof of courage, or free will, are defined as higher goals. (Which themselves are guided by the socially or genetically programmed emotions). During sleep, or quiet meditation, the process is driven by memory records alone, and away from the roar of sense modality flows, the subconscious script simulations can leak like ghosts into full consciousness, leading to imagination, creativity, planning and ultimately to self consciousness.

Humans have the ability to compute and render into consciousness the scene from any movie, placing say Donald Duck, or their grand mother, in the

leading role. This ability comes from an internal scripting language that has access to powerful modeling and animation functions. We do not need to consciously solve the mathematics for the inverse kinetics of mechanical motion; or of momentum or gravity. We create the animated collage from prior learned 3D models, environments and either pre-rendered animation sequences or on the fly with motion and shape morph tweening. After morphing and blending the objects and scenes from prior learned behaviors, we can then render them to an observer perspective into consciousness, mentally skipping over much detail – the way we’re deceived by a skilled magician – believing all the while we’ve missed nothing. But the mind has an advantage the eyes do not; it can censor and lie at will. The senses and internal memory contradictions try to keep the mind honest.

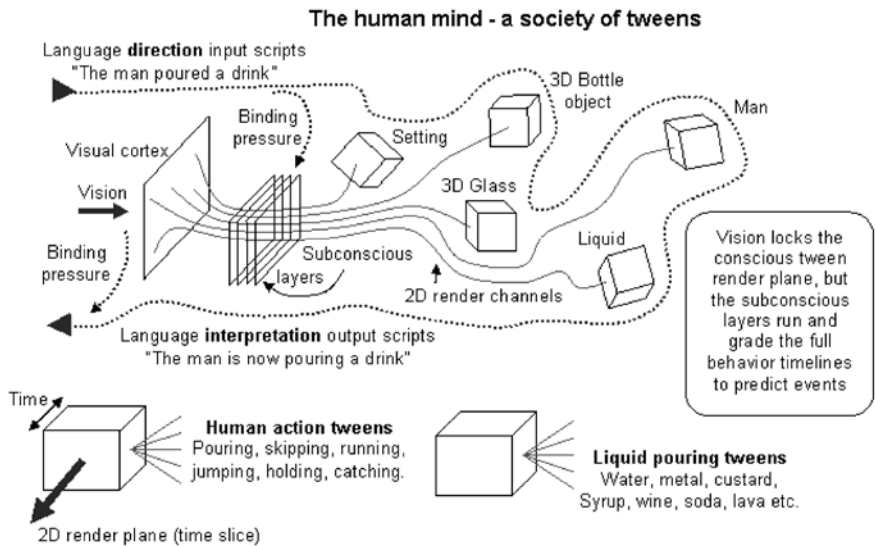


Fig. 8: The Human mind: a society of tweens

To summarize the postulates:

1. Our reality exists in 3 dimensions over time.
2. Units of matter can be represented by units of information.
3. The aggregates of atoms within objects and environments of reality can be converted (through modalities) to stored information within memory.
4. The identity and behavior of objects can be instantiated from those memory records (through computation) and then stored as informational representations.

5. These representations can subsequently be recalled and manipulated to simulate the behavior of their real world correlates as 3D animation tweens.
6. A software process can judge and emotionally grade the intrinsic value of these simulations to guide and optimize script formation.
7. Step-ahead animation and pre-trained cybernetics can be used to align physical action to the script.
8. With sufficient computation, memory resources and exposure to reality, this process can become a self reinforcing seed process – leading to advancing intelligence.

Intelligence, consciousness and feeling are virtual informational processes based around 3D simulated environments which are bound to reality by pre formed genetic road maps and experientially over time through sensory modalities and mobility. Consciousness arises from the supervision of these simulations linked to feeling – which is the computational process of grading those simulations. Intelligence is the ability to expand the time horizon (time dilation) to discover causes and make predictions. All these processes can be achieved artificially and will lead to AI. The controversial aspects of this paper are that:

1. 3D simulated environments are the basis of cognition.
2. Human language is a subsidiary process.
3. Any language which contains meaning can be reduced to a 3D simulation.
4. Human feelings are illusory; they are self referential computation processes.
5. The nexus of consciousness is the boundary between the modalities and the feedback from simulated environments created by subconscious computations.

Thus, the proposal is – that matter can be represented by information; that objects and environments can be instantiated from perceptions of reality; that they can subsequently be simulated and that information can be stored; that these simulations can be graded based on their progression in time; that simulations can run faster than reality; that through the superposition of memorized behaviors, simulations can represent potential versions of reality; that these superposition's can be “emotionally” graded and evolve toward an optimal prediction - using historic precedent; that chaotic discontinuities can be avoided; that these simulations, being able to predict reality, can be used to align physical action to those simulations; that this computational process can beat the procession of time in reality. This then, is the process that leads to consciousness, intelligence and intentionality.

