PETER B. LLOYD

GLITCHES IN THE MATRIX . . . AND HOW TO FIX THEM

Why, exactly, do the rebels have to enter the Matrix via the phone system (which after all doesn't physically exist)? And what really happens when Neo takes the red pill (which also doesn't really exist)? And how does the Matrix know what fried chicken tastes like? Technologist and philosopher Peter Lloyd answers these questions and more.

As the essays throughout this book demonstrate, the Wachowski Brothers designed *The Matrix* to work at many levels. They carefully thought through the film's philosophical underpinnings, religious symbolism, and scientific speculations. But there are a few riddles in *The Matrix*, aspects of the film that seem nonsensical or defy the laws of science. These apparent glitches include:

- **The Bioport**—how can a socket in your head control your senses? How can it be inserted without killing you?
- **The Red Pill**—since the pill is virtual, how can it throw Neo out of the Matrix?

- **The Power Plant**—can people really be an energy source?
- **Entering and Exiting the Matrix**—why do the rebels need telephones to come and go?
- **The Bugbot**—what's the purpose of the bugbot?
- **Perceptions in the Matrix**—how do the machines know what fried chicken tastes like?
- **Neo's Mastery of the Avatar**—how can Neo fly?
- **Consciousness and the Matrix**—are the machines in the Matrix alive and conscious? Or are they only machines, intelligent but mindless?

This essay addresses these questions and shows how these seeming glitches can be resolved.

THE BIOPORT

Can the machines really create a virtual world through a bioport? And how does it work? The bioport is a way of giving the Matrix computers full access to the information channels of the brain. It is located at the back of the neck—probably between the occipital bone at the base of the skull, and the first neck vertebra. Wiring would best enter through the soft cartilage that cushions the skull on the spinal column, and pass up through the natural opening that lets the spinal cord into the skull. This avoids drilling through bone, and maintains the mechanical and biological integrity of the skull's protection. A baby fitted with a bioport can easily survive the operation.

The bioport terminates in a forest of electrodes spanning the volume of the brain. In a newborn, the sheathed mass of wire filaments is pushed into the head through the bioport. On reaching the skull cavity, the sheath would be released, and the filaments spread out like a dandelion, gently permeating the developing cortex. Nested sheaths would release a branching structure of filamentary electrodes. As each sheathed wire approaches the surface of the brain, it releases thousands of smaller electrodes. In the neonate, brain cells have few synaptic connections, so the slender electrodes can penetrate harmlessly.

With its electrodes distributed throughout the brain, the Matrix could deliver its sensory signals in either of two places: at the sensory portals or deep inside the brain's labyrinth. For example, vision could be driven by electrodes on the optic nerves where they enter the brain. Artificial signals would then pass into the visual cortex at the back of the brain, which would handle them as if they had come from the eyes. Correspondingly, outgoing motor nerves would also have electrodes at the boundary of brain and skull. This simple design mirrors the natural state of the brain most closely. It is not, however, the only possibility. Electrodes could alternatively be attached in the depths of the brain, beyond the first stages of the visual cortex. This would greatly simplify the data processing. In normal perception, most of the incoming information isn't processed; information you aren't paying attention to is filtered out. If the Matrix were to deliver information directly to the output axons from the sensory cortex—as opposed to the input to the cortex—then it would save itself the job of filling in all those details.

One scene tells us which method the Matrix uses. When Neo wakes and finds himself in a vat, he pulls out the oxygen and food tubes, drags himself out of the gelatinous fluid, and—perceives the world. The fact that he can see and hear proves that the visual and auditory cortices of his brain are working. This wouldn't be possible if the Matrix had put its sensory data into the deeper centers of his brain. For then his sensory cortex would have been bypassed: it would never have received any stimulation, and would have wasted away. In that case, Neo would wake from his vat and find himself blind and deaf, with no sense of smell or taste, no feeling of touch or heat in his skin, no awareness of whether he was vertical or horizontal, or where his arms or legs were. The Matrix must have input its visual data just where the optic nerve from the eyeball passes into the skull, rather than in the midst of the brain's vision processing. Likewise, Neo's ability to walk and use his arms shows that the motor cortex is also developed and functioning. Indeed, even the cerebellum, which controls balance, must be working. So, the Matrix must be capturing its motor signals from the brain's efferent nerves after they have finished with the last stage of cortical processing, but before the nerves pass out of the skull.

