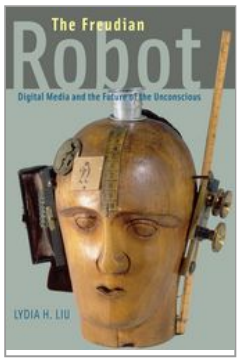

The Freudian Robot: Digital Media and the Future of the Unconscious
Lydia H. Liu

Print publication date: 2011
Print ISBN-13: 9780226486826
Published to Chicago Scholarship Online: February 2013
DOI: 10.7208/chicago/9780226486840.001.0001

The Freudian Robot

Lydia H. Liu

DOI:10.7208/chicago/9780226486840.003.0006

[–] Abstract and Keywords

This chapter examines the role of automata in psychoanalytic discourse by focusing on the problem of the Freudian uncanny. The accelerated advances in the technologies of animated pictures and automata have led many critics to return to Freud's original formulation. To grasp the full implications of automata for the understanding of the uncanny, it is necessary to reexamine the original point of contention between Freud and Ernst Jentsch and their contested readings. This work of reinterpretation aims to resituate the uncanny as the problem of the human–automata relationship, whereby the chapter clarifies the conceptual origins of the Freudian robot.

Keywords: Freudian robot, psychoanalytic discourse, technology, human–automata relationship

> What unfolds without fail before the reader's eyes is a kind of puppet theater in which real dolls or fake dolls, real and simulated life, are manipulated by a sovereign but capricious stage-setter. The net is tightly stretched, bowed, and tangled; the scenes are centered and dispersed; narratives are begun and left in suspension. Just as the reader thinks he is following some demonstration, he senses that the surface is cracking: the text slides a few roots under the ground while it allows others to be lofted in the air. What in one instance appears a figure of science seems later to resemble some type of fiction.
>
> Helene Cixous, “Fiction and Its Phantoms”

In his reflections on aggressiveness and narcissism, Lacan points to the original fracturing of the human psyche and says that at every instant man “constitutes his world by committing suicide.”[1] One of the psychological truths Lacan wanted to drive home in his study concerns the tension that marks the ego's relationship to space and its instinct of self-preservation. The extent to which the ego's instinct willingly gives way before the temptation to dominate space—examples he cited being “shock, fighter, parachute, and **(p.202)** commando trooper” (100)—says something about the malaise of a civilization for which Freud had invented a word: *Todestrieb*, or the death drive (instinct). Still, the reader may wonder why Freud thought it necessary to

adopt such a concept and oppose it to Eros or the life instinct as one of the most basic of all human instincts. And what is the death drive?

The death drive is perhaps one of the most contested of Freud's conjectures and has been related to the notion of entropy in the second law of thermodynamics. Ernest Jones has suggested that, by putting forward the death drive, Freud's aim was to establish a relationship between Fechner's principle of stability and the second law of thermodynamics.[2] In *Beyond the Pleasure Principle* and *Civilization and Its Discontents*, Freud tries to explain what he has observed as the innate propensity of all organisms toward equilibrium and destruction.[3] He argues that there is a tendency in human beings and in all living things toward their own death, toward a condition of regression to inorganic matter. This is *Todestrieb*. In thermodynamics, as we know, the tendency in a self-contained system to run down, to equilibrium, and to irreversibility is measured by “entropy,” a number that takes a quantity of heat divided by a temperature. The second law of thermodynamics stipulates the direction of the flow of energy in an isolated system, such as the flow of heat always from the hot to the cold body. Essentially, what it says is that entropy in a closed system cannot decrease unless external forces are exerted.

When Ernest Jones disputed Freud's hypothesis about the death drive, he did so precisely by invoking the second law of thermodynamics. Insofar as living beings are not closed systems, he argues, they can take energy from outside and acquire what Schrödinger called “negative entropy.” On that ground, Jones believed that Freud's attempt to bring entropy and the death drive together has been a failure (276–77). This view was shared by a good number of psychoanalysts of his time and by those who participated in a debate about the death drive and entropy in *International Journal of Psycho-Analysis* in 1931, but the debate did not seem to resolve the issue to everybody's satisfaction.[4] The relationship between the death drive and **(p. 203)** entropy has continued to engage the attention of critics from time to time,[5] and it is not by free association that in the early 1970s Anthony Wilden decided to launch his critique of Freud's entropy principle in the name of cybernetics.