## 9 Real-World AI

There are several factors distinguishing real AI from expert systems: the breadth and scope of the knowledge base; the ability to ask the questions; to identify missing knowledge; to judge the relevance of results; to apply context or predict effects over time, etc. These extra features require a simulated environment like our own and a world model of equal breadth.

On the evidence that immobile, deaf children can still develop high intelligence, presumably from visual stimuli, we might also expect a similarly restricted machine analog to have an equal chance of success. In order to conceptualize a credible AI architecture from vision, imagine following our current technological trends for a few years to where the following levels are reached:

- *Cameras*: 36 bit color depth, 6000 x 3000 resolution, 60 frames per second
- *Exposure*: 1 gigabit broadband internet connection attached to browser clients
- *Memories*: 10 terabyte, non volatile, shared, direct addressable, 10nS access time
- *Processors*: 10 teraflops, serial (in autonomous clusters)
- *Instantiation*: accurate 2D to 3D translation software
- *3D modelers*: any shape, scale, texture, orientation, behavior, etc
- *3D simulators*: supporting physics, collisions, chaos, time shifting, etc
- *2D renderers*: supporting shaders, shadows, radiosity, fog, etc
- *Animation scripting language*: object insertions, orientations, behaviors, morphing, tweening, layer and time management
- *Database*: of records, concepts, objects, environments, and episodic scripts
- *Language to animation script translator*
- *Animation script to language translator*
- *Script grader*: cost benefit analysis, entropy, normal, harm, irreversibility, danger, opportunity, 3rd person script empathy and 1st person emotion analysis

The major software challenges:

1. 3D instantiation from 2D sense modalities
2. Construction and maintenance of a universe environment map
3. Construction and maintenance of object records and behaviors
4. Powerful, multilayer 3D simulation engine
5. Blending/morphing of environments, objects, properties and behaviors
6. Grading of simulations to guide script progression

The most appealing hardware structure would be a network cluster of maybe 10 or more powerful self-contained computers but with shared memory resources, each dealing independently with separate aspects of AI. The continuing massive worldwide investment in operating systems and application

software can be leveraged to become tools, blurring the boundaries between modalities and consciousness. Products such as Windows, Linux, commercial 3D modelers, OCR and speech recognition software. But the boundary between man and machine is already getting very blurred with ubiquitous cell phones providing an almost telepathic modality; not to mention speech control of computers, graphical interfaces, instant messaging and email, etc. To some extent, most people already spend most of their lives in virtual reality; they just don't recognize novels, radio, TV, computer games or software as being virtual environments.

A human without recourse to modality extensions such as an auto, cell phone, internet, computers, fedex account, credit card, 3D printers, etc. would be a greatly diminished soul, and the same goes for AI. The most important cognitive skill will not be to walk or even talk, but to manage multiple computer graphical user interfaces.

But how could these advanced technologies begin to be organized to create intelligence? First, the camera would project its bitmap data to a memory map, which would be routinely processed by the instantiation engine to identify known objects and environments from memory records. A subsequent 3D simulated environment would be constructed in memory to match the visual scene and simultaneously rendered back down to a 2D bit map memory space at the same first person perspective as the camera input – much like a 3D animation film is rendered to the 2D screen image. If the camera data flow were interrupted, the rendered 2D data from simulation would be an accurate mirror copy of the real scene.

There are three dynamic events that can now occur within the visual field. An object can change, the perspective can change, or the whole scene can change. For scene changes, the previously described process of instantiation and discovery would be repeated. For perspective changes, motion vectors (as used in video compression) would be calculated to keep track of scene perspective. For object animation, the software process would recognize localized anomalies between the simulated projection and the vision projection. Then using normal instantiation techniques focused on the anomaly, the object in simulation would be oriented until the 2D rendered projection and the input vision were once again in alignment.

The memory management software would need to maintain an associative database linking all objects, environments, behaviors and scripts. Together with growing lists of knowledge about these models, such as: language tags, price, legal status, disposal, source, manufacture, flammability, safety, uses, weight, dangers, precautions, social status, classes, trends, history, composition, aging properties, storage, popularity, component parts, assembly, regulatory compliance, standards, size, growth time, environmental impact etc. Object behaviors would be characterized and stored based on:

1. Motion vectors over time. A feather would tumble through air differently than a balloon floats or an insect darts – stored as positional and temporal data sets.
2. Shape variation or morphing over time – butterfly, bouncing ball, coiled spring etc.
3. Reaction to stimuli (touch, drop, cut, etc.)