The rebels use the bioport to load new skills into their colleagues' brains—writing directly into permanent memory. The Matrix itself never implants skills in this way; folks in the virtual world learn things in the usual manner by reading books and going to college. So, why did the architects of the Matrix build into the bioport this capability to download skills? It is actually a by-product of how the bioport is installed. They could have attached electrodes to just the sensory and motor nerve fibers. That, though, is difficult: the installer must predict where each nerve fiber will be anchored, which is hard to do reliably, given the plasticity of the neonate brain; and it must navigate through the brain tissue to find these sites. A more robust and adaptable method is to lay a carpet of electrodes throughout the whole brain, and let the software locate the sensory and motor channels by monitoring the data flows on the lines.

That spare capacity remains available for others to exploit, and the rebels use it to download kung-fu expertise into Neo's brain and to implant helicopter piloting skills into Trinity's. If the Matrix ever learned this technique, it could create havoc for the rebels, implanting impulses to serve its own ends.

THE RED PILL

Morpheus offers Neo the choice of his lifetime, in the form of the famous red and blue pills. But what can a virtual pill do to a real brain? We have seen that the Matrix interacts with the brain only in the sensory and motor nerve fibers. It does not affect the inner workings of the brain, where a real psychoactive chemical would have to act. Minor analgesics such as aspirin would work by having their effect outside the brain centers, canceling out pain inputs from the avatar software.

The blue pill is probably a placebo. Morpheus says only, "You take the blue pill and the story ends. You wake in your bed and you believe whatever you want to believe." We never know what, if anything, the blue one would do.

So, how does the active pill, the red one, work? Since virtual aspirin can work as a painkiller, the avatar's software module must be able to accept instructions to cancel out any given sensory input. Evidently, the red pill gives the avatar a blanket command to cancel all such input. It thereby obliterates Neo's perception of the virtual world, which the Matrix has been feeding to him throughout his life. Instead of sitting on a chair in a hotel room, Neo sees and feels for the first time that he is immersed in a fluid. The perception of this filters through into his perceptions of the Matrix's own imagery. Neo touches a mirror, and finds it a viscous fluid that clings to his finger and then seeps along his arm, covering his chest and slithering down his throat. A blend of bodily perceptions and mental imagery is typical of what happens when you wake from a dream; external perceptions are distorted to fit the contents of the dream. Your dream of falling off a cliff might fade into falling out of bed. In the film, the liquefied mirror is seen only by Neo, not the others in the room. His real bodily sensations are, for the first time, sweeping into his brain, which struggles to integrate them into the stable narrative he has lived in up to that moment.

Another route out of the Matrix, besides the red pill, would be meditation. The Buddhist practice of *vipassana*[1] gives adepts penetrating insights into their own mental processes. It rolls back the barrier between conscious awareness and the subconscious. An adept of *vipassana*, living in the Matrix, would discover the interface between the Matrix's electrodes and the brain's wetware. The expert practitioner could override the Matrix's stream of imagery, and see reality. Morpheus mentions that someone did break free from the Matrix. Perhaps meditation was the key. To attain that expertise, however, would take years of effort. Leading other people to the truth would require a school of meditation to train new recruits for years, to pursue what one individual claimed was the truth, but everyone else dismissed as fantasy. No doubt this is what the Oracle is gently encouraging. But it is not surprising that the red pill was invented as a fast-track route.

---

[1] In the oldest form of Buddhism, Theravada, the two major forms of meditation are Vipassana (the Pali word for "insight") and its complement Samatha ("tranquility").

Vipassana consists in systematically attending to the individual elements that make up the contents of consciousness. It involves persistently turning away from the ceaselessly arising tide of chatter in the mind. Over time, the chatter subsides, and preconscious activity becomes more readily observed. Laboratory data support claims that long-term practitioners acquire a conscious awareness of brain microprocesses, possibly down to the cellular level. See Shinzeng Young's works.

Morpheus's team monitors Neo's progress. As he realizes that he is immersed in fluid, Neo panics, and his instinct to escape drowning compels him to drag the tubes out of his mouth. Like waking out of a dream, Neo finds the sensible world rushing in on him, and it is remarkable that his manual coordination has been so well preserved by the Matrix system. He grabs the tubes and yanks them out, using weak hands that had never before grasped anything.

When Neo's exit from the Matrix is detected, a robot inspects him and flushes him out of his pod. Too weak to swim, he must be pulled out of the wastewater pool without delay. How are the rebels to find him? In a power plant vast enough to house the human race, there would be thousands of effluent drains. As Morpheus mentions to Neo, "the pill you took is part of a trace program." Besides canceling Neo's sensory inputs, the red pill also puts a unique reference signal onto the Matrix network. When the *Nebuchadnezzar*'s computer locates that signal, it can work out Neo's physical location and order the hovercraft to the appropriate chute. In the tense moment before that reference signal is located, the worried Morpheus says, "We're going to need the signal soon," and Trinity exclaims that Neo's heart is fibrillating as the panic threatens to bring on a heart attack. Apoc finds the reference signal just in time, before Neo's brain disengages from the Matrix network and the signal vanishes.