Wilden claims that Freud's neuronal network foreshadowed the contemporary model of the brain and human beings as information-processing systems and that Freud almost anticipated cybernetics, except that he was hampered both by a bioenergetic model of the psyche and by his imperfect grasp of the differences between a closed system and an open system. In one of Freud's formulations, says Wilden, “a human being appears to be a neurotic steam-engine fluctuating between quiescence and runaway activity, with two conflicting kybernetai, Eros and Thanatos, at the controls, each haggling with the other over what is to be done with the daily delivery of coal.”[6] As if this is not enough to wake up Maxwell's Demon or the ghost of Norbert Wiener, Wilden carries his cybernetic tropes further by describing Freud's notion of the human as “a thermodynamic system condemned to the entropy of the ‘death instinct’ (the Freudian ‘principle of inertia’)” (124). He proposes that, in order to move beyond the entropy principle of classical physics adopted by Freud, one must introduce a clear methodological distinction between energy and information and that between closed and open systems (132). This sounds fine as far as the cybernetic view goes, but the historical contradiction is that Shannon did not shun the concept of entropy and in fact made it central to his own work on information theory. As we have seen in chapter 2, he openly borrowed that concept from the second law of thermodynamics and converted it from a measure of energy (heat) into a measure of information (“bits” or binary digits).

Wilden's criticism of Freud is anachronistic and flawed, although he is certainly justified in accentuating the role of cybernetics in the emerging world of digital media. Unlike Herbert Marcuse, Wilden paid attention to the rapid developments in digital technologies by working with Lacan and associating himself with Bateson and other American cyberneticians. This clearly sets him apart from the other Freudians of his time.[7] Marcuse's **(p.204)** understanding of psychoanalysis, in contrast, was very much immersed in bioenergetic discourse based on nineteenth-century physics, as if cybernetics and information theory had never happened.[8] The problem with Wilden is that he jumps over Shannon's appropriation of entropy to make a premature judgment about where Freud stands in relation to bioenergetics on the one hand and information theory on the other.

In retrospect, Shannon's translation of the thermodynamic concept of entropy into information theory in 1948 was a decisive moment in determining how the distinction between energy and information was to be drawn (and subsequently be endorsed by Wilden). This discursive shift in favor of information theory has eroded the earlier vitalist distinction between organic and inorganic—the very distinction invoked by Jones to contest Freud's concept of the death drive—and led to the formalist moment in molecular biology. Freud could not have anticipated all of this and it would not be fruitful to debate whether his formulation of the death drive conforms with or is comparable to the second law of thermodynamics. Strictly speaking, Shannon's own adaptation of entropy would have not passed the test of the second law of thermodynamics any more than Freud's.

The relationship between energy and information via the mediation of entropy is an enormously interesting problem for a historian of science.[9] Although this problem does not concern us here, we must confront the question of how the postvitalist blurring of the distinction between the organic and the inorganic complicates the human-machine relationship in digital media and, in particular, how it might impinge on our psychic relationship to automata, which had inspired Freud's famous essay on the uncanny.

Now let us consider the issue from a slightly different angle. In chapter 3, we encountered what might be called a simulation model of the death drive in Shannon's enigmatic automaton. According to Arthur C. Clarke, the Ultimate Machine was a sobering experience—if not pure epiphany—for those who came into contact with it.[10] Clarke happened to be visiting in **(p.205)** the United States to investigate the spectacular postwar developments in communication technology in the 1950s.[11] With a superb eye for detail, he wrote a few richly evocative paragraphs—excerpted in chapter 3—to describe the automaton he chanced to see when passing through Shannon's Bell Lab's office. When one presses the switch on the front of the small wooden casket, the lid arises slowly with a buzzing sound and a faux hand emerges from the inside. This hand reaches down to turn the switch off and retreats back into the casket. The lid then snaps shut and peace is restored. The whole thing does not take more than a few seconds. The embodiment of the digital principle of 0 and 1 by this machine is reflected not in its mechanics—in a sense, all robots are mechanical—but in the action of the faux hand that performs the on-and-off cycle.

