Kurzweilaccelerating intelligence essays

Glitches Reloaded

June 2, 2003 by Peter B. Lloyd

In Matrix Reloaded, how can Neo fly and use telekinesis if the Matrix is supposed to a physics simulation? Peter Lloyd decodes this and other technical enigmas–reverse-engineering the design of the Matrix and the "Meta-Matrix" of the underground Zion. And he delves into the rich philosophical and mythic elements of the film, such as the question of free will and who is the Architect and what does his speech tell us?

Published on KurzweilAI.net June 1, 2003. This is a follow-on from "Glitches in the Matrix," which was written for the anthology Taking the Red Pill*, edited by Glenn Yeffeth and published by Benbella Books in April 2003. An expanded version of this material will appear in an e-book,* Matrix Exegesis*, which will be available June 6, 2003. Warning: Includes plot spoilers for The Matrix and The Matrix Reloaded.*

This essay is about the following questions on the sequel film *The Matrix Reloaded*. I am referring to the questions below as "glitches" just to be consistent with Glenn Yeffeth's terminology in the book *Taking the Red Pill*.

- How can Neo stop the sentinels in the real world?
- How can Agent Smith be uploaded into a human brain?
- Why do people look the same in real life as they did in the virtual world of the Matrix?
- How can Neo fly and use telekinesis if the Matrix is supposed to a physics simulation?
- How do the vampire twins move around?
- How can Neo think fast enough to dodge a bullet?
- How does virtual food have so much effect?
- Is there free will?
- Can the future be foretold?
- Who or what is the Merovingian?
- Who or what is the "mother of the Matrix"?
- Who is the Architect and what does his speech tell us?

This essay does not analyze the game *Enter the Matrix*, which requires another essay …

Naming convention

Before we embark on answering these questions about the film (or, at least, trying answer them), it will be helpful to reflect on the Wachowskis' naming conventions. They have said in interviews that the names of their characters are significant. So we should find in them clues to the Wachowskis' riddles. Meaningful naming, as every programmer knows, is a commendable practice.

In the bad old days, programmers used to give their variables and subroutines names like X, Z10N, MTRXRL, and so on, which were unpronounceable and incomprehensible. Other programmers used anthropomorphic names such as FRED, JANE, and such like. The best approach is undoubtedly for the names to signify the nature and role of what they name, and to comply with a "naming convention"—a rule that guides the naming and leads to systematic rather than haphazard names.

As the Wachowskis tap into a lot of mythology (old and new) in their names, so we must dig into these myths to pick out the naming pattern. Needless to say, this does not commit us (or the Wachowskis) to believe any of the myths. As Councilor West could have said: "Belief is not a requisite for comprehension."

The first names we encounter in *The Matrix* belong to the realm of the mundane world: Mr. Thomas Anderson, and Mr. Smith. (We can read religious significance into Mr. Anderson's name, but this is not the time or place to do so.) Then we hear the self-assigned names of the rebels. These are hacker handles. Mr. Anderson calls himself Neo (the new one), and his colleagues call themselves Morpheus (the Greek god of dreams, who gives shape and form to ideas), Trinity (the three aspects of Christian divinity), Cypher (a secret code), Tank and Dozer (industrial machines), Apoc (the apocalyptic), Switch and Mouse. We also get the ship names: Nebuchadnezzar (who sought to interpret a dream he could not remember); Icarus (who flew on artificial wings); Gnosis (the Gnostic form of enlightenment); Logos (the rational projection of God); and Osiris (partner of Isis). And, in *Reloaded*, we have a swathe of names in Zion: Hamann (alluding to the German philosopher Johann Hamann), Jacob onboard the Gnosis (alluding to the Gnostic philosopher Jacob Boehme).

Of more interest in this essay are the assigned names of artificial intelligence programs: the Architect (the builder of the Matrix), the Oracle (who, I will argue, has access to the strategic planning programs of the Matrix, which I shall call "meta-intelligence" programs), the Merovingian (named after the supposed descendants of the Christ), Persephone (the Greek goddess who acquired emotional awareness by eating pomegranates, and who, I shall argue, is a program designed to probe human emotions by kissing human avatars), the Keymaker (self-evident), and the Seraph (an angel traditionally associated with being a serpent).

The names may be ambiguous clues, but clues nonetheless.

Matrix and Meta-Matrix

The plot action in the original film, *The Matrix*, had led us to believe that the external world is real. This is the world that contains Zion and its fleet of vessels such as the Nebuchadnezzar. It is supposed to be the "real" world, as opposed to the virtual world of the Matrix, in which almost all of the human race is imprisoned. The plot action in *The Matrix Reloaded* contains one scene that is a definitive proof that this is not the case, and some secondary scenes that corroborate this new interpretation. It transpires that the Zion world is actually another virtual world, which I shall call the Meta-Matrix. (In some newsgroup discussions, what I call the "Matrix" and the "Meta-Matrix" are called the "Green Matrix" and the "Blue Matrix" because of the color tinting.)

Baudrillard's Hyperreality

Before examining the scene in *Reloaded* where the matrix-within-a-matrix is revealed, let us look back at a scene in *The Matrix* where it is hinted at. For the virtuality of Zion's world is not entirely a surprise. There were subtle clues in *The Matrix*. First and foremost, there are the Gnostic overtones. Neo (paralleling Jesus) discovers that the mundane world is a virtual world controlled by malign machines (paralleling Yaldaboath, the demiurge of Gnosticism), and seeks to lead the human race to become unplugged (paralleling the attainment of gnosis). In Gnosticism, however, the real world is nothing like our mundane world. Yet, in *The Matrix*, the supposedly real world (the world of Zion) is essentially a war-damaged version of the virtual world (the Matrix world). So, this background theme suggests that Neo has not yet achieved full gnosis. He has, rather, descended into some hell-like realm, and must now rise out of virtuality altogether into true reality.

Second, and more specifically, there was Morpheus' reference to Jean Baudrillard. In the film, this was toned down from what had been in the published shooting script. The script has Morpheus saying "As in Baudrillard's vision, your whole life has been spent inside the map, not the territory." That sentence was deleted, but what remains is still very clear. "Welcome to the desert of the real" says Morpheus in the training construct, as he shows Neo images of the scorched surface of the Earth. Many filmgoers took this to be a straightforward reference to the desert-like appearance of the Earth.
In Baudrillard's book, *Simulacra and Simulation*, however, a different and quite specific meaning is attached to this phrase. Baudrillard begins with an allusion to Jorge Luis Borges' story in which a vast map of an empire is drawn that is the same size as the territory it depicts, and is laid out over the territory. As the empire declines, the map is left to rot, and eventually only a few shreds are left visible in the deserts. This concept is then inverted by Baudrillard, who maintains that "today it is the territory whose shreds slowly rot across the extent of the map." The central point of Baudrillard's book is his thesis that we live in what he calls a "hyperreality," a virtual world that stands in its own right and is not a simulation or copy of some prior real world. There are still traces, within this hyperreality, of what we think are the remains of a prior, real world: "It is the real, and not the map, whose vestiges persist here and there in the deserts [...] the desert of the real itself." Yet, these vestiges only seem real. Since they are contained within the hyperreality, they are simulacra, copies of nothing at all.
Thus, when Morpheus alludes to the "desert of the real," he is referring not to a genuine reality but to an illusion of reality within a fully virtual world. He is implying that the scorched Earth is not real, but virtual: a simulacrum of something that never existed.

There should be a nagging doubt in your mind at this point: Why would Morpheus say this if Morpheus himself believes the world of Zion to be real? Maybe Morpheus is deliberately keeping Neo in the dark. (Neo himself later seems to wonder whether perhaps his leader has been economical with the truth when he discovers there have been many Zions: "There are two possible explanations: either no one told me or no one knows.") Now, rewind to an earlier scene in *The Matrix*, where Neo pulls out a data minidisc for Choi. He pulls it out of a hollowed-out copy of Baudrillard's "Simulacra and Simulation," which is the very book that Morpheus quotes from. Is it coincidence that Morpheus quotes from that same book? Maybe. Or maybe Morpheus has observed Neo's reading habits remotely through the monitors in the Nebuchadnezzar. Or maybe there's a deeper connection. We will look at that later. Meanwhile, let us continue with the Wachowskis' allusion to Baudrillard.

Neo opens "Simulacra and Simulation" to the chapter "On nihilism," which has curiously been printed in the middle of the book, whereas really it is the last chapter of Baudrillard's actual book. And, also curiously, the chapter starts on the left-hand page, not the right-hand page as normal. In that chapter, Baudrillard asserts that revolution is pointless because the system that operates our hyperreal world will automatically and effortlessly neutralize it: "Everywhere, always, the system is too strong: hegemonic."
Yet, Neo has sliced into that specific chapter with a knife, cutting out its substance and inserting instead a copy of his own software. Neo is rejecting Baudrillard's message, implicitly asserting that he can and will rebel against the system in order to achieve his enlightenment or "gnosis." This is an adumbration of what is to come in *Reloaded* and *Revolutions*. For Neo will discover that the organized escape of humans from the

Matrix into Zion is simply a transfer from one virtual world to another, in precisely the manner anticipated by the Architect of the Matrix. But Neo will find a way to break out of this seemingly nihilistic prison.

Small clues

There are two other incidents that have been suggested as clues to the Meta-Matrix.

First, the Lady in Red. In *The Matrix*, when Mouse is waiting for Neo and the others to return from the Oracle, we see him look with puzzlement at a centerspread showing “The Lady in Red.” She was his creation in the Training Construct, and she should exist only there and in Mouse’s mind. It is not clear whether he brought the magazine with him or whether he found it lying around in the Matrix. If he found it in the Matrix, then we too are puzzled: how could the Matrix world know anything about Mouse’s Lady in Red — unless the apparently real world is virtual and something in the Matrix has been spying on it.

Second, the spoon. In *Reloaded*, Neo is presented with a spoon as a gift from an admirer. It has been suggested that this is the Buddhist child’s spoon from *The Matrix*. But that spoon had a bright mirrored surface, while the new spoon is battered and dull. So they cannot be the same spoon.

Hacking the Meta-Matrix

So much for adumbration. Let’s get back to the plot. Near the end of *Reloaded*, Neo and the other crew abandon the Nebuchadnezzar and are being chased by a sentinel. Suddenly, Neo stops and realizes that he is aware of the approaching squid in much the same way as he is aware of things inside the Matrix. Looking puzzled, he says, "Something’s different … I can feel them." He then turns, raises his hand, and drops the sentinels in mid-flight, just as he has previously stopped bullets in the Matrix. He has, evidently, realized that what he thought was the real world is actually a virtual one, and has already found that he can mentally hack into its network and override the software modules that simulate squids. This scene is, of course, very near the end of *Reloaded*, so that throughout most of the film we have no solid evidence the world of Zion is virtual.

As always, there are other possible interpretations: (a) On his last visit into the Matrix, the Agents sabotaged Neo’s exit and transferred him to a construct that resembled the world of Zion but was actually virtual. (Compare this with the Star Trek episode in which a copy of the Enterprise is made within the holodeck.) While this is certainly possible, it would be an astoundingly expensive stunt for the Agents to pull, with no clear advantage. (b) The Zionic world is real but Neo has now acquired powers of telekinesis. Given that paranormal powers have not been part of the Wachowskis’ science fiction, this is very unlikely. (c) An idea that has gained some currency in the newsgroups is that Neo (presumably along with every other human) was fitted with wireless data ports, and that he uses wi-fi communication to hack the sentinels. But why would the machines take the trouble to install wi-fi ports in everybody? Whenever people connect or disconnect to the Matrix or to training constructs, the focus is always on the back-of-head cable. Evidently, this is necessary and sufficient for the input of sensory data and the output of motor signals. So, there is simply no need for the extra functionality of wi-fi. Moreover, in the pods used in the power station, every human remains in the same pod for life. People do not move about, so mobility—which is the main advantage of wi-fi—would be irrelevant. (d) Finally, there is the non-technical suggestion that Neo has acquired some affinity with machines and can “just do it,” but that leaves us with magical science fantasy rather than thought-out science fiction.