THE POWER PLANT

During the armchair scene, we have what is probably the most criticized element in *The Matrix* story line. Morpheus claims that the human race is imprisoned in a power station, where human bodies are used as a source of bioelectricity. This is engineering nonsense; it violates the fundamental law of energy conservation. The humans would have to be fed, and the laws of physics demand that the energy consumed as food must be greater than the energy generated by the human body. That Morpheus has misunderstood what is going on is underscored by his mention in the same speech of the machines' discovery of a new form of nuclear fusion. Evidently, the fusion is the real source of energy that the machines use. So what are humans doing in the power plant? Controlled fusion is a subtle and complex process, requiring constant monitoring and micromanaging. The human brain, on the other hand, is a superb parallel computer. Most likely, the machines are harnessing the spare brainpower of the human race as a colossal distributed processor for controlling the nuclear fusion reactions. (Sawyer comes to a similar conclusion elsewhere in this volume—Ed.)

ENTERING AND EXITING THE MATRIX

The virtual world of the Matrix is not bound by physical laws as we know them, but for the virtual world to be consistently realistic, the laws of physics must be followed where they can be observed by humans. Access into and out of a virtual world is a problem, because materializing and dematerializing violate the conservation of mass and energy. Furthermore, whatever was previously in the space occupied by the materializing body must be pushed out of the way; and would be pushed with explosive speed if the materialization is instantaneous. Conversely, on dematerialization, the surrounding air would rush in to the vacated space with equal implosive force. There are no such explosions and implosions in *The Matrix*, so how do the rebels do it?

In the Matrix computer, software modules represent the observable objects in the virtual world, and these modules interact by means of predefined messages. One such message issued by a virtual human body, or "avatar," is, "What do I see when I look in the direction V?" A module whose object lies on the line of sight along V will respond with a message specifying the color, luminosity, and texture that the human should see in that direction. If a rebel's avatar is to be visible to other people who are immersed in the Matrix world, the *Nebuchadnezzar*'s computer must pick up those "What-do-I-see" requests and reply with its own "You-see-this" message.

A virtual human body does not send "What-do-I-see?" message to all other modules in the Matrix, or else it would overload the network. It refers to "registers" of modules, which record the virtual objects' shape, size, and position. Simple geometry then tells it which modules to target. For efficiency, each visible volume of space, such as the room of a building, has its own register.

The key step in materializing a body in a given space is for its module to be inserted into that space's register. For dematerializing, it is deleted from the register. Once it is registered, anyone looking in that direction will see that module's virtual body. The Matrix cannot let a software module insert itself arbitrarily into a register, since that could violate the conservation of mass if it led to an object's materializing in an area that has a conscious observer.

Registers for unobserved spaces are not constrained in this way. If nobody is watching a room and its entrances, then a body can safely materialize in it without observably breaking the simulated laws of physics.

This does not mean that the laws of physics break down as soon as all observers leave a room. The table and chair do not start to float around against the law of gravity when nobody is looking. Rather, the Matrix simply does not bother to run the simulation for a room that nobody is looking at. In its register, it retains details of where each object is, but the room is no longer rendered as visual and tactual imagery.

So, when the *Nebuchadnezzar*'s computer wants to materialize a rebel, it must find some unobserved room, and insert the data module for the rebel's body into the register for that room. Subsequently, if someone else enters the room, he will see the rebel just like any other object in the room. And the rebel can walk out of the room into any other part of the Matrix world in the normal manner. This is how rebels materialize in the Matrix without causing explosions or breaching the integrity of the simulation.

When a rebel exits, the module that simulates her body is deleted from the register. This must happen only when the body is not being observed. There is, however, an intermediate state, "imperception," which effectively takes the body out of the virtual world even while the data module is still in the register. This is an emergency procedure that the *Nebuchadnezzar*'s software uses for fast escapes.

Although the Matrix software cannot insert or delete a module while its object is being observed, it does allow any module to change its appearance. The agents use it whenever they enter the world. An agent never materializes or dematerializes, but changes the appearance of another person's avatar to match the personal qualities of the agent.