If the reader will recall, Clarke found the experience emotionally disturbing and remarked, “[T]here was something unspeakably sinister about a machine that did nothing—absolutely nothing—except switch itself off.”[12] What psychoanalytical insight did Clarke glean from the Ultimate Machine? Is it related to the idea of the uncanny that Freud developed in response to Ernst Jentsch's reading of automata a century ago? How do we explain the fact that a faux hand that

appears and disappears with a simple feedback mechanism could exert such a powerful effect on the psyche? Is Shannon's Ultimate Machine a kind of Freudian robot?

In the previous chapter, we saw that both Lacan's reinterpretation of Freud and his development of the notion of the symbolic order were mediated through his innovative engagement with cybernetics, game theory, and information theory. This chapter reopens the Freudian problem of the uncanny and brings it to bear on the recent developments in automata, image making, and digital media. There are several important reasons for rethinking the uncanny in this light; we will examine only three. First, theorists of visual culture and postmodern art tend to fall back on Freud's notion of the uncanny as they try to explain how animated pictures and automata address our visual and cognitive systems, and they seem to find **(p.206)** Freud's “Das ‘Unheimliche’” (The “Uncanny”) particularly relevant—and even indispensable—to our modern aesthetic sensibility.[13] Second, if this tendency is indicative of the inherent difficulty of doing media analysis without psychoanalytical insights, we need to step back and ask what object of knowledge the psychoanalyst has sought to establish with the idea of the “uncanny” in the first place. This question can have significant bearings on the study of digital media because Freud's original contribution to that discussion was embroiled in a contested reading of the automaton in E. T. A. Hoffmann's “Der Sandmann” (The Sandman). Finally, and more important, the current work in automata, digital media, and AI engineering is beginning to throw fresh light on some aspects of Freud's explication of the uncanny that have heretofore remained obscure.

I aim to reestablish the seminal points of linkage between the automaton and the uncanny in the original context of Freud's engagement with Jentsch. Through a reinterpretation of Freud, Jentsch, and other theorists, this chapter will bring the relationship between Freud and digital media to light and further examine the Uncanny Valley research in the field of artificial intelligence in that light. The chapter will close with a consideration of the Neurotic Machine (K. M. Colby) and the Emotion Machine (Marvin Minsky)—computer simulations of human cognitive behavior of one kind or another—to reflect on the place of the Freudian robot in the human-machine ecology and its alternatives.

The Uncanny in the Automaton

The starting point of our inquiry is where Freud himself began in “The ‘Uncanny’”; namely, his vigorous repudiation of Ernst Jentsch's interpretation and, in particular, his rejection of Jentsch's argument of ambiguity or intellectual uncertainty as the source of the uncanny. This point of departure or indebtedness, which will be crucial to our unraveling of Freud's position on the subject of automata, is often slighted if not completely suppressed in the extensive commentaries that have accumulated around Freud's essay since its first publication in 1919. In the majority of the existing literature, the uncanny mysteriously figures the return of the repressed or, in Schelling's much quoted phrase, the uncanny is that which “ought to have remained secret and hidden but has come to light.”[14] This reading makes sense in the context of Freud's argument, but the unintended **(p.207)** consequence is that Jentsch's “Zur Psychologie des Unheimlichen” (1906) is seldom read to enlighten us on why Freud could not tolerate the suggestion of ambiguity and uncertainty as a source of the uncanny.[15] Bill Brown is certainly right to quip that “Jentsch has suffered the fate of being most familiar as a dismissive footnote.”[16] Instead of scrutinizing Freud's repudiation of Jentsch, most critics are content to mimic Freud's manner of introducing an argument through German or European etymologies while reaffirming the universal claim that *das Unheimliche* (the uncanny) is what becomes unconcealed as *das Heimliche* (the familiar).[17] Jentsch's essay was neither reprinted in German nor made available in English until 1997, and this curious fact has led Forbes Morlock to call the

situation “doubly uncanny,” commenting that “the repressed, however, always returns—the uncanny in ‘The Uncanny’ is about nothing else.”[18]