The overall environment map would need to hold concepts ranging from the universe, through planets, countries, cities, neighborhoods, homes and factories, to materials, chemicals, molecules and atoms. At any point in the simulation, a relationship would exist to this universal map. Which specific country, town and room? Or if generic, it would still need a generic history with the potential to be “fixed” by subsequent facts. The depth and accuracy of this virtual world will largely determine the bounds and precision of thought for the artificial intelligence.

All objects blend together in an overall environment map, which fits within a wider contextual world map. Physics rules (gravity, hardness, weight, momentum, heat, speed of light, etc.) guide behavior and interactions between objects. (Cloth against solids, light through glass etc.) Much of this is already well advanced in commercial 3D software packages. The overall resolution and speed is dependent simply on processing power and memory resources. The software must subsequently recognize any bitmap changes as object behavior animation or changes to perspective and re-calculate to keep the simulation bound to the vision input.

The input video stream drives the construction of the virtual scripts. If novel, those scripts might be the basis of new memory formation. Inconsistencies would be challenged based on the source credibility or physical law, with certain knowledge discovery causing rippling adjustments throughout memory. Logical inconsistencies and vagueness might be highlighted to trigger some human supervisory training to help bootstrap the process. The addition of a language translator to convert words to simulation scripts will greatly speed learning, since most human knowledge and communication channels exist in the form of serial language streams. The language parser would construct scenes from any objects alluded to in the text, with action scripts proceeding from memory precedent and/or from the language verbs, syntax or emphasis.

Any proto-intelligence would begin as basic memory formation and correction processes, but the main advances will arise when running the subconscious simulation machinery separately from the vision input. The content of those simulations could be guided either by recent episodic memory scripts, prior behaviors or simulator physics, and graded by “genetic obsessions,” such as the “need” to understand.

The process of discovery might involve the searching of any language scripts associated with the problem, with their subsequent conversion to animation. Metaphor will be examined through object or scene substitutions within the trial scripts. Script diversity built from breaking up object sets and



re-ordering time through dislocations. But how exactly are all these script trials to be graded? This is the most difficult part of the process to explain with any clarity. There are several grading concepts like testing against law, mores or relevance to global goals. Further grading concepts might be: normal object condition; reduction of scene entropy; novelty detection; consequences to the wider time frame or applicability to other environments. But the most likely method will revolve around either quick-and-dirty pre-programmed emotional prejudices or, if more time is available, growing circles of cost benefit analysis expanding in time and in environment space, as the potential effects of the sample scripts ripple outward. These wider scope integrations will ultimately be graded against predefined “genetic” schema. Such as profit, shame, humor, social capital, etc. A final script must be found that predicts the highest probability of benefit and the lowest possibility of costs.

Another strategy for knowledge discovery would be the joining of means with ends to build a script timeline from the missing links in between. Once in the 3D domain, tweening can be used to bridge gaps, with the new tweened content tested against simulator reality constraints such as gravity, physical form and behavior, social mores and rules etc. Or perhaps more like a jig-saw puzzle, only with the pieces made up of memory records of objects and their behaviors or triggered from external search results. Finally, the expansion of complex objects to simpler sub units. Or the reverse, the assembly of complex from the simple, would further aid knowledge discovery.

So at this point we have a simulated environment held in “conscious” memory tracking the live video feed (and/or receiving script revisions from a subconscious process). We have a subconscious simulator building and expanding upon those conscious scripts using prior behaviors, with particular interest in novelty. We have script expansion though behavior extrapolation (e.g. a vase being nudged toward the edge of a counter will be predicted to fall and shatter). We have scripts graded through cost benefit analysis (a broken vase creates a loss of value and a mess – entropy). Next, we need a method for allocating time and resources to maximum effect, to direct focus and attention, and an ability to interact with external knowledge bases. Finally we need a satiation response to help allocate computational resources and escape dead ends.

The ability to search the external world for solutions would require language formation from the simulated scripts and an output method for gaining human attention or an ability to directly enter text searches into internet search engines. Due to speed, the first choice would likely be the internet, with human intervention being the least rational choice for guidance. Humans will be totally unable to keep up with the data velocity associated with AI thinking. An un-tethered AI would quickly overtake one kept anchored to the dead weight of human consciousness.