To make a rebel imperceptible, the *Nebuchadnezzar*'s computer changes the body's visible appearance to be transparent; and the body's mechanical resistance to that of the air. From an observer's perspective, the body has melted into air. From a software perspective, the data module is still on the register but simulating a body indistinguishable from thin air. Later, when the scene is no longer being observed by anybody, the module will be deleted.

We see this happen only once, when Morpheus leaves the subway. Once the *Nebuchadnezzar*'s computer has located his avatar, it sends an instruction to make it invisible. This does not affect the whole avatar at once: the module has to calibrate its appearance to match exactly its surroundings. The first part of the body to receive the instruction is the nervous tissue of the ear, and this at first glows bright white, before settling down to a state of transparency. The rest of the body follows. Its appearance oscillates around whatever is visible in the background, settling down to transparency: where the Morpheus stood, we see the background shimmer momentarily. The solidity of the body then fades: moments after Morpheus's body has become invisible, the telephone handset that had rested in his hand drops, slowly at first, toward the ground. The observed sequence is consistent not with the sudden deletion of the body's module, but rather with its changing its appearance.

HARD LINES

Telephones play a key role in entering and leaving the Matrix. But the rebels do not travel through the telephone lines as energy pulses. There is no device at the end of the telephone for reconstructing a human body from data: all you would get is noise in the earpiece. Furthermore, the bandwidth of a telephone line is too narrow to ship an entire human being. Finally, nothing at all ever really travels along the lines in the Matrix world, as they are only virtual.

Instead of being a conduit for transporting dematerialized rebels, the telephone line is a means of navigation. It pinpoints where a rebel is to enter or leave the Matrix.

To enter the vast Matrix requires specifying where the avatar is to materialize. To get an avatar into the Matrix world, the rebels must use some strictly physical navigation. This is done with the telephone network, which has penetrated every corner of the inhabited world with electronic devices, each of which has a unique, electronically determined label. Without knowing anything of human society and its conventions, the physics modules of the Matrix can determine where any given telephone number terminates.

How are the rebels to give a telephone number to the Matrix? They must dial it, but they cannot simply pick up a handset and make a call to a number inside the Matrix world, for any handset in the *Nebuchadnezzar* is connected to the real world telephone network, not the Matrix's virtual network. Inside the Matrix, a call must be placed subtly, without observably breaching the simulated laws of electromechanics.

To see how this can be done, we need to know something of the infrastructure of the Matrix. Monolithic computer systems are unreliable, so the Matrix is instead an assemblage of independent modules, each having a unique "network address." For a module to communicate with another, it will put a data message on the network with the address of the intended destination. Neither module need know where the other one is inside the electronic hardware of the Matrix computer. They might be inches apart, or a mile away.

This scheme is robust and flexible. There is no central hub, and individual modules can be plugged into, or taken out of, the network without disturbance. Conversely, the rebels can easily hack into it. Once they are linked into the network, their equipment can simply pretend to be another module. It can place data messages onto the system, which will be routed just like authentic messages, and be received and read by the addressed module. So, to initiate a telephone call, the crew will place a data message on the network, addressed to any module that simulates an aerial for receiving calls from cell phones. Some such node will pick up and read the counterfeit data message just as if the message had been sent by a bona fide source. On getting this message, the aerial module will carry out its role in handling a telephone call.

The *Nebuchadnezzar*'s operator maintains contact with rebels who are in the Matrix even while the hovercraft is moving, so they must use radioports onto the network. The rebels might have installed their own rogue radio receiver—mechanically securing it in some dark corner, and plugging its data cable into a spare socket of a router. More likely, the Matrix itself uses radio as part of its network infrastructure, and the rebels broadcast their counterfeit messages on the same frequency.

Materializing or dematerializing, however, needs a network address, which is gotten as follows. When the *Nebuchadnezzar* makes a "phone call" into the Matrix, it places on the network a packet saying "Place this call for (212) 123-4567" or whatever the telephone number is, together with the *Nebuchadnezzar*'s own network address as a return label, such as 9.54.296.42. When the call is picked up, the Matrix will return a data packet, addressed to the *Nebuchadnezzar*, saying "Message for 9.54.296.42: call connected to telephone (212) 123-4567." All the *Nebuchadnezzar*'s computer has to do is listen out for its own address, and it will find attached to it the network address of the telephone equipment.

As soon as the answering machine picks up the incoming call, the *Nebuchadnezzar* will get the network address of that destination.