It is important to bear in mind that Jentsch, not Freud, was the first to link the phenomenon of the uncanny to automata by way of Hoffmann's “The Sandman.” Fascinated by the psychic power of automata, Jentsch noted the rise of the uncanny feeling and sought to explain it by emphasizing the role of *intellectual uncertainty between animate and inanimate*. Freud acknowledged this work as the starting point of his own inquiry but tried very hard to repudiate the argument of intellectual uncertainty. But his dismissal of Jentsch's thesis should by no means be taken at face value or be accepted at all until one is prepared to engage with the men's shared interest in the automaton as a simulation of the uncanny. The ways in which their interpretations converge or diverge on the question of *intellectual uncertainty between animate and inanimate* can yield some vital clues as to how the Freudian uncanny can be made relevant to the contemporary discussions of the psychic powers of image, medium, and automata in digital media. It is only by bringing Jentsch back into the discussion of the uncanny that we can reassess the stakes of Freud's own argument concerning animated figures and automata more clearly than before. In this **(p.208)** regard, Freud is as strangely indebted to Jentsch in developing the idea of the uncanny as he is to Sir James Frazer in the matter of the totem.[19]

Jentsch's “Zur Psychologie des Unheimlichen” (1906), translated as “On the Psychology of the Uncanny” (1997), sets the tone and initial direction of Freud's investigation of the relationship of aesthetics and the uncanny. This includes his numerous references to automata and fantastic literature. Jentsch speculates that of all the psychical uncertainties that may cause uncanny feelings to arise, one in particular is likely to develop into a regular, powerful, and general effect. This usually concerns the doubt as to whether an apparently living being is animate or not; and, conversely, it may involve the uncertainty about whether a lifeless object may not in fact be alive, as, for example, when a tree trunk is perceived to suddenly move and shows itself to be a giant snake or a primitive man experiences his first sight of a locomotive or a steamboat. The horrors people feel toward a dead body, especially a human corpse, a death's head, skeletons, and similar objects are likewise attributed by Jentsch to the thought of a latent animate state that lies close to objects that used to be alive. Amongst the other lifeless objects that fall into the same category of the uncanny, he mentions wax figures, panopticons, and panoramas and, above all, draws our attention to automatic toys, life-size automata that can perform complicated tasks like blowing trumpets and dancing, and dolls that can close and open their eyes by themselves. Yes, dolls and automata, etc. Among such objects should fall Shannon's suicidal Ultimate Machine as well as countless other cybernetic toys of our own time, all of which would aptly belong to Jentsch's catalogue of dolls and automata and supposedly inspire the feeling of the uncanny in the observer.

Anthony Vidler, who has derived the term “the architectural uncanny” from psychoanalysis, approaches the interaction of the mind and space in precisely this manner, for he grasps the aesthetic dimension of mental projection toward spatial configurations as that which “elides the boundaries of the real and the unreal in order to provoke a disturbing ambiguity, a slippage between waking and dreaming.”[20] This understanding of the uncanny **(p.209)** enacts a return of the repressed by reasserting Jentsch's view of ambiguity and intellectual uncertainty in the name of Freud.[21] As a matter of fact, Jentsch's return in the guise of Freud appears almost inevitable as can be seen in the works of a growing number of cultural critics who contemplate the social implications of the increased blurring of boundaries between human and machine. Vidler sums it up as follows: “[N]ow, the boundaries between organic and inorganic, blurred by

cybernetic and bio-technologies, seem less sharp; the body, itself invaded and reshaped by technology, invades and permeates the space outside, even as this space takes on dimensions that themselves confuse the inner and the outer, visually, mentally, and physically" (147). The bio- and techno-uncanny is evoked by Vidler to help make sense of our posthuman world of cyborgs who live with their prosthetic devices and technologies. This is where Jentsch, and Freud, become once again relevant.