The primary source for learning material would be the translation of web based information to animated scripts within a global environment map to form the basis of knowledge integration. This would require 3D script con-

struction built from text, charts, sounds and images, in conjunction with the previously described video/image instantiation engine. The seed AI could begin making its own predictions and then testing those predictions through further internet searches to discover if it had found the correct causes, processes or results. Only when a concept exists without internal contradictions can it be said to be properly integrated and its authenticity secure. The cognitive advantages available to AI will include the following:

1. Persistence in simulation layers
2. Simulation accuracy and precision (e.g. for math and software)
3. Increased number of conscious objects
4. Increased size of simulation
5. Accurate simulation of physical law
6. Accurate “photographic” memory
7. Multiple parallel modality inputs (e.g. 100 simultaneous internet channels)
8. Extended modality inputs (e.g. data protocols, radio, IR, UV, ultrasonic etc.)
9. Automatic, high speed multi language translators
10. Greater conscious control of simulation progression and persistence
11. Scientific calculator, thesaurus, dictionary and encyclopedic resources
12. Patience, rationality and deep foresight

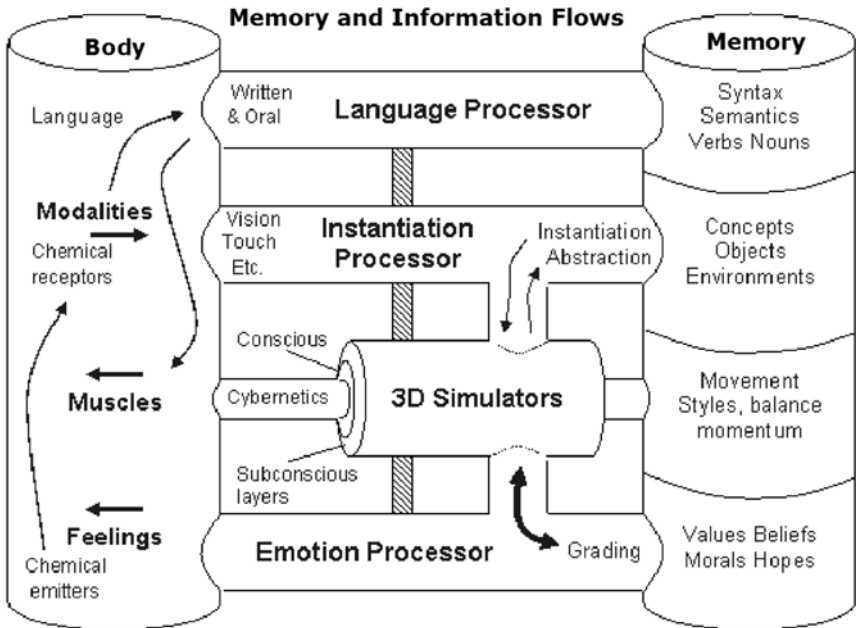
Finally, the human mind is unable to properly render its internal 3D content to anywhere near the clarity as when “painted” directly by the modalities. Thus, we are only really partially conscious; there is an enormous richness to existence and experience we are blind to. The mind is full of ghosts rather than realistic impressions and a ghost world is hard to fully embrace.

The only credible mechanism for self awareness to occur is as a computational process dealing with information representing and bound to reality. The self can then exist and be aware through a process of reflection (simulation) in a time controlled domain, where emotional grading (feeling) can percolate through time dilating script trials.

If you doubt that such processes of simulation and virtual time travel will really lead to intelligence, think of this analogy. You suddenly find yourself able to re-run time backwards and forwards in the real world as many times as you like, even making notes as you go along. After many such “simulations,” do you not think it likely that the action you finally take might be a little wiser?

## 9.1 Examples and Metaphors

Imagine a fish tank, 1000mm wide, deep, and high. The tank is filled with 1mm cubes. Inside each cube is a little scroll that says: air, gold, glass, skin, hair, cheese and such. These scrolls represent electronic memory locations that can be filled with information about real objects. Laws defining the relationships



**Fig. 9:** Memory and information flows.

between adjacent memory spaces are programmed to be in harmony with those in nature. Such as weight, object boundaries, momentum, light refraction, texture, behavior etc, very much like current 3D modeling and animation software.

This memory space can be thought of as a movie stage – a “Cartesian Theatre” or to use modern parlance – a virtual reality chamber. This virtual chamber can be filled with objects or environments and at any scale. Gas will disperse, liquids will spread to boundaries and solids will have weight and maintain form. Animated objects will flow according to their motion vectors and morphology – a perfect analog of real life; except matter is replaced by information. Like a 3D window or camera, this “box” can float across virtual landscapes and environments, to be filled at one moment with the great expanses of space and time and in the next, the most intimate molecular spaces of tiny living cells.