Essentially the same job must be done when a rebel leaves the Matrix world. In order to disengage the rebel from his or her avatar, the *Nebuchadnezzar*'s computer must again get a fix on the avatar's location within the virtual world. As before, it is not enough to locate the avatar's virtual body in terms that relate to human culture. It is no use to say that Neo is at 56th and Lexington. Rather, it needs a network address that the Matrix's operating system can follow. Of course, the *Nebuchadnezzar* gets it by calling a telephone in the Matrix world, which must be answered for the network address to be passed back to the *Nebuchadnezzar*. Once that has happened, the avatar's module can be deleted from the register for that location.

Why don't the crew navigate their exits with the stylish cell phones that all the rebels carry? Why hunt for a land line (called a "hard line" in the film) under hot pursuit from the agents? The answer is that the cell phones are not part of the Matrix world and do not have network addresses known to the Matrix software. The cell phone is projected into the Matrix world by the *Nebuchadnezzar*'s computer, along with the avatar's body and clothes—and the weapons that Neo and Trinity eventually bring in with them. The software that simulates the cell phones is running inside the *Nebuchadnezzar*'s computer, not the Matrix's computer, so the rebels must find a land line—which are somewhat scarce in an era when everyone has a cell phone.

THE BUGBOT

Before Neo is taken to meet Morpheus, the agents insert a robotic bug into him. Trinity extricates the bugbot before it can do any harm. But what was the bugbot for? Given that it operates inside the human body, the bugbot should be as small as possible. Yet, it is clearly much bigger than a miniature radio beeper needed for tracking Neo's whereabouts. Trinity says that Neo is "dangerous" to them before he is cleaned. We can infer that the bugbot is actually a munition, probably a semtex device that will detonate when it hears Morpheus's voice, killing both Neo and Morpheus and everyone else in the room.

Just before it is implanted, the bugbot takes on the appearance of an animate creature, with claws writhing. Yet, after Trinity has jettisoned it out of the car window, it returns to an inert form. It is another illustration of the agents' limited use of the shapeshifting loophole in the Matrix software, that lets an object transform its properties under programmed commands.

PERCEPTIONS IN THE MATRIX

At dinner on the *Nebuchadnezzar*, Mouse ponders how the Matrix decided how chicken meat should taste, and wonders whether the machines got it wrong because the machines are unable to experience tastes.

A nonconscious machine cannot experience color any more than taste. A computer can store information about colored light, such as a digitized photograph, but it does so without a glimmer of awareness of the conscious experience of color. The digitized picture will evoke conscious colors only when someone looks at it. All other sensations that you can be conscious of will elude the digital computer.

The feel of silk, the texture of the crust of a piece of toast, feelings of nausea or giddiness: these are all unavailable to insentient machines. This being so, Mouse could have doubted whether the Matrix would know what anything should taste, smell, look, sound, or feel like.

But the Matrix doesn't need to experience the perceptual qualities to get them right. As we have seen, the Matrix feeds its signals into the incoming nerves where they enter the brain, not into the deeper nerve centers. So when you eat (virtual) fried chicken inside the Matrix, the Matrix will activate nerves from the tongue and nose, and the brain will interpret them as taste sensations. What the Matrix puts in will be a copy of the train of electrical impulses that would actually be produced if you were eating meat. Because of the way that the Matrix has been wired into the brain, it has less freedom than Mouse assumed. Whilst the Matrix cannot know tastes itself, it can nonetheless know which chemosensory cells in a human's nose and mouth yield the requisite smell and taste.

NEO'S MASTERY OF THE AVATAR

For purists of science-fiction plausibility, Neo's superhuman control over his avatar body is a troubling element in the film. The final triumphal scene, where Neo flies like Superman, has especially come under criticism. But is it completely at odds with what we have inferred about the Matrix? And how does Neo transcend his human limits?

The Matrix interacts with the brain, but the brain in turn affects the body. When Neo is hurt in training, he finds blood in his mouth. He asks Morpheus, "If you are killed in the Matrix, you die here?" and gets the cryptic reply: "The body cannot live without the mind." But it cuts both ways; ultimately, Neo's avatar is killed inside the Matrix, causing the vital functions to cease in his real body.

Mental states and beliefs can affect the body in several ways. In the placebo effect, the belief that a pill is a medicine can cure an illness; in hypnosis, imagining a flame on the wrist can induce blisters. In total virtuality, the mind accepts completely what is presented. If the Matrix signals that the avatar's body has died, then the mind will shut down the basic organs of the heart and lungs. Actual death will inevitably ensue, unless fast action is taken to get the heart pumping again.