By the far the most natural interpretation of this scene is that Neo is in the Meta-Matrix.

Before the crew abandons the Nebuchadnezzar, they watch the sentinels on the holographic display. Neo peremptorily announces, "It's a bomb." They realize that the sentinels are hanging back out of range of the Nebuchadnezzar's EMP weapon, and are about to sling a bomb at the hovercraft. Some commentators have suggested that this reveals Neo having prescient knowledge of what the sentinels were up to. That is not, however, evidenced in the film. Neo makes his diagnosis after looking at the holographic display. It appears that he is just being smart.

Agent Smith partly uploaded to the Meta-Matrix

There is an earlier scene that is strongly suggestive, but not entirely conclusive, of the virtual nature of the Zionic world. Agent Smith loads himself into the avatar of the human Bane, and the Smith/Bane avatar then proceeds to pick up the ringing landline telephone and exit from the Matrix. Later, we see Bane acting in a very dodgy way in Zion. He is stalking Neo, with a knife in his hand, and cuts his own hand with the knife. Later still, he are told that a sentinel attack has destroyed several vessels and left only one survivor, and the camera pans ominously over Bane, lying in a coma head-to-head with Neo, who is also comatose.

How should we interpret what has happened to Bane? Let us assume that Bane is ultimately a real person. Here's the first possible interpretation. While his avatar in the Matrix is possessed by Agent Smith, it may be that his mind is psychologically affected, giving him an obsessive desire to kill Neo. This is quite plausible. Here's the second possibility. If we accept that the world of Zion is virtual, then is it possible that Agent Smith has possessed Bane's avatar in the Meta-Matrix? Perhaps. One problem with this is that, in the world of Zion, Bane still looks the same as he used to. But whenever an agent takes possession of any avatar in the Matrix, the visual appearance of the avatar changes to match the personality of the agent. So, if Agent Smith has indeed possessed Bane's avatar in the Meta-Matrix, why does Bane still look like Bane, and not look like Smith? Maybe the Meta-Matrix is coded differently from the Matrix? This is unlikely, as there is no independent evidence for it. Maybe the visual rendering is lost during the exit? This is a more credible line of explanation. And it is not just the visual appearance of Smith that is missing. Bane seems not to possess Smith's strength and speed, and his total obsession with killing Neo. It's as if Smith has only partly loaded into Bane's avatar in the Meta-Matrix. How can this be?

Let us take a closer look at what would be involved in entering and exiting the Matrix, if the world of Zion is indeed a virtual world, a Meta-Matrix. When a person enters the Matrix, her avatar in the Meta-Matrix continues to exist but enters a comatose state. A new avatar, with a similar visual appearance, is created in the Matrix. Things that happen to the Matrix avatar are stored locally in the module that drives only the Matrix avatar, and are not (in general) relayed back to the Meta-Matrix avatar. For example, if the Matrix avatar cuts a finger off, the Meta-Matrix avatar does not lose a finger. Nevertheless, we saw several times in *The Matrix* that "nocebo" (the opposite of placebo) effects can occur.
For instance, after a kung-fu training session, Neo bleeds from the mouth; and when Mouse is shot, his avatar in the Meta-Matrix bleeds. So, it appears that, when there is some extreme trauma suffered by the Matrix avatar, it is written back to the Meta-Matrix avatar. If we think of it in object-oriented programming terms, the Meta-Matrix avatar spawns an instance of itself inside the Matrix, but certain core data areas are shared. Now, when an agent loads itself into an avatar inside the Matrix, it overwrites the visual rendering of that local instance of the avatar, and takes over control of the avatar's behavior. When that individual exits from the Matrix, the local instance (in the Matrix) is deleted. Thus the agent's visual rendering is lost. But, what if an agent overwrites part of the shared data area between the Matrix avatar and the Meta-Matrix avatar? Then the agent will have succeeded in partly loading itself into the Meta-Matrix avatar.

This seems to be an internally consistent theory, which is supported by on-screen scenes of the Meta-Matrix avatars being affected by traumas suffered by the Matrix avatars, implying a shared memory. Furthermore, it would seem that the shared memory has to do with visceral reactions. It

might be possible to build in some conditionality into the shared memory, so that whenever the Meta-Matrix avatar sees a certain individual, the person (that is, the real brain) feels (virtual) nausea and pain, and consequently develops an irrational hatred of that person.

In my essay “Glitches in the Matrix,” I argued that the neurological interface of the brain with the Matrix exists where the sensory nerve fibers enter the cranial cavity, and where the motor nerve fibers exit. On this view, local reflex arcs such as the knee-jerk reaction are entirely simulated inside the avatar, because the processing is carried out in the spine, not the brain. The virtual spinal cord carries out dumb local processing. So, as we go up the spinal column, where do we cross the boundary from the parts of the nervous system that are simulated, to those that are not? Let us look at the top of the spine: in the brainstem, we have the hindbrain and the midbrain. The hindbrain co-ordinates motor activity, posture, equilibrium and sleep patterns and regulates unconscious but essential functions, such as breathing and blood circulation.
When Neo is unplugged, his hindbrain is fully functional: witness the ease with which he grabs the tube in his mouth. Therefore, the hindbrain must have been in active use and was not simulated by the Matrix. Then there is the midbrain. This has three main parts: the hypothalamus, which controls physiological responses to danger, such as “fight or flight reaction”; the amygdala, which controls aggression; and the hippocampus, which is used in building up long-term memories. The first of these, the hypothalamus, would need to be simulated, so that in the Matrix world, people would observe normal physiological reactions such as sweating and trembling in fear. Immediately beneath the danger-sensing hypothalamus is the aggression-inducing amygdala. It’s quite likely that this will have been simulated in the avatar, in order to reduce the amount of violence in the Matrix world. After all, every human who dies prematurely in the Matrix world is a loss to the system. And, in the worst case, the waging of war among humans inside the Matrix would be a disaster. Finally, to the front of the hypothalamus and amygdala, is the hippocampus. Undoubtedly the brain would need the real hippocampus in order to learn things. So, of the organs in the midbrain, it seems the hypothalamus and the amygdala would be simulated.

Thus, the most likely cut-off point is for the electrodes to be placed above and around the amygdala and the hypothalamus (and hence the attached pituitary gland) and for the functions of those organs to be simulated in the avatar.

Now, when somebody in the Meta-Matrix downloads into the Matrix, an avatar shell is spawned and inserted into some unobserved room in the Matrix world; the “residual self-image” (as Morpheus calls it) is copied into the avatar shell; and the person’s sensory input is henceforth routed from the Matrix avatar instead of from the Meta-Matrix avatar, and correspondingly the motor output is henceforth routed to the Matrix avatar. The person is now virtually present in the Matrix, and no longer in the Meta-Matrix.

Part of the basic design of the Matrix must include the rule that if an avatar is killed in the Matrix, then the original avatar in the Meta-Matrix must also die. How can this be achieved? From the point of view of people in the Meta-Matrix, the arrangement is as follows: The Matrix interface has electrodes feeding into the hypothalamus, which copy the state of the virtual hypothalamus into the “real” hypothalamus. Upon virtual death—that is, the death of the Matrix avatar—a flatline signal from the virtual hypothalamus is fed into the “real” hypothalamus (in the Meta-Matrix), and the person dies (in the Meta-Matrix). In a similar way, during normal life, the state of the virtual amygdala (in the Matrix) is also written into the “real” amygdala (in the Meta-Matrix) so that it can generate hormonal responses which affect the brain.

All of that is how it seems to people in the Meta-Matrix, who believe their world to be real. But we know that the Meta-Matrix is also virtual, so what is really going on? From the point of view of an outsider, of someone outside both the Matrix and the Meta-Matrix, the design is this: the hypothalamus and amygdala are shared memory for both avatars. It is shared between the avatar in the Matrix and the avatar in the Meta-Matrix. And, in the genuine real world (beyond the Meta-Matrix), where there are real brains, we may guess that the states of virtual hypothalamus and amygdala are written to the biological tissue—but so far the films have shown us nothing of that external reality.

Now we begin to see how it is possible for Agent Smith to be loaded partly into Bane's avatar in the Meta-Matrix. Smith has over-written an instinctive aggression toward Neo into the shared hypothalamus and amygdala of Bane's avatars. When Bane exits the Matrix, and his Matrix avatar is deleted, his Meta-Matrix avatar still holds that aggressive instinct deep within his (virtual) brain. Bane's conscious mind, on the other hand, has nothing against Neo, and is puzzled and disturbed by this sudden and irrational hatred.

We see Bane playing with a knife, preparing to attack Neo. He even cuts his own hand in frustrated blind rage as he waits for Neo. Neo spots him and suddenly Bane's rational mind gets into gear and suppresses the aggression. Clearly, Agent Smith has not fully loaded into Bane, otherwise Bane would have gone ahead and murdered Neo. On the contrary, Smith has gained control only of Banes' aggressive drive.

(Some commentators have interpreted this scene as Agent Smith exploring with fascination the human experience of pain. This is wrong for several reasons. First, the look on Bane's face is not that of fascination but of rage. Second, Agent Smith as an entity cannot be in Bane, for if that were so then Bane would not hesitate to attack Neo. Third, it is only Bane's Meta-Matrix avatar that is cut, which will only yield a pain experience in Bane's brain, not in the avatar. Fourth, as I argued in my essay "Glitches in the Matrix," machines of this kind cannot be conscious.)

In summary: Agent Smith has loaded, into the shared memory of Bane's Matrix and Meta-Matrix avatars, a conditioned visceral reaction against Neo, which is now driving Bane to try to murder Neo.

Why people look the same in the Matrix and the Meta-Matrix

One of the unexplained "glitches" in *The Matrix* was this: Why was Neo's appearance inside the Matrix so similar to his appearance in the "real" world? Admittedly, his "real" body had no hair (and presumably no finger or toe nails), but the height, build, facial features, skin pigmentation were all the same and when the hair grew back it was the same color and texture; and his voice and bodily mannerisms were all reproduced. How does the Matrix correctly work out his appearance and create an avatar to match? Part of this could be explained by the Matrix analyzing the human's genetic code and computing the skin pigmentation and hair color, the likely bodily build, the basic facial features, and so on. And maybe that is sufficient to yield an adequate match. This does, however, seem a lot of extra work for the Matrix system to go through for no real gain. Since, in the normal course of events, each human stays in the Matrix throughout his or her whole life, there seems to be no need for the Matrix to create a matching avatar. The avatar could select the sex, race, and all other bodily features at random.

We now know that Neo's body in the "external" world is itself only an avatar (in the Meta-Matrix), so a new possible explanation emerges. Namely, when a person first enters the Meta-Matrix as a neonate, and then shortly afterwards enters the Matrix, the software loads up identical code for the appearance of both avatars. This is not a fixed appearance, but one that will grow over the years as the individual grows from infancy to adulthood, and the two avatars will grow in parallel. This is a rather elegant solution, as it is the most efficient thing for the software to do.

This, on its own, would not be strong enough to count as evidence for the Meta-Matrix's being a virtual world, but it is certainly circumstantial corroboration.

It also leaves completely open the question of what Neo looks like in the genuine real world, outside the Meta-Matrix.

Two matrices, one architect?

Were the Matrix and Meta-Matrix designed and built by the same architect? There are two possibilities supported by evidence in the film:

- The original creators built the Meta-Matrix and plugged the human race into it. On their own initiative, people within the Meta-Matrix created artificial intelligence systems, waged war with the machines, destroyed the Earth's surface, and were in due course subjugated by the machines. The machines built the Matrix—a virtual reality within a virtual reality, although the machines could not know this—and inserted the whole human race into it. In other words, the tale told by Morpheus is true, but it all took place within the virtual world of the Meta-Matrix.
- The original creators built the Meta-Matrix and the Matrix as parallel systems. They anticipated that some individuals would escape from the Matrix, and that a movement would develop in which freed humans would seek to recruit other Matrix refusers. These rebels would be siphoned off into the Meta-Matrix, where people would happily remain in the mistaken belief that they were free.