In "On the Psychology of the Uncanny," Jentsch argued that the figure of the automaton traverses the boundaries of animate and inanimate and tends to provoke a certain kind of anxiety. This psychic phenomenon requires explanation and Jentsch gives it as follows:

> The mass that at first seemed completely lifeless suddenly reveals an inherent energy because of its movement. This energy can have a psychical or a mechanical origin. As long as the doubt as to the nature of the perceived movement lasts, and with it the obscurity of its cause, a feeling of terror persists in the person concerned. If, because of its methodical quality, the movement has shown its origin to be in an organic body, the state of things is thus explained, and then a feeling of concern for one's freedom from personal harm arises instead—which undoubtedly presupposes, however, a kind of intellectual mastery of the situation for the sake of all other forms of intensity. Conversely, the same emotion occurs when, as has been described, a wild man has his first sight of a locomotive or of a steamboat, for example, perhaps at night. The feeling of trepidation will here be very great, for as a consequence of the enigmatic autonomous movement and the regular noises of the machine, reminding him of human breath, the giant apparatus can easily impress the completely ignorant person as a living mass.[22]

(p.210) However, human minds are not equally affected by the uncanny. According to Jentsch, the head of a pillar or the figure in a painting can come alive in the minds of those who are delirious, intoxicated, or superstitious and who would address the figure, carry on a conversation with it, or attack it. In contrast to rational men, presumably including himself, whose intellectual maturity provides better psychological defense against such fancies, Jentsch believes that children, women, primitive men, and dreamers are particularly susceptible to such stirrings and to the danger of seeing spirits and ghosts. On the point of intellectual mastery, Freud is in agreement with Jentsch, for the former begins his own essay on the uncanny by confessing that it is a long time since he has experienced or heard anything that would give him an uncanny impression.[23]

The social drama of how women and the psychically vulnerable confront a clever machine or a simulation model of the human mind has been staged repeatedly in recent times by computer scientists who design neurotic machines or simulation computer programs either to model a theory of the mind or to advance the cause of artificial intelligence. The most famous of the early programs was written by MIT computer scientist Joseph Weizenbaum in 1964–66. His simulation program is named ELIZA—with a script called DOCTOR—and it requires those who interact with the machine to play the role of a patient as DOCTOR simulates the speech of a nondirective therapist. Having observed how people talked to the machine, Weizenbaum was surprised to find how quickly and how deeply people became emotionally involved with the computer. His own secretary had followed his work for many months and knew that it was a computer program, but she forgot all this when she began conversing with DOCTOR. "After only a few interchanges with it," recalls Weizenbaum, "she asked me to leave the room. Another time, I suggested I might rig the system so that I could examine all conversations anyone had with it, say, overnight. I was

promptly bombarded with accusations that what I proposed amounted to spying on people's most intimate thoughts."[24] Weizenbaum is not implying that only women are susceptible to such delusions, but the popularity of his story within the communities of computer science and artificial intelligence does reveal something about shared assumptions about gender and intelligence. **(p.211)** If the story sounds strangely familiar, it reflects back on the mental delusions that Jentsch tried to explain in the essay on the uncanny.