In the climactic scene, Agent Smith kills Neo's avatar within the Matrix. Neo's brain accepts this fate: it stops his heart and loses conscious awareness. His real brain, however, retains enough oxygenated blood to keep it functioning for approximately three minutes, after which it would begin to suffer irreversible damage and, a few minutes later, brain-death. During this time, the auditory cortex keeps on working and digests what Trinity says, albeit unconsciously. Trinity's message is comprehended by Neo's subconscious mind, and a deep realization that the Matrix world is illusory crystallizes in his mind. At an intellectual level, Neo already believed this, but now he knows it at the visceral level of the mind, the level that interfaces with his physiology. Empowered by the insight that his avatar's death is not his death, Neo regains control of his avatar—not only resurrecting it but attaining superhuman powers: the avatar can stop bullets, and fly into the air.

Neo's new powers contrast with the rigid compliance with simulated physical laws that the Matrix generally adheres to. It reveals that Neo has gained direct access to the software modules that simulate his avatar body. That raises two questions: Why does the avatar software accept commands to transform itself, when normally it strictly follows a physical simulation? And, how can Neo's brain issue such commands, which are obviously outside the scope of the normal muscular signals?

The software that simulates the avatar must have a special port, intended for use only by agents, which accepts commands to change the internal properties of the avatar's body. Agents use this facility to embody themselves in human avatars. Like all software, the avatar will obey such commands wherever they originate, provided that they are correctly formulated. We saw earlier how the *Nebuchadnezzar*'s computer used this transformative power to make Morpheus disappear from the subway station. Now Neo's brain is directly using the same command port.

Commands to transform the body cannot travel on the wires that carry the regular muscular signals from the brain to the avatar module. So, they use some of the many other, seemingly redundant, data lines that terminate throughout the rest of the brain. That those lines are hooked up at all on the Matrix end is a spin-off from the Matrix architect's use of general-purpose interfaces. When a newborn human baby is connected to the software module that runs its avatar, there is no way to predetermine which wires carry which data streams. So, at the Matrix end, each line is free to connect to any data port of the avatar module. Some data ports emit simulated signals from virtual eyes and other sense organs, and they will connect with the brain's sensory cortex; others will accept motor commands to carry out simulated contractions of virtual muscles, and they will link up with the motor cortex. In a feedback process that mirrors how the natural plasticity of the brain is molded to its function, useful connections are strengthened and the useless are weakened. As a baby grows into an infant, it gains feedback through using the simulated senses and muscles of the avatar, and therefore its brain builds up the normal strong connections to the conventional input and output channels. But it lacks the abstract concepts needed to use the special port that accepts transformation commands. So the brain's connection with those lines atrophies. Nevertheless, the hardware for that potential connection remains in place. In Neo's kung fu training, his brain rediscovers the abandoned data lines, and he starts to issue rudimentary transformations, giving his avatar's muscles superhuman strength. Only with the deep insight that he gains from being woken after his avatar's death, does he acquire the mental attitude needed to harness that transformative function fully.

The existence of the transformational back door into the avatar software is a security hole that the architects of the Matrix never imagined would be used by mere humans—but now it threatens the very existence of the Matrix, as Neo exploits the power it gives him.

CONSCIOUSNESS AND THE MATRIX

The last question I will address in this essay is a complex one, and one that continues to be explored and debated in scientific and philosophical circles. Can machines be conscious? In everyday life, the machines are so dumb that we can ignore this question, and so we do not have an established criterion for judging whether the intelligent machines of science fiction are conscious. How similar must a machine be to a human for it to be conscious? Humans have a cluster of properties that always hang together: they have conscious perceptions and emotional feelings, they have opinions and beliefs, intuition and intelligence, they use language, and they are alive and warm-blooded, and have a biological brain. We do not, in everyday life, have to separate out those concepts and decide which ones are necessary and sufficient for sentience. The properties all come as a package. In contrast, the lower animals are like us but do not use language and are not as intelligent as we are. So, it is believed that the higher animals probably have basic conscious perceptions—such as colors and sounds, heat and cold—much as we do, but they lack the superstructure of thought. But what about machines that are intelligent and use language, but are not made of biological tissue? Could they be conscious?

To respond rationally to this emotive challenge, we need to be clear about the ideas that are involved. The commonest and most damaging conflation is that of "intelligence" and "consciousness." Alan Turing, in his celebrated paper that introduced the Turing Test, used the terms interchangeably—but mathematicians are notorious for playing fast and loose with their terms. Philosophers, whose trademark is the careful delineation of concepts, have always insisted on maintaining the distinction. Intelligence is the capacity to solve problems, while consciousness is the capacity for the subjective experience of qualities.