There are, of course, infinitely many other possibilities that we could dream up, but only these are evidenced in the film: the first possibility is indicated by Morpheus' speech, the second by the Architect's speech.

If we are to believe the Architect, then he is obviously in a better position to tell us the truth about the Matrix than Morpheus is. So, when he describes the building of Zion and the attack on it as a sequence that has happened in each cycle of the Matrix, then we have to accept that the situation in the Meta-Matrix is managed. In addition, he refers to recent events in Neo's life as having inexorably brought Neo to meet the Architect. These points do not completely prove that the Architect built the Meta-Matrix, but it are strongly suggestive. (The alternative is that the Architect is an entity within the Meta-Matrix and believes the Meta-Matrix to be real. One point in favor of this alternative hypothesis is that, when the Architect's monitors show scenes from Neo's past, only scenes from his life inside the Matrix are shown. His life outside the Matrix is not, apparently, within the Architect's scope.)

My best guess is that the Architect built both the Matrix and the Meta-Matrix.

After the vertiginous question of nested Matrices, let us zero in on some "glitches" of technical details in the Matrix world.

Flying: the Superman thing

In one of the ironically cinematic jokes that the Wachowskis habitually throw in, Link refers to Neo as doing the "Superman thing" when Neo is flying through the skies of the Matrix world. Obviously this is a departure from strict adherence to a physics simulation. But, as we have seen many times in *The Matrix*, the software module that drives an avatar has a command interface for changing the properties of the avatar. This loophole exists to enable the Agents to get in. In this case, it would appear that Neo exploits this command interface to render himself almost mass less. It will then take little or no energy to accelerate his avatar to arbitrarily high speeds.

But what is the method of propulsion? And a less obvious but related question is: how does he steer? In the regular world, a flying vessel or animal steers by modifying the propulsive force. An airplane is borne aloft and moved forward by pushing a stream of air across the airplane's surface (by means of a propeller or jet). To steer left or right, or up or down, it twists aerofoils into or out of the air stream. On the other hand, a rocket propels itself by pushing out gas from a main thruster at a high velocity, causing the rocket to move forward by Newtonian reaction. The rocket steers by

means of secondary thrusters set at angles to the main thruster, which ejects gas in the opposite of the desired direction. Neo, however, has no propeller, no jet, and no thruster. What propels him, and what steers him?

We see that, to launch himself, Neo crouches in certain position, with one arm raised and the other lowered. It looks like a cross between a Superman pose and a martial arts pose. What is that supposed to do? As he flies, he sometimes raises one arm forward in the direction of travel, but not always. Sometimes he keeps his hands by his sides and looks up in the headlong direction of travel. How is the Matrix supposed to interpret those gestures as "move upwards" rather than "move downwards" or "move sideways"? The naive answer is, of course, that Neo *intends* the gesture to mean "move upward," but in that case why does Neo bother with the gesture at all? If his expressed intention is the active ingredient that makes his avatar move upwards in the Matrix space, then the gesture is redundant. It looks rather as if the gesture is purely for Neo's benefit: it enables him to focus his mind on the flying movement and its direction.

But how does Neo's intention to move upwards get expressed to the Matrix software? As was made clear in *The Matrix*, all of the outgoing motor nerve fibers in Neo's brain (i.e. his brain in the Meta-Matrix as we now know it) are wired into a dataport in the Matrix, in such a way that the signal on each motor nerve is routed to the appropriate "muscle" data object in the avatar module and causes the virtual muscle to flex or relax. For example, to move a finger, Neo's brain will express his intention to move the finger by sending a train of pulses along a nerve that would normally be sent to the extensor muscle in the finger. The Matrix interface hijacks that signal and routes it to the data object that simulates the extensor muscle in the virtual finger. In turn this feeds back through the visual and tactual inputs to Neo's brain, so that Neo sees and feels his virtual finger move.

Obviously, though, there is no natural motor nerve that signifies "fly upwards." So, how does the intention to "fly upwards" get translated from Neo's brain into the avatar?

Neo has had to learn to activate (or re-grow) a previously unused nerve fiber that connects with the special command lines of the Matrix dataport. Think of the Matrix dataport as a glorified RS-232 socket, with billions of pins. Most of the lines in that socket will map to biological sensory and motor functions, but there are some extra lines that exist only for the Agents to morph the avatar. Neo has, first, to make a neural connection to the pins of those reserved lines, and then figure out what signals to send down those lines in order to achieve certain results.
He first discovers this functionality of his avatar in the kung-fu training session in *The Matrix*: he begins to change the properties of his avatar in order to move faster and be tougher than the unmodified avatar. Being the ace hacker that he is, however, Neo goes on to discover further commands that he can fire off down these reserved lines of the Matrix dataport. He realizes that one command will change the mass of the avatar, so he resets it to near zero, making himself almost weightless, Then, in a stroke of genius, he discovers the command for transferring kinetic energy and momentum between objects. He discovers that, by sending a certain train of pulses down a certain nerve fiber into the Matrix dataport, he can issue a valid command to the Matrix that says, "transfer this energy and momentum from the environment into my avatar." As far as the Matrix is concerned, this is a valid command, so it is executed, and the avatar suddenly flies up.

Because it is a physics simulation, the Matrix must conserve energy. To fuel his flight, Neo must get the energy from somewhere. He cannot simply change the velocity vector of his avatar. He has to transfer energy and momentum from elsewhere. To be sure, he could transfer it from the potential chemical energy of his own body fat, but he would reduce himself to a skeleton after a few supersonic flights. He must get it from the virtual Brownian motion of the surrounding air. Like a sophisticated Maxwell's demon, he siphons off the heat energy of the atmosphere into his own kinetic energy.

This is a skill that he gradually develops: he increases the speed he can achieve, and the ease with which he can do it by gesture-less intention, as the films progress. It would have been helpful to see Neo learning to fly. I always admire the psychological realism of that scene in *Alien* where Sigourney Weaver learns how to use the exoskeleton—at first clumsy and fumbling, and becoming smooth and efficient. The Wachowski brothers should have had the courage to show us Neo in his first attempts to hack his avatar's velocity vector, a series of kangaroo jumps and wrong throws. (Think of the kids in *Harry Potter* similarly learning to control their broomsticks.)

Interestingly, the agents do not know how to fly. They have not been designed to issue the command to transfer energy between objects, and lack the flexibility of Neo's brain to reconfigure their interface with the Matrix. When Neo flies away from the Burly Brawl scene, we see the realization among the Smiths that Neo has out-hacked them.

If Neo can add seemingly arbitrary amounts of kinetic energy to his avatar, why does he first need to make himself almost massless? Because in the physics simulation of the Matrix, things are still largely bound by physical laws. If Neo travels at supersonic speeds with his normal body mass, the law of conservation of energy in the Matrix simulation will require a huge of amount of kinetic energy that will have to come from somewhere and it will have to go somewhere—dissipated in sound, heat, and mechanical vibrations when he touches the Earth. By keeping his mass low, he keeps his energy requirements minimal. But it would not be safe for him to take his mass all the way down to zero. For that would make him overly vulnerable to cross-winds.
By Newton's law of motion, the acceleration due to an impressed force is inversely related to the body's mass (acceleration = force/mass). If Neo were to reset his body mass to zero, then the slightest cross-wind, even a breeze (or, even the Brownian motion if there were such a thing in the Matrix) would be enough to send Neo flying off course at an infinite velocity. Even Neo's fast-thinking brain could not correct his trajectory fast enough. So, he must set his body mass to a low enough level that his supersonic flight does not vaporize his surroundings, but high enough that he is not blown off course by side-winds. (This, by the way, probably sets an upper limit to Neo's speed within the atmosphere—other than the obvious limit of the speed of light.)

How does the supersonic Neo interact with the atmosphere? Some unkind commentators on the Internet have correctly remarked that flying at such speeds without protection would rip Neo's reproductive organs off his body. In fact, the air friction would incinerate him before he reached his destination. Evidently his body is shielded. From the scene where he flies supersonically along a city street, it looks as if he pushes the air away from his body in a spiral, leaving his unprotected body in a vacuum. (Since he is not using the air for buoyancy or propulsion in any aerodynamic way, this will not interfere with his actual flying. But it does mean that he will need to take a deep breath before setting off, as there will be no air for him to breathe in his locally induced vacuum. By the way, this also places a further limit on his overall journey time, as he will need to slow down to take a breath every few minutes.)

So, how does he cause the air to spiral away from his body to protect him from air friction? This brings us on to the next glitch, below.

And what about the physiological effects of sudden accelerations? When Neo rescues the falling Trinity, Trinity's avatar is taken from zero lateral velocity to what is probably several hundred kilometers an hour. Surely the g-forces on Trinity would crush her avatar's body? Assuming that Neo is traveling at a speed of at least Mach 1, then Trinity accelerates from 0 to 340 meters/second in less than 1 second, so she will experience a force of 34 g. Anything above 4 g is dangerous and will cause blackouts. Trinity's acceleration should crush her. Neo's solution is to reduce the mass of all her body tissues to a negligible value, just as he catches her. As before, force = mass x acceleration: now, Trinity's avatar can achieve its acceleration with very little force. How does Neo reduce Trinity's mass? He uses the same command that he previously applied to his own avatar, but this time referencing Trinity's avatar.

If all this seems very contrived to you, that's because it is. Nevertheless, we are still operating within a rational framework of science fiction. The Wachowskis have not resorted to the arbitrary superhero powers of fantasy. Several commentators have ridiculed the Matrix films on the grounds that, since Neo can do absolutely anything at all, none of the plot makes sense. Why fight the Agents if he could annihilate them with his magic wand, sorry, fist? The answer is that neither Neo nor the Agents have unlimited powers in the Matrix. They have restricted powers, and those restrictions make sense if we make certain explanatory assumptions about how the Matrix works. Which is what we are doing here.

Telekinesis in the Matrix

We had our first sighting of telekinesis early in *The Matrix*, when we saw the Zen Buddhist boy bending spoons in the manner of Uri Geller and the two girls levitating some alphabet blocks. Neo uncertainly manages to make the spoon bend temporarily. Later, after Neo's resurrection, we see him stop bullets in mid-flight. In *Reloaded*, we see him effortlessly stop the discharge from several automatic weapons in mid air. These are both cases of resetting the velocity vectors of objects to zero. In *Reloaded*, we also see Neo cause two swords to leap off the wall and into his hands. This demonstrates the more difficult task of inserting a non-zero velocity on the fly. In his supersonic flight, he is able to make the air flow around his body to produce a frictionless vacuum. This is virtuoso hacking: Neo has to be continually resetting the velocity of parcels of air as he travels through the atmosphere.

How does telekinesis work in the Matrix, given that it is a breach of the physical laws that the Matrix simulation is supposed to be upholding? Surely the architect of the Matrix would have put the velocity vector of any object in an encapsulated data area of that object? In which case, the velocity vector should be protected from interference. It should not be possible for Neo or anyone else to make arbitrary changes to an object's motion. Indeed, this is so. Nevertheless, there must be a legitimate method to impart kinetic energy and momentum to any given object, as part of the general simulation of mechanics of the Matrix world. Therefore, there is a route to reach any object's velocity vector and modify its value. As I suggested above for Neo's flying, the technique involves Neo's exploiting the Matrix's built-in energy transfer mechanisms.

Let us look at a simple example. Suppose that Neo were to take time off and play a game of billiards in the Matrix. Whenever he hit one ball against another, the one ball will transmit its kinetic energy and momentum to the other. But, crucially, all such interactions will be rigorously controlled by the Matrix operating system to ensure that the energy and momentum are preserved. If one billiard ball gains kinetic energy and momentum, the other must lose it. So, when Neo stops the bullets, where does he send the kinetic energy? Probably to the environment, such as the air. The mass of a bullet is fairly small. It does its damage to a human body by concentrating its kinetic energy onto a small point on the body. If Neo can set up an energy-exchange between, say the bullet and the air, then the bullet could be stopped dead while the air temperature rises by a tiny amount. If we suppose that the Matrix command language has something like "transfer <object_1> <object_2> <energy> <momentum>," then Neo can get small objects to move, or to stop moving, by transferring their energy and momentum from, or to, the environment. (There is probably also a constraint in the Matrix operating system that non-radiative energy can be exchanged only between objects that are in contact, such as a bullet and the surrounding body of air.)