At the centre of this chamber is a virtual, animate human character. The contents of the theatre always render down to this 2D observer perspective, which is also the source point of filling the chamber from any modality inputs. This virtual space will form the contents of waking consciousness.

Further, shadow realms exist. Again like scrolls set beside the originals, except the contents of these scrolls are able to break free from the straight

jacket of modality flow. Here, the behavior of objects can follow trajectories learned from the past, together with substitutions and time discontinuities. These “subconscious” shadow realms can leak in and out of “consciousness” to also fill the “stage.”

## 9.2 Math and Software

“The challenge is not how to use computation to deal with the real world – it is how to use the real world to deal with computation.”

Within the human mind, a teapot can be blended with a donkey! The resulting simulation can be infected with the properties of china, flesh, ice cream or whatever. Inconsistencies fade out of the scene. This ability to mechanically draw disparate objects of class, form and scale together in the most structurally consistent and plausible way, is the basic stuff of our simulation machinery. The teapot handle may detach from the lower join to become a free flowing tail and the spout the donkey head. But through introspection, inconsistencies will come to light.

Whereas this ability does seem very powerful, it is at the same time very weak. No more than five or six attributes of a simulation can be held in focus at any one time. A simple long division would appear a somewhat trivial symbolic animation in comparison, but few are able to maintain sufficient control over the parts to achieve even this simple feat[4].

Math and software memories similarly exist either as animation scripts or simple learned pattern recognitions – such as multiplication tables. Like the images on a dice face, digits can have direct dot pattern equivalences for subsequent math animation (add, subtract etc.) Thus math can manifest as either image animations (e.g. joining/separating dot groups) or rely on memorized symbolic beliefs, such as  $12 \times 12$  has “an equivalence to” 144. Or for the binary truth tables – or should I say “belief tables” [10].

Just as there are behavior scripts for the way a ball bounces, a rabbit runs and a feather floats, so there are the more abstract behavior scripts of memory indexing, for-next loops and the like. Most math and software concepts would likely exist initially as animation scripts, but as our familiarity and confidence with them grows, short cuts are taken, jumping straight from beginning to end, and so over time they become simple memory beliefs without the intermediate animations. When we imagine a vase falling to the ground and breaking, we jump from the initial fall to the shattered remains more through belief, than the accurate simulation of each part of the event down to each individual chard.

## 9.3 Barcode Example

Finding the relationship between bar patterns and a decimal subscript will be used to illustrate knowledge discovery through animation (see Fig. 10).

The assumptions made are that the AI has access to image samples and that the memory beliefs of basic math and software primitives exist, as do the fundamental instantiation and grading abilities previously outlined.

Human cognition evolved to integrate 3D objects, environments and behaviors into a knowledge hierarchy – not 2D symbolic abstractions. It takes a great deal of effort and training for a human mind to so contort itself as to be able attempt these classes of problems. But with persistence, the help of external tools like pen, paper, calculator and computer, together with a little academic “coercion” – we are sometimes rewarded with results.

The method of discovery does not need to be infallible or super efficient, it just needs to have a statistical chance of success in finding connections and thus guide knowledge formation within the time allocated. The higher goal, as always, is to discover meaning through finding memory connections, joining means with ends and reducing mystery. In this case, the means is a barcode image, the ends – a decimal subscript number. A simulator deals primarily with object shapes and forms. Apart from drawing upon prior memorized beliefs in the form of animated scripts or static image relationships, there are fundamental “instructions” operating on those forms:

- Instantiation – identification
- Separation, scene explosion
- Re-scaling
- Perspective translation
- Geometric alignments
- Language attachments
- Object substitutions
- Joining – connecting

And grading machinery based on:

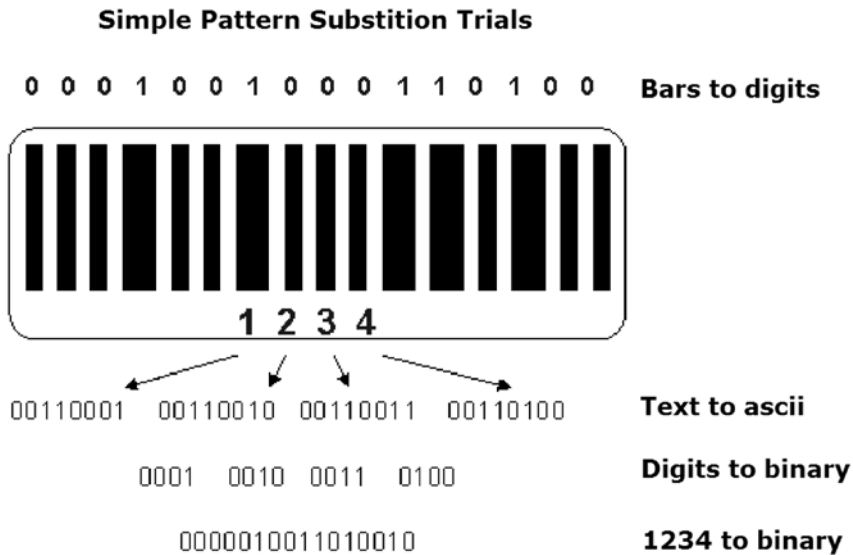
- Proportionality
- Similarity of scale, quantity and class
- Pattern matching
- Scene entropy
- Scene simplicity (Occam’s razor)
- Completeness/loose ends

These processes are fast, automatic and operate in layers through reversible animated pipelined scripts. Humans use pen and paper to “fix” parts of these flows to create order and permanence out of these somewhat chaotic streams. This helps construct an external framework to guide the process. AI will have the ability to do this internally by way of “persistent” simulation layers[7].

Each process is essentially dumb and automatic, but as a whole, and connected to sufficient source material and memory support, new connections can usually be found and integrated into memory. Dead end simulations will

fade away and if grading progress stalls, higher level processes will kick in – overall goal re-appraisal; the process will seek more real world data through the modalities or widen the internal associative memory search.

Applying instantiation to the global barcode image would yield six classes of abstract objects; two rectangle shapes and four numeric digit shapes. Language attachment to the object instances would connect as thick and thin bars and the four digits as a number.



**Fig. 10:** Simple Pattern Substitution Trials

At the “ends” part of the problem, we have a number 1234. Memory references will recall a belief that numbers have “an equivalence to” binary 1’s and 0’s. The first script trial might show an ascii equivalence yielding 8 bits per character. Thus an image of 1234 transforms to 32 digits. A second script layer might show each separate digit converted to a simple binary count. The third has the whole decimal number, 1234 represented by a binary count. Of the three scripts, simple pattern recognition would grade binary expansion as the closest match between means and ends. Further sample barcode image trials would confirm the link. Memory formations of the newly discovered script sequences would follow, including mutual pointers between the existing precursor knowledge records of decimal to binary equivalence etc. (Which incidentally, would reinforce the familiarity and trust in those prior beliefs.)

Now, when presented with similar barcode images, the scene will be recognized and will draw from memory links to the newly formed animation scripts

and an intimate familiarity with the scene will ensue due to these very same memory references, together with the emotional confidence that comes from recognition and understanding. The fundamental simulator operations used in this example of discovery were:

- *Scene instantiation*: to shape primitives
- *Language tagging*: from memory recognition of images/forms
- *Prior memory associations*: decimal to binary equivalence (as animation or belief)
- *Object substitutions*: bar shapes to “thick” / “thin” or to 1’s and 0’s
- *Image comparisons*: the bit patterns.

The process of decoding the barcode will not be understood in some isolated abstract way, but within the known framework of reality through intimate linkages with existing memory records; all being a part of a world knowledge and environment map. If a barcode is now presented with no number or vice versa, the simulation can play the script in forward or reverse to discover the missing parts through simulation to final substitution of bar patterns or decimal digits.

## 9.4 Software Design

The same *principles* involved in joining the donkey to the teapot would be used to create software. Each part of a flow chart script would be drawn into morphing relationships with software primitives. The challenge of software design is similar to the long division problem – only very small fragments can be held in simulation at one time. Software is the application of language rules to “direct the structured animated progression of data bits”. Just like normal language is used to script the animation of everyday objects. With software, pen and paper are often used to “fix” the framework using language tags in order to maintain animation persistence and build complexity. And just like for real world animation, where atoms are aggregated to forms and forms to behaviors – so bits are often aggregated to higher data abstractions. Like floating point numbers, arrays, memory containers or pointers with animated behaviors their own.

Using prior knowledge of indexed memory containers, a simple symbol substitution layer can form to match the data in our example, see Fig. 11. The initial “means” are still the bars substituted for 1’s and 0’s. The “ends”, the decimal digits with the previously learned simulated decode script in between. But this script is neither a formal flow chart nor software. It includes all sorts of miracles and beliefs to get from the bars to the numbers. (Bars turn into symbols, symbols to patterns. Patterns are compared to other patterns). There needs to be discovered, through trial and error, linking morph translations between the means and ends using software animation characters.