When the mental powers ascribed to children, women, primitive men, and dreamers are called upon to help stage the primal psychic drama of the uncanny, they effectively separate the investigator, whether Jentsch or Freud, from the object of investigation. This has been one of the originary social conditions of psychology and psychoanalysis acknowledged by their pioneers and practitioners. It seems that the psychic drama of social discrimination is what drew Bill Brown's attention when he tried to excavate Jentsch's analysis of animated objects to understand the American history of race, slavery, capitalism, and reified people and things. Brown argues that "[t]he animation of the lifeless object, in the case of the black collectible, reinstates Jentsch's argument, fully leavened by part of Freud's. For the point is not only that the inanimate comes to life but that the history of this ontological ambiguity—human or thing—is precisely what remains repressed within U.S. culture" (199). This observation implies an unmistakable grasp of Jentsch's original proposition—even if the latter's social statement is ironically downplayed by Brown—but the dynamic between Jentsch and Freud on the interpretation of Hoffmann's literary work is something to which we should pay closer attention, because it is precisely the ambiguities within the text of "The Sandman" that had given Freud the initial opportunity to repudiate Jentsch.

Jentsch has speculated that the uncanny is a semi-conscious projection of the self onto an object even as the object returns to terrify the self in the image of that self-projection. For this reason, human beings are not always capable of exorcising the spirits that have been fabricated out of their own heads and this inability produces the "feeling of being threatened by something unknown and incomprehensible that is just as enigmatic to the individual as his own psyche usually is" (14). This weakness is often exploited by poets and writers who aim to arouse the same uncanny effect in their readers. In storytelling, Jentsch suggests, one of the most reliable artistic devices for producing uncanny effects is to leave the reader in a state of uncertainty about whether he is dealing with a human being or an automaton. "In his works of fantasy," he adds, "E. T. A. Hoffmann has repeatedly made use of this psychological artifice with success" (13). Although the title of the Sandman story is omitted from his account, the reference leaves Freud in no doubt as to the focus of uncertainty Jentsch alludes to here, namely the doll Olympia. It bears pointing out that in another Hoffmann story, "Automata," mechanical dolls also raise troubling questions about **(p.212)** the uncanny or "images of living death or inanimate life."[25] Hoffmann's focus in this story is the psychic power of the automaton over human beings and the machine in question is a Talking Turk on exhibition. I would not rule out the possibility that Jentsch had this other story in mind as well, although Freud focuses his attention exclusively on the Sandman story in his repudiation of Jentsch.

As we know, Jentsch attributes the uncanny effects of the Sandman story to the artistic skill with which Hoffmann manipulates intellectual uncertainty in Nathanael and in the reader who identifies with him as the narrative unfolds: Is Olympia animate or inanimate? Interestingly, the word *unheimlich* appears in Hoffmann's story a number of times, often in conjunction with Olympia, which apparently constrained Jentsch's interpretation. For example, Nathanael's close

friend Sigismund uses this word to describe the peculiar look and gesture of the doll with whom Nathanael has fallen madly in love:

> It's very strange, however, that many of us have come to the same conclusion about Olympia. She seems to us—don't take this badly, my brother—strangely stiff and soulless. Her figure is symmetrical, so is her face, that's true enough, and if her eyes were not so completely devoid of life—the power of vision, I mean—she might be considered beautiful. Her step is peculiarly measured; all of her movements seem to stem from some kind of clockwork. Her playing and singing are unpleasantly perfect, being as lifeless as a music box; it is the same with her dancing. We found Olympia to be rather *unheimlich*, and we wanted to have nothing to do with her. She seems to us to be playing the part of a human being, and it's as if there really were something hidden behind all of this.[26] (my emphasis)

It is safe to assume that Jentsch took both his idea and his language from Hoffmann's narrative to make an argument about the hesitation or uncertainty between animate and inanimate beings as the cause of the *unheimlich*. But as Freud correctly observes, the uncertainty surrounding the identity of Olympia is gradually dispelled toward the end of the story, at which point both Nathanael and Hoffmann's reader are informed that she is an **(p.213)** automaton and that her clockwork was made by physics professor Spalanzani and her eyes were put in by Coppola, from whom Nathanael bought his pair of telescopes with which he has spied on her. On this ground, Freud rejects Jentsch's argument and decides to look elsewhere for an answer to the sources of the *unheimlich*.