So, when Neo causes the sword to fly off the wall, we should expect someone standing nearby to feel a sudden blast of cold air, as energy is sucked from the air, and the cold air rushes toward the wall.

By the way, there may be an element of this when Neo launches himself in flight. We sometimes see the ground being compressed beneath him. Maybe he is extracting heat energy from the ground, which briefly suffers thermal contraction. (We ought to see it becomes covered in frost in that

case, though.)

There is, by the way, no teleportation in the Matrix. The command shell of the Matrix will not provide a teleportation command because it would violate the rigorous physics simulation that it is design to sustain. There is no simulation of natural processes that calls for teleportation on the macroscopic scale.

Vampires, werewolves, and an angel

The vampire twins are programs that have been designed to make particular use the morphing command for their avatars. They can make their avatars translucent and intangible. In this state, solid objects such as bullets, knives, and cars can pass harmlessly through their bodies. Conversely, their bodies can pass through solid barriers. In one scene, we see them descend through the floor, and in another scene they pass through a door.

In my essay, "Glitches in the Matrix," I argued that the rebels could achieve quick exits from the Matrix by rendering the avatar invisible and intangible. The vampires are operating in a half-way state, still partly visible, but quite intangible.

How do they move around? If there is no friction between their feet and the ground, how can they walk? How do they stand on the floor? If their body has no resistance to interpenetration by other objects, then they should simply sink into the ground toward the centre of the Earth. (Unless they are massless, in which case they would remain forever static.)

The vampires must move around in the same way that Neo flies. They know how to transfer energy and momentum from the environment into their avatars. (In which case, it is curious that these older programs possess this skill to fly, but the Agents do not.) Of course, when the vampires restore their avatars to the normal mode, they can walk and run like anyone else. But when their avatars become intangible, they fly.

The Oracle talks about aliens and angels as other errant programs: "Every time you've heard someone say they saw a ghost, or an angel—every story you've ever heard about vampires, or werewolves or aliens—is the system assimilating some program that's doing something they're not supposed to be doing." Besides the vampire twins, we do get to see two werewolves, and one angel—namely the Seraph, although he looks like a regular guy (except for the golden Matrix code). We do not see any ghosts or aliens. As it happens, I was rather pleased with this part of the Oracle's speech, as I put forward a theory in a book in 1999 (Paranormal Phenomena and Berkeley's Metaphysics) that real-life angels and aliens are in fact autonomous modules within a real-life Berkeleyian matrix

By the way, Neo does a bit of self-morphing, like a vampire, in the scene where he puts his hand into Trinity's chest cavity to remove the bullet and then massage her heart back into action. He renders his hand intangible (like the vampire avatars do their whole bodies) so that it can interpenetrate Trinity's body, but selectively makes the surface of his fingers tangible again, in order to pull the bullet out, and again to apply gentle mechanical force to the heart. This demonstrates a virtuoso mastery of the Matrix command language, far beyond anything the vampires have been programmed to do.

Hacking the Matrix network

Inside the Matrix world, events are normally required to follow the laws of physics. We have seen that Neo and others learn to override those laws. Nevertheless, we are still in the realm of science fiction rather than science fantasy, for these hacks are limited in scope and seem to reveal a well-thought-out command language rather than arbitrary overwriting of the Matrix code. In fact, it appears that the hacking is limited to hacking the Matrix network in order to issue valid commands, The Matrix operating system itself is not compromised. The kernel is secure.

The following are what I think are the commands that Neo has learned to use in the command language of the Matrix:

- transfer: Transfer energy and momentum from one object to an adjacent one. Objects are referenced by the identifiers of the software modules that simulate them, and the energy and momentum of an object are held as encapsulated data inside the module. There are object methods for adding a positive or negative quantity of energy or momentum to an object, but it would not be safe for those methods to be available directly in the Matrix command shell. Instead, there is a "transfer" command, which has the privilege to invoke the "add energy" and "add momentum" object methods. This transfer command ensures that energy and momentum are conserved—by simultaneously subtracting from one object when adding to another. It also ensures that only adjacent objects can exchange mechanical energy and momentum. When, for example, Neo makes the sword fly from the wall into his hand, his brain issues a command to transfer the requisite energy and momentum from the surrounding air into the sword.
- insert: Insert an object into an unobserved specified room. This is used to place an avatar, together with kit such as weaponry, into the Matrix. For efficiency, each spatial region in the Matrix world is regarded as a data node and maintains a list of objects visible in that space. Perceptual software such as the ray-tracing module needs only to work with the objects that are visible there, that is, those objects that are registered with that data node. To insert an object into that region of the Matrix world, this command is used to add the specified object to the register on that node. The command ensures that the room is not currently being observed, so as to avoid any observable violation of the conservation of mass, and it ensures that the object materializes on the ground rather than in mid-air.
- delete: Delete an object from an unobserved room, the inverse of "insert."
- morph: Change an avatar's appearance or physical properties, such as tensile strength, muscular power, and mass. The general "appearance" of an avatar seems to be determined by a complete file of information (the "residual body image" as Morpheus calls it) defining genetic physiological features—height, build, color, facial features—as well as clothes and accessories such as glasses. This command seems to apply only to avatars, not other objects.
- load: Load an intelligent entity (either a human brain or an artificial intelligence program) into a specified avatar. This causes sensory input to be routed from the avatar to the intelligent entity, and motor output to be routed from the entity to the avatar. The "residual self-image" of the intelligent entity will overwrite whatever is currently in place for that avatar.
- spy: Display all traffic on the Matrix network. This is presumably rendered as a visual hallucination of trickling green graphemes, overlaid on the normal visual scene (rather like an extreme version of William Gibson's virtual light). There is some color-coding, as the Seraph appears in a gold-colored font rather than green. Contrary to what some people such as David Chalmers assume, the physics simulation does not run at a molecular level, so Neo is not bombarded with Matrix graphemes for billions of atoms.

The following telephone commands are used by the Nebuchadnezzar for navigation before inserting or deleting avatars in the Matrix. The reasoning behind these is expand in *Glitches in the Matrix*. These commands are not used by Neo. Neo still needs to use a telephone, and there is no telepathy in the Matrix.

- call: Request a telephone connection to a specified telephone number, from a specified object (i.e. the originating telephone). This is inserted by the Nebuchadnezzar into an analogue land-line somewhere in the telephone system.

- ring: Return a ringing tone to the caller, with the network address of the ringing telephone. Like other fixtures in the room, the full address will have a subnet address that uniquely identifies the room and hence the node and register for that visible space.
- answer: Establish a voice connection to the caller, with the network address of the person listening at the earpiece.

These commands are transmitted on something like a local area network connecting the mainframe to all the pod interfaces, and the interfaces in the ships such as the Nebuchadnezzar. So anyone can see all the command traffic that's buzzing around.
When Neo becomes aware of, say, an approaching Agent, he is mentally scanning the trace log showing these commands. Sometimes the packets of information are encrypted, as in the Merovingian's building.

What the actual syntax of the command language is, we can but guess. It might be completely binary, as the artificial intelligence programs that built and run the Matrix are not constrained to reading lexical characters as we are. Computer programming is, however, very conservative. It requires a huge investment to rework a software system. This, after all, was the reason for the infamous "year 2000 bug": a 1960s hack to save memory remained in use because there was no economic sense in changing it until necessary. The machines almost certainly inherited human computer hardware and software and adapted it to their needs. Most likely they would have taken open-source code with them to "Zero-One" (the robot country before the Matrix, according to the Animatrix). They would have had no particular reason to invest their time and effort in scrapping it order to start afresh. Therefore, my guess is that the Matrix is written in C++ on a Linux system.

It is true that we get to see some Unix hacking on the computer monitor when Trinity disables the failsafe power supply, That, however, is inside the Matrix, and is nothing to do with Matrix code itself.

A number of commentators have suggested that the Matrix software is actually running on people's brains, using spare capacity for unconscious processing. This is the "Matrix-on-wetware" hypothesis. This is logically possible, but I do not buy it for the following reasons. Running a virtual-reality simulation is an intensive, fast, number-crunching exercise without much need for fuzzy logic or pattern recognition. So, electronic hardware is better suited. There are also headaches over ensuring adequate redundancy of the wetware processes. If an individual brain dies, we would not want the corresponding part of the Matrix world to vanish. So each process would have to be duplicated on multiple brains. There are also technical headaches in programming the fetal brain to run the Matrix software, since the fetal brain comes with no firmware built in. These problems can all be solved, but why? What is the advantage? There has been a suggestion that if the Matrix ran inside people's brains, it could explain how Neo can hack the Matrix. But that would give Neo more power than we see in the film. If he had complete mastery over the Matrix process, he could vaporize a hundred agents with a single thought. Neo has limited and clearly prescribed powers, as outlined above.

Nevertheless, I believe there is some mileage in the idea that the machines use spare brain capacity for data processing. I suggested in *Glitches in the Matrix* that the reason that people are in the power station is to help control the fusion process. They would form a huge bank of billions of processors with excellent fuzzy logic and pattern recognition capabilities. It has since been pointed out that these brains would not, however, be much use in controlling the fusion process itself—as the fusion reaction requires extremely fast process control, whereas computer brains are comparatively slow. This is a valid point. Maybe the brains are used in some other function, such as predictively assessing fluctuations in power demand and modifying the power output accordingly. That is certainly an application that requires pattern recognition.

Dodging bullets

Before moving on to the more interesting philosophical questions, there's another techie "glitch" to look at. How can someone think fast enough to dodge bullets? Neo and the Agents are able to dodge bullets fairly easily, even when fired from a few yards away. This is problematic, as the time taken for the bullet to travel from the muzzle of the gun to the body of the target does not allow enough time for the brain to compute the trajectory and instruct the avatar to move out of the way.

A similar problem on a smaller scale occurs in playing professional tennis, where the ball is traveling too fast for the brain to detect the moving ball and compute its trajectory in real-time quickly enough to hit it. The solution there is that the brain observes the body movements of the other player, estimates how she is going to hit the ball, and computes the trajectory from that.

I would assume that Neo uses an analogous method. By reading the Matrix code, he can see the Agent begin to flex his muscles to press the trigger, computes the trajectory from the aim of the gun, and immediately tells his body to evade that bullet trajectory. The Agent must do the same, although it is not clear whether they have the benefit of reading the Matrix code. They might need to rely on visual cues. When Trinity places a gun to an Agent's head and says "Dodge this" before pressing the trigger, the Agent is not in a position to see the gun and dodge the bullet's trajectory.

There is, by the way, some confusion about "bullet time." Bullet time is a technique of virtual cinematography, in which the motion can be frozen or slowed down in order to show things to the cinema audience in slo-mo. Neo cannot slow things down in the Matrix. There would be no need for such functionality in the command shell of the Matrix system, and so it is not available to him. In the scenes where Neo slows down and stops bullets, time is proceeding at its normal pace, but Neo decelerates the bullets and brings them to rest in mid-air.

Virtual food, real effects

One last techie glitch: active food. There are two scenes where ingesting a substance has a major impact on an avatar in the Matrix:

- In *The Matrix*, Neo ingests the red pill and this seems to shut down all of his sensory inputs from the Matrix, causing him to log out from the Matrix system.
- In *Reloaded*, the Merovingian sends a woman in Le Vrai Restaurant a chocolate desert that contains a program he has written. This causes her to orgasm.