As we shall see, intelligence can be attained without consciousness.[2] A digital computer can be programmed to perform intelligent tasks such as playing chess and understanding language by well-defined deterministic processes, without any need to introduce enigmatic conscious experiences into the software. On the other hand, a conscious being can have subjective experiences—such as seeing the color red, or feeling anger—with needing to use intelligence to solve any problems. An android could be vastly more intelligent than

[2] For an alternative perspective, see Kurzweil's essay in this volume. —Ed.

any human and still lack any glimmer of interior mental life. On the other hand, a creature might be profoundly stupid and still have subjective experiences.

Agent Smith is an example of a machine that manifests human-like behavior—which, if you witnessed such words and gestures in a human, you would immediately regard them as showing conscious emotions and volitions. Indeed, it is the immediacy of the interpretation that is deceptive. When you see someone laugh with joy, or scream in pain, you do not knowingly infer the person's mental state from those outward signs. Rather, it is as if you see the emotions directly. Yet, we know from accomplished actors that these signs of emotions can be faked. Therefore, you are indeed making an inference, albeit an automatic one. It is a job of philosophy to scrutinize such automatic inference. When you see another human being emoting, your inference is not based wholly on what you see, but also on background information (such as whether the person is acting on the stage). More fundamentally, you are relying on the reasonable assumption that the person's behavior arises from a biological brain just as yours does. Whenever those premises are undermined, you inevitably revise any inferences you have made from the emoting. If the emoting stops and people around you clap, you realize it was a piece of street theatre, and the person was only acting out those emotions. Or, if the person has a nasty car accident that breaks open his head, revealing electronic circuitry instead of a brain, you realize that it was only an android and you may conclude that it was only simulating emotions.

A key step in the inference is the premise that the emotion plays a role in the causal loop that produces the outward words and gestures. If, instead, we have established that the observed words and gestures are wholly explained in some other way, without involving those emotions—then the inference collapses. The exterior emoting behavior then ceases to count as evidence for an interior emotional experience. If we know that an actor's words and gestures are scripted, then we cease to regard them as evidence for an inward mental state. Likewise, if we know that the words and gestures of an android or avatar are programmed, then they too cease to support any inference of a mental state.

In an android, or in a software simulation of a human such as an agent, words and gestures are produced by millions of lines of programmed software. The software advances from instruction to instruction in a deterministic manner. Some instructions move pieces of information around inside memory, others execute calculations, others send motor signals to actuators in the body. Each line of code references objective memory locations and ports in the physical hardware. It may do so symbolically, and it may do so via sophisticated data structures, for example, using the tag "vision-field" to reference the stabilized and edge-enhanced data from the eye cams. Nevertheless, nowhere in the software suite does the code break out of that objective environment and refer to the enigmatic contents of consciousness. Nor could the programmer ever do so, since she would need an objective, third-person pointer to the conscious experience—which, being a subjective, first-person thing, cannot be labeled with such a pointer.

Everything that the android says and does is fully accounted for by its software. There is no explanatory gap left for machine consciousness to fill. When the android says, "I see colors and feel emotions just as humans do," we know that those words are produced by deterministic lines of software that functions perfectly well without any involvement of consciousness. It is because of this that the android's emoting does not provide an iota of evidence for any interior mental life. All the outward signs are faked, and the programmer knows in comprehensive detail how they are faked.

This point is systematically ignored by the mathematicians and engineers who enthuse about artificial intelligence. You have to go next door, to the philosophy department, to find people who accord due importance to it. Even if, by some unknown means, the android possessed consciousness, it could never tell us about it. As we have seen, everything the android says is determined by the software. Even if, somewhere in the depths of its circuit boards, there was a ghostly glimmer of conscious awareness or volition, it could never influence what the android says and does.

Could it be that the information in the computer just *is* the conscious experience? This argument is popular with information engineers, as it seems to allow them to gloss over the whole mind-body problem. It is flawed because information and conscious experience have different logical structures. Namely, information exists only as an artifact of interpretation; but experience does not stand in need of interpretation in order for you to be aware of it. If I give you a disk holding numerical data (21, 250, 11, 47; 22, 250, 15, 39. etc), those numbers could mean anything. In one program, they are meteorological measurements—temperature, humidity, rainfall. In another, they are medical—pulse rate, blood pressure, body fat. The interpretation has no independent reality; the numbers have no inherent meaning by themselves. In contrast, conscious experience is fundamentally different. If you jam your thumb in a door, your sensation does not need first to be interpreted by you as pain. It immediately presents as pain. Nor can you reinterpret it as some other sensation, such as the scent of a rose. Conscious experiences have real, subjectively witnessed qualities that do not depend for their existence on being interpreted this way or that. They intrinsically involve some quality over and above mere information.