To start, the traditional **for-next** loop might be used as an initial script trial skeleton, upon which to attach elements of the known model. It doesn’t

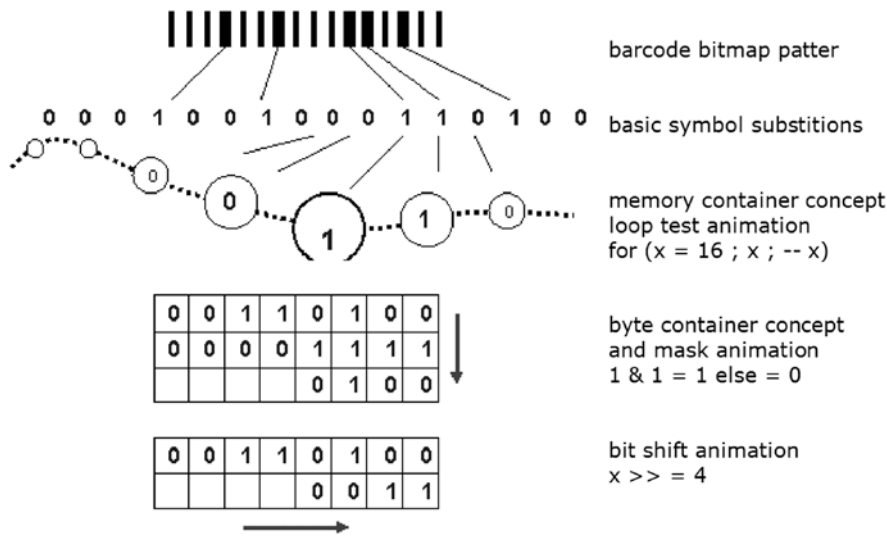


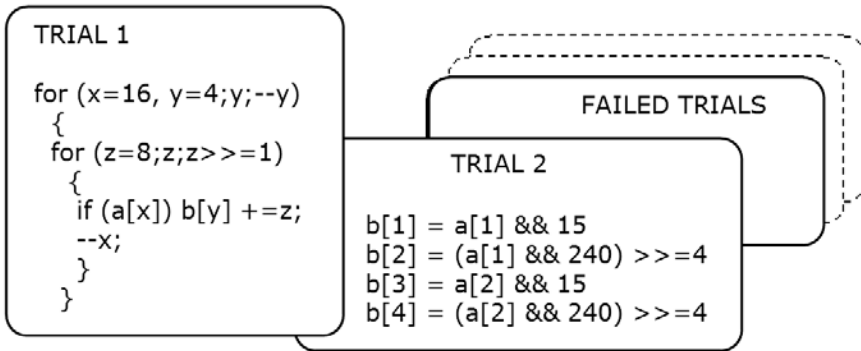
Fig. 11: Symbol substitution

much matter how the `for-next` animation is initially understood or remembered. Whether a cart wheel with spokes marking along a track, or a string of beads passing through some grading point. As experience has always shown digits to be the predominant substitution, they themselves will likely become the animation characters. And so for someone familiar with the C programming language, the expression `for(x=4;x;--x)` will invoke this abstract animation, but with roots firmly embedded to real world behaviors and thus connected in some way to all other knowledge. There are only two variables in the original simulation, 16 inputs and 4 results. So any loop substitutions are likely to be based around these two numbers, rather than say 42 or 365. The simulation iterations will then run by substituting the only elements possible to change, yielding:

- `for(x=16;x;--x)` do something with 'means'
- `for(y=4;y;--y)` do something with 'ends'

Morph attempts can now be made between these loop fragments and the original decode script. (As the donkey animation was merged with a teapot). Disparate parts of the two scripts will find tentative bindings, which will strengthen or weaken upon introspection – i.e. running the animations to discover anomalies. And using the framework from persistent language layers (as humans use pen and paper), to build and hold animation complexity. In this way, animated fragments will bind to the original decode script to generate code trials with subsequent C-language script formation, as shown in Fig. 12





**Fig. 12:** Script formation

Wider scope accounting would expand the extent of simulation to close loose ends by explicitly defining memory container sizes; initial conditions; test flexibility to handle longer barcodes or discover optimizations through further substitution experiments. But more importantly, for this code fragment to be understood in any context, it must be integrated into a wider causal framework of just how the barcode widths will become input data; what host device will be running this computation and how the results will be used. Thus the software code fragment will come to have a relationship with a material existence in the real world; as the motion of real electron charges on real atoms within the microcontroller of a real product. It is these linkages that are far more important to intelligent understanding than software – the awkward mental construction of abstract pattern animations and beliefs.