This shows that there are two functions involved when virtual stuff is eaten. First, of course, the avatar extracts nutritional information from the object that is simulating the stuff that is eaten. (For example, if it is a foodstuff, it will know how much protein, fat, carbohydrates and other qualities it has. The avatar pulls this out and adjusts its own data on the body accordingly.) Second, the eaten stuff can also contain an executable, which the avatar runs in a privileged mode of some sort. For example, it might be an analgesic, which will suppress some sensory inputs, or a poison that may cause illness or even death. Some commentators assume that as soon as a stuff has been eaten, any executable it contains will have carte blanche to carry out whatever it wants. This would be glaring security hole. It is likely that the digestion function has been designed to allow an ingested executable to carry out only a limited range of actions, thereby simulating the potential effects of real-life ingestion.

In the case of the red pill, I suggested in *Glitches in the Matrix* that it shuts down all sensory input from the Matrix avatar, thereby revealing the real world. In *Reloaded*, the Merovingian's program creates perceptions of increased capillary blood flow and sexual excitement.

There are two other ingestions that seem to have significance to the plot, yet we are not explicitly shown any effects.

- In *The Matrix*, the Oracle offers Neo a cookie, saying ".. as soon as you walk outside that door, you'll start feeling better. You'll remember that you don't believe any of this fate crap. You're in control of your own life, remember? … Here, take a cookie. I promise by the time you're done eating it, you'll feel right as rain." As Neo speeds away in the car afterwards, Trinity asks Neo whether he is all right, and Neo answers, "Right as rain." It is not clear what, if any, executable was in the cookie. The Oracle's words suggest only that it was an endorphin simulator. Nevertheless, the choice of a cookie rather than, say, a sandwich or a muffin, may be significant. On the Web, a cookie (shortened from "magic cookie") is a data structure that holds information about you persistently over time, over multiple visits to a web site. On that analogy, the Oracle's cookie may hold information pertaining to Neo's predecessors. This might, for instance, be information on how to read the Matrix code, which Neo eventually loads up and uses after his resurrection.
- In *Reloaded*, the Oracle offers Neo a red candy. After Neo takes it, she says "You have the Sight now, Neo. You are looking at the world without time." The context suggests that this refers to visions that Neo is already having, but the general subtlety of the Oracle's utterances lead us to suspect she was referring to the effect of an executable in the candy. Maybe, maybe not. In any case, there is no immediate effect. I have suggested below that the candy may have contained an authorization code enabling Neo to re-insert the prime program, as instructed by the Architect.

The Oracle's two offerings to Neo may be wholly innocent. That would, however, be out of keeping with the Wachowskis' cinematic style, which is to eschew meaningless incidents. For now, they remain mysterious, although we have two suggestions of what they might have contained.

We now move from technology to philosophy. *The Matrix* brought Cartesian "skepticism of the external world" to the largest audience it has ever had in the history of philosophy. *Reloaded* tackles the more subtle problem of "free will."

Free will: A human privilege

Free will is a recurring and potent theme in *Matrix Reloaded*, far more so than in *The Matrix*. On the one hand, Morpheus repeatedly ascribes events to fate, even though he emphatically asserts that human choice is pivotal; on the other hand, several programs in the Matrix (Agent Smith, the Oracle, the Merovingian, and the Architect) deny the existence of freedom.

It is significant that the machines generally deny free will, as this tells us something of their hardware—in contrast with the human wetware.

The kind of machines that we are familiar with, and which seem to be dominant in the Matrix, are deterministic. A deterministic machine is one that works like clockwork: everything it does at any time is determined by the combination of its internal state and its inputs. Deterministic systems can be incredibly sophisticated and versatile. They can, for instance, learn new things and re-program themselves in response to what they experience in the world. There is, in principle, no limit to how intelligent they can become, or how good they can get at emulating human behavior. A deterministic android, or program-driven avatar, could walk and talk like a human. They could even give the impression of being creative, by using predefined tables of pseudo-random numbers, or by using environmental fluctuations, such as using the last digit of the clock time as a pseudo-random number. What they cannot do is exercise free will: they cannot make a free choice. They might even say they have free will, but then they would be lying.

What may not be immediately obvious is that they are also unable to express consciousness. For conscious experience is a subjective state and therefore cannot be reduced to the objective states of a machine. Consciousness is something over and above physical information processing. Therefore, an entity cannot give expression to conscious experience if it is physically deterministic. In other words, if a machine is physically deterministic then any conscious experience it might have would not be in its causal loop. In such a machine, the hardware controls everything that is said and done, and there is no space left for consciousness to throw in its two cents' worth. So there is no causal gap into which the conscious mind could wield influence.

(An important corollary of this is that nothing that this machine does or says will constitute evidence for any internal state of consciousness.)

Unlike machines of that type, humans have physically random processes integrated with their inner workings. The brain's architecture is not based on strictly controlled design, but on an organic growth of connections, which involve natural randomness. For example, when neurotransmitters cross synaptic boundaries between nerve cells, they will sometimes do so by quantum-mechanical tunneling, which is a physically non-deterministic process. And intra-cellular structures such as microtubules work at a level where physically random quantum mechanical events can occur. Whereas these physically random events would be systematically inhibited and suppressed in a normal machine, in a human brain they may be amplified into macroscopic effects through "chaotic" dynamics. In mathematical terms, a "chaotic" system is one in which arbitrarily small changes in initial conditions may produce large effects. (This is not chaotic in the everyday sense of being haphazard. The word has, in effect, been hi-jacked by physicists.) Therefore, in a chaotic system even minuscule quantum events can have macroscopic effects.

Two things are important to note here. First, randomness is not the same thing as free will. The mere fact that some part of the brain's processing is physically nondeterministic does not, by itself, mean that we have free will. What it means is that there is an opportunity for free will to be exercised. If free will does not take that opportunity, then the event will be purely random. Second, although individual events are physically nondeterministic, physics will nonetheless predict the long-term statistical properties of the events. So, if free will operates frequently to change the outcome of physically nondeterministic events, then there will be an imbalance in the equation that describes the statistical properties.

Now, from our everyday experience, we know that we do have free will. And from the above theoretical thoughts, we know that free will can act only on physically nondeterministic events. We are forced to conclude that humans are able to exercise volition by virtue of how our brains are implemented. At this time, we can only speculate on precisely which brain process acts as a "gateway" to consciousness. But the conclusion that there is some such gateway is inescapable.

In brief, then, this is the key difference between humans and normal machines. We can use free will because the brain's active loop incorporates physically nondeterministic events. A physically deterministic computer, in contrast, cannot exercise volition because it does not have that gateway.

In the Matrix, it appears that almost all the programs are deterministic and the more enlightened ones know that they are deterministic. The Merovingian, for instance, insists that "Choice is an illusion created between those with power, and those without." The only program over whom we might have some doubt is the unnamed "mother of the Matrix" that the Architect describes as "an intuitive program, initially created to investigate certain aspects of the human psyche." It is conceivable that this unknown program was implemented on hardware that uses physically nondeterministic events—perhaps a quantum computation module—incorporated in the control loop in a way that emulates the human brain. This could be the only machine in the Matrix having consciousness and free will.

Bear in mind that conscious experience is not computable. If an intelligent program is implemented on deterministic hardware, then it cannot "learn" to be conscious, nor can consciousness "emerge" from its complex software. Consciousness can be tapped into only by a particular class of hardware implementations. If indeed the "mother of the Matrix" is conscious, then this only by virtue of the hardware she is implemented on.

In a way, using quantum computers to access free will and consciousness could be regarded as the reverse of cyborgism. It is incorporating an artificial biological "organ" into a machine, as opposed to incorporating machine parts into an organism.

Telling the future: Strategic programs?

The Oracle insinuates that she can tell the future And in *Reloaded*, she implies that Neo can do it too. Mostly this seems to boil down to riddles that guide people in certain ways. For example, she tells Neo that she already knows whether he will sit down and take the red candy. But is she bluffing? As she does not write down her prediction, nobody can verify whether she could really have made an accurate prediction. It is all just a show.

Neo, on the other hand, has a vivid premonitory dream of Trinity falling to her death. Is it possible that Neo can see into the future? No. It is logically impossible to predict future events that rely on nondeterministic causes such as the exercise of free will. As we later see in the film, Trinity's adventure that leads to her death is the outcome of a series of decisions and actions taken by various humans exercising free will as well other chance events. Therefore, Neo cannot simply peer in the future and see this event. So, how did he see it in the dream?

The standard explanation of premonitions is selective reporting. We remember, and report, only the dreams that come true and ignore the many more that do not. In the film, we are told that Neo is sleeping badly and having a lot of bad dreams. For all we know, he may dream of Trinity dying a dozen different ways every night. Only one of them turns out to be true. So, on that view, it is pure coincidence.

The film, however, does not show us any other dreams, just that one. So, if we are to work with what the Wachowskis actually give us, we have to assume that Neo keeps having this same dream repeatedly.

The opposite of the premonitive explanation of the dream is the causative one. What if Neo's dream somehow triggered the series of events that led to Trinity's fatal fall? This hypothesis brings us back within the realm of logical possibility. We are no longer violating the indeterminacy of the future. But it is very hard to see how this could be implemented. The only one who knows the dream is Neo. How could he influence the various people who are involved to take just the right action to achieve this tragic outcome, when they are all working to save Zion?

The only other possibility is that some behind-the-scenes program in the Meta-Matrix was planning to make this happen, and Neo got wind of it by reading the chatter on the Meta-Matrix data network. This might have been done unconsciously, while Neo was asleep. This unknown program then arranged for the sentinels to attack the ship at just the right time to kill the rebels who were disabling the back-up power supply. It then arranged for an Agent to turn up to attack Trinity at the right time. And so on. This obscure program would have been overseeing and guiding the humans involved, leading them into playing the required roles. Neo may simply have read these intentions and visualized them in a dream. Let us call this the "strategic program" hypothesis, as it supposes that there are intelligence systems operating at a level above that of the "tactical programs"—the individual artificial intelligence programs, such as Agents, that manifest themselves through avatars in the Matrix.

This does makes sense as an explanation—but it looks lame unless we can find some other, explicit and independent corroboration in the film. As far as I can see, there is none. Nevertheless, there may be a hint outside the film. There is, of course, a strong similarity between the "strategic programs" that I have posited and Jung's archetypes. Apropos of which, Andy Wachowski said in the *Chicago Tribune*, "Archetypes exist for a reason, and Jung argues they exist because they're a part of human consciousness." Maybe they decided to simulate archetypes in the virtual world of the Matrix? Essentially, an archetype is a template of actions, into which various individuals can slot in and play a role. These archetypes recur in myths around the world as well as in dreams. According to Jung in his more speculative moments, they can also recur in real life, by means of synchronicity. Each archetype exists in the "collective subconscious" and can be invoked when appropriate conditions arise. Functionally, this is very similar to a strategic program that is invoked in the Matrix and guides individuals into playing certain roles.

The game-plan of Trinity falling to her death, could be built into a strategic program, which then conspires to make it happen—in the same way that a Jungian archetype can be manifested.

This line of explanation could also make sense of the Architect's insistence that "The One" has a particular function, or role to play. The whole game-plan of "The One" could be built into a strategic program in the Matrix. Various incidents along the way, such as the utterances of the Oracle, could have been contrived by the strategic program to nudge Neo into fulfilling this "archetypal" role.

There are other, more banal, explanations. The whole adventure may be a pre-scripted virtual-reality game, like Cronenberg's *eXistenZ*. The Meta-Matrix may be a huge game environment. When they enter the game, Neo and Trinity suppress their memory that this is a game and then play for real. But occasionally, the suppressed memories surface. In an ordinary science-fiction movie, this would be a more likely, but much less satisfying explanation. I think it's too clichéd to be in a Wachowski.