Another popular argument is to appeal to "emergence." Higher-level systems are said to "emerge" from lower-level systems. The simple classic example is that of thermodynamic properties, such as heat and temperature, which emerge from the statistical behavior of ensembles of molecules. Yet the concept of "temperature" just does not exist for an isolated molecule, although billions of those molecules collectively do have a temperature. In like manner, it has been suggested, consciousness emerges from the collective behavior of billions of neurons, which individually could never be conscious on their own. But emergent properties are, in fact, artifacts of how we describe the world, and have no objective existence outside of mathematical theories. An ensemble of molecules may be described in terms of either the trajectories of individual molecules or their aggregate properties, but the latter are invented by human observers for the sake of simplifications. The external reality comprises only the molecules: the statistical properties, such as average kinetic energy, exist only in the mind of the physicist. Likewise, any dynamic features of the aggregate behavior of brain cells exist only in the models of the neuroscientists. The external reality comprises only the brain cells. Yet, as you know, when you jam your thumb in the door, the pain is real and present in the moment; it is not a theoretical construct of a brain scientist.

So there are good reasons for believing that machines are not conscious. But—wouldn't these arguments apply equally to brains? Surely a brain is just a bioelectrochemical machine? It obeys deterministic programs that are encoded in the genetic and neural wiring of the brain. Yet, if our argument that machines are not conscious can also apply equally to brains, then the argument must be flawed—since we know that our own brains are indeed conscious!

The answer is that there are certain processes in brain tissue that involve nondeterministic quantum-mechanical events. And, working through the chaotic dynamics of the brain, those minute phenomena can be amplified into overt behavior. The nondeterminism opens a gateway for consciousness to take effect in the workings of the brain.

As we saw earlier, you can report only the conscious experiences that are in the causal loop that gives rise to the speech acts. If you can report that you are in pain, then the pain sensation must exert a causal influence somewhere in the chain of neural events that governs what you say and write. A step that is physically nondeterministic provides a window of opportunity for consciousness to enter into that causal chain. Since we, as humans, know that we do express our conscious perceptions, we can infer that there must be some such nondeterminism somewhere in the brain. So far, quantum-mechanical events constitute the only known candidate for this. For example, Roger Penrose and Stuart Hammeroff have formulated a detailed theory of how quantum actions in the microtubules of brain cells could play this role. The jury is still out on whether the microtubules really are the locus at which consciousness enters the chain of cause and event.

A conventional, deterministic computer has no such gateway into consciousness. So androids, and virtual avatars, that are driven by computers of that kind, cannot express conscious awareness and their behavior therefore can never be evidence for consciousness. But, if a machine were to be built that used quantum computation in the same way that the brain does, then there is no philosophical reason why that machine could not have the same gateway to consciousness that a living being does. This is not because the quantum module lets the machine carry out computations that a classical computer cannot do. Whatever the quantum computer can do, a classical one can also do, albeit more slowly. Rather, it is the specific implementation of the quantum computer that provides the bridge into conscious processes.

In *The Matrix*, there is no reason to think that the machines are equipped with the kind of quantum computation needed to access consciousness. Quantum computation is not mentioned in the film, and there is circumstantial evidence that the Matrix and its agents are devoid of conscious thought.

Therefore the agents—which are software modules within the Matrix—are intelligent but mindless automata. For the most part, the agents behave unimaginatively, and we might naively think that this corroborates their lack of awareness. Yet, Agent Smith exhibits initiative and seems, in his speech to Morpheus, to evince a conscious dislike of the human world. But is he genuinely conscious, or only mimicking humans? In fact, Smith gives himself away when he says about the human world, "It's the smell, if there is such a thing . . . I can taste your stink and every time I do, I fear that I've somehow been infected by it." Smith's own logical integrity obliges him to doubt the existence of that noncomputable quality that humans talk about: the conscious experience of smell. When Smith says, ". . . the smell, if there is such a thing," he is exhibiting the mark of the automaton. This is corroborated when he then tells Morpheus that he can "taste your stink," revealing that Smith simply does not understand the differentiation of senses in the human mind. For a computer, data are interchangeable, but for a human, tastes, smells, colors, sounds, and feels, are irreducibly different. This fact eludes Agent Smith.

Smith is mimicking human behavior as a tactic to trick Morpheus into cooperation. As the interrogation is getting nowhere, Brown suggests, "Perhaps we are asking the wrong questions." So Smith pretends to talk like a human, to gain Morpheus's empathy. Needless to say, the tactic fails completely.