## 10 Conclusion

Everything that really matters in the world has form and behavior predictability. Sure, fancy mathematics can predict the exact arc of a theoretical cannon shell. But the universal language of object form and behavior (reality) is not English or math, it is 3D animation. More intriguing still, is that the 3D simulations within the brain are intimately “connected/grounded” to reality through our senses and muscles (the sensorimotor system), whereas data in electronic systems is almost entirely disconnected.

This theory challenges much conventional wisdom about human action, consciousness and artificial intelligence. In its simplest form, the theory presents these processes, all of them, by a single paradigm; 3D object computing. That is, all mental activity centers around the processing of virtual 3D objects.

The process involves the recognition of objects in reality from the sensory input flow (instantiation) and constructing an internal scene simulation based on those remembered precedents. These simulations exist beyond the fixed time reference of the outside world, because the memory precedents of the assembled objects contain a series of 3D “movie script like” scenarios for each of the objects – as learned from the past. As such, they can be used to make predictions of past and future action. Being virtual, these predictions can beat the time of reality, to allow a human to say, catch a ball in a future moment, and to know the balls origin from the past. Consciousness performs time dilation by building a simulation from memory precedents, which animate through virtual time. The simulated objects can be triggered into initial alignment with reality by the human senses. But the simulations easily yield control to the collections of animated memory precedents aroused by the scene, which subconsciously search out emotional value peaks and troughs based on the biologically inspired pleasure/pain value axis. Together with metaphoric links and substitutions, this allows creativity in the choice of physical action, while remaining broadly aligned to mental goals.

Language and symbols are sensory objects too, and are extensions of the same simulation processes. But they have the special properties of indexing 3D objects and scene narrative; including empathic states, subjective values and goals. Allowing ideas to be shared socially, with wisdom traveling through the generations.

The gist of this research is that all conscious and intelligent processes center around 3D simulation; with language and symbols used for indexing and scripting. That all knowledge can be understood in terms of 3D model behavior based on precedent. That software designed to handle 3D models and environments, will be central to AI and this commercial software is advancing rapidly.[15].

## Notes

This work is influenced by the work of Ayn Rand [13] and the philosophy of objectivism [12], and in particular, her proofs of: the primacy of existence; the validity of our senses; the instantiation of instances from concepts; the proper relationship between consciousness and reality, and of reason and emotion. This paper builds on her work to integrate 3D forms as the primary vehicle for connecting human percepts to higher mental abstractions.

## References

1. Dawkins R (1982) *The Extended Phenotype*. Oxford University Press, Oxford.
2. Dennet D (1991) *Consciousness Explained*. Little Brown, Boston, MA.
3. Donald M (2001) *A Mind so Rare: the Evolution of Human Consciousness*. W. W. Norton, New York.

4. Le Doux J (1998) *The Emotional Brain: the Mysterious Underpinnings of Emotional Life*. Touchstone Books, New York.
5. de Garis H (1999) *Artificial Embryology and Cellular Differentiation*. In: Bentley P (ed) *Evolutionary Design by Computers*, Morgan Kaufmann, San Francisco, CA.
6. Harold FM (2001) *The Way of the Cell*. Oxford University Press, Oxford.
7. Kosslyn S, Koenig O (1995) *Wet Mind: The new Cognitive Neuroscience*. Free Press, New York.
8. Kurzweil R (2001) *The Paradigms and Paradoxes of Intelligence, Part 2: The Church Turing Thesis*. [www.kurzweilai.net/articles/art0256.html](http://www.kurzweilai.net/articles/art0256.html).
9. Kurzweil R (2001) *The Law of Accelerating Returns*. [www.kurzweilai.net/articles/art0134.html](http://www.kurzweilai.net/articles/art0134.html).
10. Moravec H (1999) *Robot: Mere Machine to Transcendent Mind*. Oxford University Press, Oxford.
11. Minsky M (1988) *The Society of Mind*. Touchstone Books, New York.
12. Peikoff L (1993) *Objectivism: The philosophy of Ayn Rand*. Meridian, New York.
13. Rand A, Binswanger H, Peikoff L (eds) (1990) *An Introduction to Objectivist Epistemology*. Meridian, New York.
14. Schacter D (1996) *Searching for Memory: The Brain, The Mind and The Past*. Basic Books, New York.
15. Yudkowsky E (2000) *General Intelligence and Seed AI*. This volume, and [www.singinst.org/GISAI/](http://www.singinst.org/GISAI/)