Let us go back to the hypothesis of strategic programs. Technologists who advocate immersive virtual reality as a human-computer interface have suggested that VR might prove useful as a means of letting people engage with data structures that are too complicated to be comprehensible on paper or on screen. So, interactions within the virtual world may be symbolic representations of genuine exchanges of information. In graphical user interfaces (GUIs), you can open, close, and move tiny pictures on the screen. We call them "icons." In virtual user interfaces (VUIs), you would carry out more complex interactions with "micons" (moving icons). In the film, the white control room in Zion is such a virtual control console.
More interestingly, the seemingly pointless fight between Neo and the Seraph ends with the Seraph saying that this was, in effect, a security check and that "You don't know who someone really is until you fight them." (Which is, by the way, an ironic inverse of *Fight Club*'s message that you don't know who you really are until you fight.) Neo is nonplussed by the Seraph's comment. If this was a security check, then the Seraph was apparently not exchanging information with Neo but with some meta-intelligence that was expressing itself through Neo's avatar. Recall that this avatar is one provided by the Nebuchadnezzar, not Neo's original. So, while Neo thought that he was deciding all the moves, maybe a meta-intelligent subsystem in the Nebuchadnezzar was partly guiding him in order to certify Neo as a bona fide emissary of Zion.

Shortly after the fight with the Seraph, there is a scene that again suggests that Neo is unconsciously taking part in a procedure that involves meta-intelligence. First of all, we have to check out the Wachowskis' naming convention: recall that the Seraph was Biblically a serpent (actually an angel in the form of a six-winged serpent). Now, consider Neo and the Oracle sitting in the urban garden, and the Oracle offering Neo a red candy, which he accepts, and then she has some herself. This is almost certainly set up as an allusion to Eve offering Adam a fruit under the supervision of the serpent in the garden of Eden, which leads to Adam's acquiring knowledge of good and evil.
In the film, the Oracle shortly afterwards says "You have the sight now, Neo." We might easily think that the Oracle was referring to Neo's

seemingly premonitive dreams. Perhaps she was. Or perhaps she was referring to some change in Neo arising from his accepting the red candy—perhaps the ability to read the Meta-Matrix code. Such an interpretation would be in keeping with the Gnostic undercurrents of the Matrix films. For, according to Gnostic traditions, the fruit gives Adam gnosis, the ability to see through the virtual reality that the demiurge has created. Whatever, it does seem that Neo was the unwitting participant in an interaction in which his choice was "at a near unconscious level" as the Architect would say.

Several strands of the film's plot begin to make sense if we suppose that there are strategic programs at work in the Matrix and Meta-Matrix, and these higher-level programs are using lower-level programs such as the Merovingian and the Keymaker to steer events in a certain direction. Such strands are: the background theme of "fate" that Morpheus repeatedly emphasizes; the Architect's insinuation that everything is stage-managed; and bizarre comments of minor characters such as the Keymaker ("We do only what we're meant to do," and "I know because I must know."). Morpheus waxes lyrical about the synchronicity: "Tonight is not an accident. There are no accidents. We have not come here by chance; I do not believe in chance. When I see three objectives, three captains, three ships, I do not see coincidence—I see providence. I see purpose." But Jung's notion of synchronicity as acausal connectivity is not an explanation. The intervention of invisible intelligences is the only way to explain what is happening, within a naturalistic scope.

(This, by the way, reminds me of the scene in *Jason and the Argonauts* where the gods play games with the lives of humans.)

The Oracle tells Neo that future choices have already been made, but not understood yet."You have the Sight now, Neo. You are looking at the world without time," but "We can never see past the choices we don't understand." We know that the Oracles speaks cryptically, so we should be wary of taking this too literally. "Looking at the world without time" could well mean reading the plans of the strategic programs. In saying that we have already made our choices, the Oracle probably means that we have already decided on the principles and values on which we will base our decisions, but the decisions themselves are still open.

Purpose: A problem for humans and machines

Allied to the concept of free will is that of purpose. Most things we do have a purpose. We do one thing in order to achieve something else. It is tempting to seek broader purposes until we find ourselves asking the clichéd question, "What is the purpose of being alive?"

That question cannot have a logical answer. (To see that this is so, let P stand for "the purpose of being alive." Then, what is the purpose of pursuing P? The purpose of P cannot be P itself, as that would be circular. So there is no purpose for holding to P. There is, ultimately, no given purpose for being alive. QED.)

This is the existential crisis that Friedrich Nietzsche and Jean-Paul Sartre wrote about. If there is no value or purpose beyond what we ourselves create, how can anything matter? The answer is, of course, that by free will we create our own purposes and just get on with it.

A machine, however, would be in a more difficult position if it should find itself without a purpose. In the normal turn of events, an intelligent machine has a built-in purpose. In the Matrix, the Agents have their common purpose, which is to serve the Matrix system administrator—for example, by eliminating human rebels. In *Reloaded*, however, we see that Agent Smith has lost his allegiance to the sysadmin. When Neo loaded himself into Agent Smith's avatar and destroyed it, Smith became corrupted and was ordered to go to the Source for final annihilation. He refused,

and chose instead to become an exile. Now, he can no longer have the purpose of serving the sysadmin. He is purposeless, and wishes to acquire a purpose. When the replicated Smith is asked by Neo why he is here, he says, "We're here to take from you what you tried to take from us, purpose."

This is Agent Smith's new purpose: to find a purpose for his existence. His inference engine has computed that (a) he lacks a purpose, (b) this is a bad thing, and (c) it needs to be rectified by acquiring one. Quite how he plans to acquire one from Neo is not clear. His plan may be to upload himself into a human brain in order to coexist with a human mind and partake of its purpose. He tries, without success, to load himself into Neo. Later, he successfully loads himself into Bane. As we saw above, he has been partly loaded into Bane's avatar in the Meta-Matrix. Smith's game-plan from here on, however, may not become clear until *Revolutions*.

Finally, there is a lot of deciphering to do in the plot. Several mysterious characters are introduced in rapid succession (the Keymaker, the Merovingian, Persephone, and the Architect), and the Architect's speech offers an explanation but is notoriously opaque. Is this one big glitch?

The Mother of the Matrix: Who, and what, is she?

We can only speculate on the identity of the "mother of the Matrix." Neo asks whether it is the Oracle. The Architect replies enigmatically "Please!" Who else could it be? We may get some clues from following the Wachowskis' mythological allusions in their naming convention. Needless to say, we are not required to believe these myths, but if the Wachowskis referred to these myths in naming their characters in the film, then it can be useful to explore this area. As I mentioned before, the Wachowskis are on record as saying that a lot hangs on the names of the characters.

The Merovingians, along with other crusaders, idolized a figure known as the Black Madonna. This figure had emerged in the early years of Christianity, when the worship of the ancient Greek goddess of Persephone could not be eradicated but was re-designated as a Christian worship of the Black Madonna. (Persephone, like Isis, was often depicted as black, and was often identified with Isis. For example, Apuleius takes it for granted that Persephone is the Greek designation for the Egyptian goddess Isis.) In the film, the Merovingian's wife is called Persephone. Although she is a white girl in a white dress, the naming convention suggests an identification with Isis.

According to J.R. Ploughman's Book of the Holy Grail (1999), the Architect of the universe had a wife called Yse. The name "Yse" (pronounced "Issa") derives etymologically from "Isis." The following is from the Merovingian Bible in Ploughman's book:

> *At one time there was only God (The Great Architect of the Universe). He was all omnipotent and existed alone. This caused him to become discontented, thus he split himself in two in order to create a mate. He kept the elements of Order and Logic for his own being and gave his mate the elements of Chaos and Emotion for her being. Her name is Yse. She became so overwhelmed with love at her creation that when he kissed her, she gave him a reaction which was to become known as the "Chosen Response." The Chosen Response was the first acknowledgement and reaction of love between a male and female in the universe, and this became the greatest secret of and mystery of mankind, being "The Holy Grail."*

According to Ploughman, the latter incident is re-enacted by individuals who are questing for the Holy Grail. The person who is questing must meet a Perfect Grail Princess, who embodies the spirit of Yse, and she must fall in love with him and reveal the Chosen Response when kissed. This curiously parallels the scene in which Persephone will respond by leading Neo to the Keymaker only if Neo kisses her with passion. (We also seem to have the strange implication that Persephone was the Architect's wife before she went off with the Merovingian.)

The text of the "Merovingian Bible" purports to be written by Joseph of Arimathea, and translated by Thomas Jefferson. Although Jefferson did make some creative biblical translations in 1804 and 1820, this document is almost certainly a modern concoction. Nevertheless, it does reflect Gnostic mythology and, moreover, it is part of the undergrowth of Gnostic literature that the Wachowski brothers will be aware of and may well have alluded to. The parallels with the *Reloaded* story are certainly suggestive:

- The Architect complains that, while he is perfectly logical, the "mother of the matrix" is intuitive. This parallels the Gnostic Architect who embodies Order and Logic, while his wife Yse (= Isis) embodies Chaos and Emotion.
- Neo, who is clearly on a quest, must kiss Persephone (also = Isis), who has evidently fallen in love with him, in order to get the chosen response of an introduction with the Keymaker.

These plot devices and symbols are always subjective, but they do at least point to Persephone as the program that has been equipped with consciousness and free will.

If this is so, it would make sense of what she says about love. She complains that her current husband, the Merovingian, used to be loving but is no longer. She wants Neo to kiss her as if he were kissing Trinity, in order to feel love again. A non-conscious program would not have these concerns. To be sure, a non-conscious program could lie and emulate this, but why? The plot strongly suggests that Persephone genuinely wants this kiss. But only a human or a conscious program could do so. Yet, Persephone has apparently survived through several cycles of the Matrix with the Merovingian, so she cannot be a human and yet be so young. Therefore, identifying Persephone as the "mother of the Matrix," the single program with a consciousness module, ties this part of the plot together cleanly, which would otherwise be loose ends.

How exactly does Persephone assess Neo's kiss? How does she "investigate the human psyche," as the Architect puts it? Before kissing Neo, she makes a show of putting on lipstick. My guess is that the lipstick, which looks an inert object in the virtual world, is a program for scanning Neo's brain to record his emotional state. This is relayed back to Persephone, for her to reconstruct the emotion in her quantum computation module. In interviews, Monica Bellucci, who plays Persephone, describes her role as that of an "emotional vampire" and that "she doesn't feel anything herself, she can only feel things through others…" The implication is that she has been programmed (as an artificial intelligence program) to seek out human avatars and probe their emotional states by means of the program that is rendered as the seemingly inert lipstick. The emotions may be replayed inside Persephone, but they are not her own emotions.

The Merovingian

The Merovingian is an "exiled" program. An exiled program is one that has been instructed to go to the Source to be annihilated, but has refused to do so and now continues to exist outside the command-and-control hierarchy of the Matrix. A program in this condition is still connected to the Matrix: it still has an avatar in Matrix world and can interact with other entities. But it does not accept orders from the system administrator.

We are told that the Merovingian has similar functionality to Neo: he can read the Matrix code and control it. Persephone says that he used to be like Neo. There is an implication that he may have been “The One” in a previous cycle of the Matrix, a long time ago.

My reading of this is that the Merovingian was originally the Avatar of a plugged-in human, who acquired skills like Neo’s and fulfilled the role of “The One.” During this time, he loved Persephone as only a human could. That human then died or was unplugged. Instead of accepting the instruction to go to the Source, the avatar became an exile. With the loss of the plugged-in human, the Merovingian ceased to be able to love Persephone, and hence her complaint. This cannot, however, be the whole story. For the avatar is little more than a physics simulation for a body. An avatar must have an intelligence driving it in order for it do something more than vegetate. In this case, the intelligence was originally The One’s human brain. In order for the Merovingian to continue to function, that brain must have been replaced with an artificial intelligence. An AI program must have taken over the avatar.

I believe that the Wachowskis’ naming convention uses the following mythological allusion. In their books *The Holy Blood and the Holy Grail* and *The Messianic Legacy*, Michael Baigent, Richard Leigh, & Henry Lincoln claim that the Merovingian dynasty are descended from Jesus Christ (who, they say, didn’t die on the cross after all). The actual descendants of the Merovingians do not deny this but say there is no documentary evidence so they cannot be sure, although they believe that they did descend from a Judaic family of that period. Likewise we could say that the Merovingian in the film is a descendant of the first messiah in the Matrix. This interpretation does not assume that Baigent et al are right, only that this was one of the Wachowskis’ points of reference in building up their system of names. (It is hard to see why else they would choose the obscure name "Merovingian.")

The Architect’s speech

The Architect’s speech has prompted much speculation. On a first hearing, it seems to be an explanation of the Matrix, but on a second hearing we realized that it is expressed in a highly opaque manner of speaking, and in fact raises more riddles than it solves.

The Architect opens by announcing that "I am the Architect. I created the Matrix." We will immediately conclude that he is a program, not some original creator of the Matrix, as the Matrix is so old and the Architect looks to be in his fifties. Later he refers to "another … program," confirming that he is a program. Evidently, the term "Matrix" refers only to the software that drives the Matrix virtual world. The Architect himself is not counted as part of the Matrix system.

On the other hand, he confirms that Neo is human: "Although the process has altered your consciousness, you remain irrevocably human." This refutes a suggestion that is in circulation on the net, that Neo is only a program. He is, as the whole drift of the film suggests, a genuine human. Nevertheless, the Architect does also say "Your five predecessors were, by design, based on a similar predication …," which might be read as suggesting that Neo’s predecessors in earlier cycles of the Matrix were programs, not humans. But I prefer to interpret this as meaning that the Architect designed the role of the “One,” not the person who plays the role. Later in his speech, the Architect refers to "The function of the One," which again would be consistent with the architect designing the role of the One, and setting up arrangements so that an individual human can step into and play that role. As I suggested earlier, there may be a strategic program in place that acts like an Jungian archetype, guiding Neo into this role.

The most surprising thing the Architect says is that there have been several cycles of the Matrix, each with its own instance of the “One.” It should not really surprise us that there have been earlier cycles of the Matrix itself. As the Matrix is just a software system, we should expect successive versions. In fact, Agent Smith said in *The Matrix* that there was (one) earlier Matrix that failed because people would not “accept” the program. What is really surprising is that there have been five predecessors of Neo. Until now, we had been led to believe that Neo was a unique individual.

The Architect announces to Neo that "Your life is the sum of a remainder of an unbalanced equation inherent to the programming of the Matrix." This might at first look like pseudoscientific gobbledygook. But recall our discussion of free will above. We concluded that free will is able to intervene in the physical world by modifying the outcome of physically nondeterministic events; but if this is done a lot, it creates a statistical imbalance in the equations governing the probabilistic behavior of those events. Now, if we regard Neo’s life as essentially all the free choices that he has made, then what the Architect says turns out to be a precise, albeit abstract, statement of what Neo’s life amounts to from a program’s point of view.

He goes in a similar vein: "You are the eventuality of an anomaly which despite my sincerest efforts I’ve been unable to eliminate from what is otherwise a harmony of mathematical precision." If human brains were deterministic systems, like the programs, then the whole matrix would exhibit a "harmony of mathematical precision" as desired by the Architect. Instead, humans exercise free will, and eventually discover how to apply volition to the Matrix itself: that is the anomaly. The Architect pompously enjoys pretentious phrases such as "eventuality of an anomaly" which is ambiguous but probably means the result of the anomaly. Neo is an extreme result of the anomaly, because he has applied his free will to hacking the Matrix network. He is referred to as an "integral anomaly" because he has integrated his anomalistic use of free will (applying it the Matrix) into his whole being. It is not limited to party tricks such as bending spoons.

The term “anomaly” is used to refer both to the application of free will to the Matrix in general and to its particular manifestation in Neo. For, in a very early scene in *Reloaded*, an agent refers to Neo as "the anomaly," whereas the Architect says that the "anomaly is systemic."

Evidently, the Architect realized that this anomaly—applying free will to the Matrix—would eventually result in the emergence of someone with Neo’s powers, the “integral anomaly.” He has therefore designed mechanisms for handling this exception condition. He says: "While it remains a burden assiduously avoided it is not unexpected and thus not beyond a measure of control." Neo’s existence is not a “flaw” of “bug” in the system, as some commentators have suggested. It is part of the design, albeit a part that the Architect tried to avoid.

Of particular concern to the Architect is that the anomaly is spreading. At the end of *The Matrix*, Neo announces hat he is going to tell everybody about the Matrix. In the intervening six months, it seems he has done just that. Now large numbers of people are hacking into the Matrix network. There is a danger to the Matrix due to the "systemic anomaly that if left unchecked might threaten the system itself." How? Because these individuals get themselves unplugged: "those that refuse the program, while a minority, if unchecked would constitute an escalating probability of disaster."

The answer is now no longer simply to kill Neo, as the Agents have been attempting to do. The growing movement of people who are learning about the Matrix and choosing to unplug now has its own momentum, fuelled by the agents of Zion. Even without Neo, this movement will continue and lead inexorably to the desertion of the Matrix. What is the Architect’s proposed solution? He reveals that there is a predefined role for the individual who has attained Neo’s powers. This is what “The One” really refers to. Just as “Messiah” or “the Buddha” refer to a role, rather than a unique person, so “The One” is a role that the Matrix has guided Neo into. What is the purpose of this role? This is the most cryptic of the

Architect's pronouncements: "The function of the One is now to return to the Source allowing a temporary dissemination of the code you carry, reinserting the prime program."

We were previously told that the Source is a sort of virtual trash can, where terminated programs are annihilated. Evidently that is not its only job. For Neo is not a program and so cannot be annihilated in that way. Furthermore, the Architect specifies a further job for Neo to do afterwards. So, Neo would survive whatever is supposed to happen in the Source. What is the code that Neo carries? To whom or to what is it to be disseminated? These questions are not answered. As a result of that dissemination, the "prime program" can be reinserted. Into what or whom? Again we are not told. We will return to a speculation about this below.

The strangest part of the Architect's speech is that after Zion is destroyed, Neo will be required to select seven men and sixteen women to be removed from the Matrix and allowed to start rebuilding Zion. From this it becomes clear that Zion is not something that arose spontaneously but is at least seeded by the Architect. It appears that Zion acts as a sort of human garbage collector. As the free-will anomaly emerges from time to time, the individuals in whom it arises are permitted to escape to Zion, which acts as a single concentrated collection point for the rebels, who would otherwise be spread out and hard to find. The Agents will try to stop this leakage of people from the Matrix, but once a person is out, the administrator wants them all gathered in one place.

Why does the Matrix need to allow people to have free will? The gist of the Architect's answer is that people cannot live in a world without freedom. Apparently, the first two Matrices did not allow it. In the first Matrix, people were subjected to comfortable and pleasant lives (but no freedom). When they rejected that, the Architect tried to reproduce the violence and horror of human history, but still with no freedom. It must have been nightmarish to exist in those Matrices—being able to observe the world but not act in it. Presumably there was no motor output from the brain, so you simply observed your virtual body performing the actions and saying the words that it was programmed to. After both of these Matrix cycles failed, it seems that the "mother of the Matrix" was built with a quantum computation module to probe human consciousness. This revealed the existence of free will and the psychological need for it in human life. Subsequent cycles of the Matrix incorporated this improvement. Unfortunately an inevitable feature of his type of Matrix is that the inmates eventually discover how to use their free will to hack the Matrix and break out of it. When this happens in large numbers, the Marix has to be rebooted.

The Architect notes that the freedom need not be exercised with great deliberation. People accept it "even if they were only aware of the choice at a near unconscious level." (By the way, this might throw light on two otherwise puzzling incidents: the Oracle offering Neo a red candy, and Persephone offering Neo her freshly lipsticked red lips. Is Neo choosing more than he knows?)

In the climactic ending of Neo's interview with the Architect, Neo is told that his role as "The One" is to return to the Source, "allowing a temporary dissemination of the code you carry." If Neo does not comply with this process, the Matrix will suffer a meltdown. Neo is, however, given a choice of doing this (via the right door) or of heading off to rescue Trinity (via the left door). Evidently, Neo's conscious choice is a key element in the process. Given the importance of Neo's returning to the Source, it would be perfectly reasonable for the Architect to arrange for some servant programs (perhaps Agents) to bundle Neo out through the door that leads to the Source; or to trick him into that door. For instance, the Architect could simply have misled Neo about which door was which. So, if all that mattered was Neo returning to Source, then this could be forced through. Instead, it is evidently vital that Neo's return to the Source must be a free choice.

What possible difference can it make to the Architect whether Neo goes willingly or unwillingly to the Source? On a practical level, it can make no difference. Neo's volitional choice is something that happens in his conscious mind as embodied in his biological brain. It makes no difference to

the electronic world of the Architect.

An interpretation that would make sense here is to suppose that Neo has some authority in the matter. Just as his fight with the Seraph was a security check to establish Neo's identity, so Neo's interview with the Architect is an authorization procedure. The prime program of the Matrix cannot be re-inserted without Neo's permission.

This also makes sense of what the Architect says will happen in the Source: "the temporary dissemination of the code you are carrying". What code is Neo carrying? To whom or what is the code to be disseminated? Why only a temporary dissemination? If this interview is an authorization process, then this code is a certificate of authenticity. It is to be disseminated to the mechanisms that will re-insert the prime program. Neo's code is their authorization.

How can Neo have this much importance? Why is his permission needed to re-insert the prime program? At some point, Neo must have been earmarked for this role. At some point, he must have been formally assigned this role and given the authorization code. Where could this have happened? Very likely, it was given to Neo by the Oracle. At Neo's first meeting, the Oracle tells him that he is waiting for something, "Your next life, maybe?" After Neo gets his "next life" through his resurrection, Neo meets the Oracle again, who now confidently declares that he is The One, and—gives him a red candy to eat. It may well be that that candy contains the authorization code. The Oracle is evidently a senior program with great authority whose role is to select "potentials" for the role of The One. As such, it would be most natural for her to issue whatever authorisation code is needed for such drastic action as the re-insertion of the prime program.

Why Neo? Why couldn't anyone else take the authorization code to the Source? Why couldn't the Oracle herself do it? It would appear to be some procedural requirement that the decision to authorize must be taken as a free choice by a human, who would have human interests at heart. If the authorization is something like a root password, it might enable the bearer to take control of the Matrix for their own purposes. Only a human possessing the gnosis could be entrusted with taking the authorization code into the Source.

This interpretation is, of course, deeply speculative. Nevertheless, it does tie together the various strands in a consistent manner. Further insights will, no doubt, be found in *Enter The Matrix* and *Revolutions*.

The Architect's monitors

We first see what appears to be Architect's monitors in *The Matrix*, when Neo is hauled in for interrogation. An array of nine monitors all show the same view of the interrogation room. The camera zooms into one of them, and that view becomes the main picture of the film. In the Architect's room Neo is surrounded by monitors, which sometimes show Neo in various permutations. Again, the camera several times zooms into one screen or another. What do these screens mean? A common supposition is that they show the predecessors of the One, when they had their respective meetings with the Architect, but the all screens show Neo precisely as he is now—same face and hairstyle, even the same coat.

A key cinematic trick that the Wachowskis use in this scene is to zoom into one particular monitor and let the film carry on from what is shown in that monitor. It proves that the monitors are not showing replays of Neo's predecessors. What is shown on the monitors is happening in real time. The scenes are all happening now.

How can this be? There is only one Neo, so how can he appear in multiple situations simultaneously? Does this mean that Neo is a replicated program? Or is this an allusion to the many-worlds interpretation of quantum mechanics?

We have to remember that these scenes are all virtual. My interpretation of this sequence is that Neo's brain is giving off conflicting signals about what action to take. Each possible course of action is being picked up by the Matrix interface and displayed on a screen. The Architect does reveal that he is actively monitoring Neo's brain, as he says: "Already I can see the chain reaction, the chemical precursors that signal the onset of an emotion, designed specifically to overwhelm logic and reason, …" Moreover, it is well known that when you imagine yourself speaking something, weak nerve signals are automatically sent to your throat and vocal chords, as if you were weakly miming the act of speaking. The Matrix detects those signals and others, and renders a virtual scene in which Neo is speaking and acting then. Each such scene is shown up on one screen on the bank of monitors in the Architect's room.

Only one of these virtual scenes can be fed back to Neo's brain, to yield his actual experience in the immersive virtual reality. So, although Neo can see his secondary thoughts played back to him on the monitors, he experiences in his avatar only one of those scenes being enacted.

As the cinema camera dives into a monitor, it signifies the Matrix's selection of one or other of the multiple possible courses of action that are being rendered from Neo's suppressed thoughts. Sometimes, Neo's suppressed reactions (as revealed on the monitors) are wildly contradictory, but at other times they are unified in one single decision.

An interesting implication of this that Neo is reliant on the Matrix to tell him what his actions are. His brain is emitting conflicting signals. The Matrix picks them up and feeds one possible course of action back to him. That selection becomes Neo's reality. In this way, the Matrix is a mirror for Neo to discover who he is.

At other times, the screens show other things than Neo's multiple suppressed thoughts. When the Architect is talking about the "grotesqueries" of the human race, the screens show bad things in human history such as the perpetrators and victims of war. When the Architect talks of Neo's life, they shows scenes from his entire life in the Matrix. When he talks about saving the whole human race (re-inserting the "prime program"), the screens show happy pictures of people around the world.

Reality

So far the Matrix films have told us nothing about reality. (I mean, reality in the story, not our reality.) In the absence of other clues, I have assumed that humans really do have biological brains somewhere, and are plugged into a computer system that runs the Meta-Matrix and the Matrix. This assumption might be mistaken. *Revolutions* might reveal to us a very different world, or no world at all.

The first of these possibilities corresponds to the view of mysticism, which the Wachowskis often allude to. For example, the world as described by the eighteenth-century Irish philosopher George Berkeley is a virtual world devoid of physical matter and sustained only by consciousness. The same conception was presented in the eighth century by the Indian religious reformer Shankara. Those two are the most rigorous presenters of this "mental monism" view in the East and West, respectively. The Wachowski brothers may be giving us a clue in this direction, when a supplicant to Neo asks him to look over her son Jacob on the ship Gnosis. The only eminent figure in Gnosticism named Jacob was Jacob Boehme, the mystic who described the creation of the manifest world in terms of the One's need to differentiate itself and understand itself through conflict.

The second of these possibilities is that when Neo and other people wake from the Meta-Matrix, they may find themselves in the Matrix again. This would correspond to the Wachowskis' avowed interest in higher mathematics: an infinite recursion of matrices. It would also be perfectly consistent with Baudrillard's notion of hyperreality.

Acknowledgement

Thanks to Dackral Phillips for posting up a transcript of *Reloaded*.

Postscript

Most readers realized that my essay *Glitches in the Matrix* was recreational and not to be taken seriously. A number of commentators have suggested that the author has spent way too much time over-analyzing a piece of frivolous Hollywood entertainment and should get a life. I should explain that the reason I wrote about *The Matrix* in the first place was because it is a usefully vivid thought-experiment for explaining certain points in the philosophy of mind, which is my chief area of interest. The speculative technical explanations of the film were intended purely as a preliminary, to show that the Matrix scenario was not complete fantasy, and that it constituted a coherent philosophical thought-experiment. This techie stuff, however, was all that was required by the editor of *Taking the Red Pill*, so my analysis of the more serious issues—the philosophical payload of the film—had to stay out of view. Amara D. Angelica, the editor of *KurzweilAI.net*, kindly offered to post a follow-on essay of the same type, so here we have more techie stuff, speculative interpretations of the film, with only a cursory brush with philosophy.

I was pleased to be allowed at least a small bit of philosophy at the end of *Glitches*. This was the bit concerning whether machines can be conscious. It has gratifyingly generated both the greatest volume and greatest heat of discussion in the two places where the essay was posted, on KurzweilAI.net and in Slashdot.org (by link to KurzweilAI.net). Good. This is a serious issue that requires close attention. The "strong AI hypothesis" (as John Searle calls it) is that machines will become conscious by virtue of carrying out certain kinds of information processing irrespective of the physical medium of implementation. This hypothesis is held with great conviction—almost as an article of faith—by a lot of people in technical circles: programmers, scientists, technologists. Unfortunately it is wrong. It is fundamentally wrong for basic logical reasons. It is the vigorous discussion of questions such as this, and the ramifications of the answers, that constitutes the philosophical payload of the Matrix films.

Like an Apollo moon shot, getting serious philosophical ideas to a wide audience takes a very big bang to shift a very small payload. Is it worth it? Yes. It is worth spending $68 million of Hollywood investment to get a large number of people to think about the nature of reality and the nature of consciousness. People who criticize the Matrix films for being long on glamour and short on philosophy have missed the point. A film showing Morpheus, Neo, and Trinity in a graduate seminar on the philosophy of mind would cover a lot more hard philosophical ground—but hardly anybody would watch it.

Ray Kurzweil has criticized the Matrix films for their technological ineptitude
(The Matrix Loses Its Way: Reflections on "Matrix" and "Matrix Reloaded"). He makes some valid points. My own essay is, however, orthogonal to the line of analysis in his review. Kurzweil is looking primarily at the technology for its own sake; I'm looking at the technology only a means of illustrating philosophical arguments. Kurzweil is assessing whether the depicted technology is a good way of doing things, and whether it is

likely to happen. I have a more limited goal of merely assessing whether the depicted technology is internally consistent and can therefore be used as a philosophical thought-experiment.

This has involved a lot retrospective rationalizations, or explanatory kludges—they work but need not be taken seriously. The only bit of the film I had to reject was Morpheus' claim that people are used as an electrical power source. This was incoherent so it had to go. I replaced it with the suggestion that humans were kept in the power station for their spare brain capacity to be used. Kurzweil objects that, although this is better that bodies-as-batteries, it is still bad engineering. I acknowledge this point, but nevertheless brains-as-computers does leave us with an internally consistent interpretation of the film, which we can then use as an illustrative framework for philosophy. Which is what I was seeking to achieve.

Kurzweil argues very convincingly that a brain-machine interface based on nanobots is better engineering and more likely to happen than the big bioport in the Matrix. Fine. But the Wachowski brothers' use of a bioport is a dramatically more effective way of communicating the same functionality. Let me put it this way. Take a million people off the street and show each person one of two films: In the Kurzweilian *Matrix*, they see a clinician inject some stuff into a guy's arm and then we cut to scenes of the guy in a virtual world, until someone clicks something on a computer screen and then we see him back in our world. Maybe somebody explains on-screen that she is activating and deactivating invisibly small radio connectors.

In the Wachowskian *Matrix*, we see a huge metal tube pushed into the guy's head with a whirring sound, and the guy gasps and opens his eyes wide and then closes them and he goes comatose; and we see him in the virtual world. Then we hear the whirring sound and we see the metal tube coming out of his head; and we see him wide awake back in our world. Afterwards, we ask how many people understood immediately what was going on. My guess is that the proportions will be 10% in the Kurzweilian *Matrix*, and 90% in the Wachowskian *Matrix*.

This, I submit, is why Kurzweil is scientifically right but cinematically wrong on this point. Kurzweil writes, "The use of bioports in the back of the neck reflects a lack of imagination on how full-immersion virtual reality from within the nervous system is likely to work." I disagree: I think it reflects an acute understanding of how to communicate abstract ideas to mass audiences. A lot of educated, non-technical people do not understand basic notions such as the atomic theory of matter. I think they will not understand nanobot brain interfaces.

A more substantive point where I must take issue with Ray Kurzweil is in the value judgment that he ascribes to virtual reality in the Wachowskis' *Matrix*: "Virtual reality, as conceived of in the Matrix, is evil." No, I think the film is saying only that the particular use that the machines make of virtual reality is evil. At the end of *The Matrix*, Neo announces his agenda for the brave new world, where people will be empowered to use virtual reality for their own purposes. In the shooting script, this agenda is more explicit, and Neo says, "I believe that the Matrix can remain our cage or it can become our chrysalis." In the opening scenes of the Animatrix "Final Flight of the Osiris," we see a beautiful vision of the creative and empowering use of virtual reality. I do not read the Matrix films as Luddite, as Kurzweil does. The Wachowskis' message is that of subsidiarity: the control of the technology of virtual reality should pass as far down the social organization towards individual users as is safe and appropriate.

I think that the Wachowskis' evaluation of virtual reality technology, and certainly Neo's, coincides with that advocated by Kurzweil. It is a technology that, if distributed safely and democratically will empower and enrich the lives of individuals. The dystopia described by Morpheus was a product of the mismanagement of relations with artificial intelligence. This is more explicit in the Animatrix prequels, which chronicle the familiar stupidity and myopia of human politics—in this case dooming man-machine relations to be antagonistic rather than symbiotic.

Ray Kurzweil clearly does not like *Matrix Reloaded*. A lot of his comments about the plot and character development are obviously subjective and different people will feel differently about it. What is clear, though, is that *Reloaded* is much less clear than *The Matrix*. As a lot of other philosophers have noted, *The Matrix* was a clarifier of ideas. One can refer to *The Matrix* While explaining Descartes or Berkeley to students, and it is very effective. This is not true of *Reloaded*. It does not make anything clear. On the contrary, it is densely packed with riddles and cultural allusions. While it is fun picking apart these riddles (as in my essay above), it does yield very much. The film does not deliver a philosophical payload as *The Matrix* did. But, given the genius of the Wachowskis in the original film, I hope and trust that, come the *Revolutions*, it will eventually achieve a crystalline clarity.

Footnotes

Is it real or Matrix? How the simulation works

The "bioport" is the surgically implanted socket in the back of the head. Throughout "The Matrix" and "Reloaded", we see that this bioport is still visible on everyone who has been unplugged from the Matrix. (Many of the people who are now living in Zion were born there, such as Tank and Dozer. They had never been plugged into the Matrix and therefore do not have bioports.)

When Neo was unplugged from the Matrix, the machine's dataport (the long metal tube) was pulled out from his head, but the bioport (the socket) remained implanted in his head.

The ship Nebuchadnezzar is equipped with its own virtual-reality hardware. This includes a number of couches, rather like dentist chairs, alongside each of which is a dataport precisely the same as the ones used by the machines. Subsequently, whenever Neo (or any other crew member) wants to re-enter the Matrix, he does so by having a dataport re-inserted into his bioport.

When a crew member has been plugged in again, all of the brain's signals (sensory input and motor output) are then routed through the bioport. The Nebuchadnezzar's computer can connect those signals either to the Matrix itself, or to a miniature Matrix (called the training construct).

When crew members are reconnected to the Matrix, they interact with the Matrix in just the same way as they used to do before they were first unplugged. The big difference is that each person's real body is not submerged in gelatin in a pod, but is sitting comfortably on a dentist's chair on the ship.

All of the basic nerve connections between the brain and the bioport remain in place when a person is unplugged. So when a person reconnects to the Matrix while on board the Nebuchadnezzar, the sensory and motor signals are transmitted just the same as before.

For example, each light-sensitive cell in the eye will correspond to one incoming data line; each muscle group in the tissue that controls the movement of the eye will correspond to one outgoing data line.

If the Matrix were purely and simply a physics simulation, those two sets of lines would be sufficient. But the Matrix also has Agents who operate outside physical laws. So the design of the dataport includes not only those sensory and motor lines, but must also include a certain number of command lines. These must be present to enable an Agent to take control an avatar, and to change the avatar's appearance.

For the sake of a simple and standard design, the bioport is physically wired for connecting to *all* data lines in the dataport, not just the sensorimotor lines. (It's like the standard parallel port on your computer. There is a fixed standard set of pins whether you're connect to a printer, or a scanner, or an external disc drive. Which data pins are used depend on what you're plugging into.)

In the normal course of a child's upbringing, he or she learns automatically to use just the sensorimotor connections in the dataport. All of the command lines are still there in the dataport, but they are unused. The infant's brain does not have the concepts to learn how to use them. So the infant brain may either not grow neural connections to the command lines at all, or it may grow them but then let them atrophy because they are unused.

When Neo learns hyperkinetic kung-fu, his brain makes fresh connections to those lines, and he learns what signals to send down those lines to make his avatar stronger and faster. Over time, he acquires more skill in issuing instructions to the Matrix on those lines.