Freud, who cannot tolerate the idea of intellectual uncertainty as an explanation, introduces his alternative reading by displacing the automaton as the problem of the uncanny with something else, namely, Nathanael's ocular anxiety about the Sandman.[27] The move to decouple the uncanny and automata is made perfectly clear in the following statement: “[U]ncertainty whether an object is living or inanimate, which admittedly applied to the doll Olympia, is quite irrelevant in connection with this other, more striking instance of uncanniness.”[28] What is this other instance of which Freud speaks? He calls it castration anxiety and proceeds to demonstrate that Nathanael's fear of the loss of eyesight can be read as a symptom or substitute for his repressed castration anxiety. This fear has been instilled in him as a child and is embodied by the Sandman, Coppelius, and the eye-glass peddler, Coppola. The multiple interplay of doubles or the doppelgänger in Hoffmann's tale from Nathanael/Olympia and Spalanzani/Coppola to Coppelius/Coppola and Klara/Olympia leaves ample room for Freud to develop a parallel narrative that seems to make good sense from the standpoint of psychoanalysis.

In “The Uncanniness of the Ordinary,” Stanley Cavell tries to wrestle the problem of the uncanny away from the hegemony of Oedipus to refocus our attention on “an uncertainty in our ability to distinguish the animate from the inanimate.” If this critique appears to echo Jentsch's earlier argument, it has not led Cavell back to Jentsch's essay; he doesn't come to grips with the world of dolls and automata that had fascinated Jentsch and Freud alike. Instead, Cavell confines himself to speculations about “the sense of the human as inherently strange, say unstable, its quotidian as forever fantastic,”[29] instead of considering, as suggested by Bill Brown, that “the repression at work may be the repression of the unhuman object-world itself, which psychoanalysis compulsively translates into the human.”[30]

(p.214) W. J. T. Mitchell proposes that the notion of the living image—which is more than a metaphor or an archaism—deserve our attention because it requires thinking through the category of life itself. The latter poses a semantic rectangle that governs its dialectics as follows:

living	dead
inanimate	undead[31]

This would have appealed to Jentsch, whose own study of the uncanny obeys similarly construed logical opposites: the dead object that was once alive, and the inanimate object that never was alive. In Mitchell's formulation, there is also the negation of the negation, or the return and arrival of life in the nonliving substance where human beings impersonate the inanimate figures of painting and sculpture. Presumably, the “undead” is where Hoffmann's automaton truly belongs and, indeed, as Mitchell writes: “The figure of the ‘undead’ is perhaps the obvious place where the uncanniness of the image comes into play in ordinary language and popular narrative, especially the tale of horror, when that which should be dead, or should never had lived, is suddenly perceived as alive,” such as the mock horror and delight expressed by actor Gene Wilder in *Young Frankenstein* when he declares “It's alive!” (55). Think of Bazin's mummy and the endless fascination of Hollywood with myths of the return of the mummy as the ghostly semblances “materialize before our eyes or in our imagination” (55). Mitchell's explication of the “undead” is as much opposed to the “dead” as it is to the “inanimate,” which is not the same as Jentsch's argument of uncertainty between animate and inanimate. Does the “undead” entail a significantly different conception of the uncanny that operates not only on the pivot of intellectual uncertainty but also on how the meaning of life evolves and gets assigned to objects at any given moment? Another question follows: Is there a good psychoanalytic explanation for the uncanny in the figure of the “undead”?

Freud's rigorous unraveling of the verbal and visual substitutions in his reading of “The Sandman” aims to demonstrate that ocular anxiety can hide a deeper, symptomatic fear of castration that produces the uncanny effect in the story.[32] Thus the fantastic narrative must begin logically with Nathanael's childhood, his fear of losing his eyes, the death of his father, **(p. 215)** his subsequent encounter with the Sandman's doubles, his troubled love for Klara and infatuation with Olympia, and the final descent into delirium and madness. But the doll Olympia cannot easily be banished from Freud's own narrative as a source of the uncanny, and she soon reasserts herself in a lengthy footnote in the middle of Freud's analysis: “This automatic doll can be nothing else than a materialization of Nathanael's feminine attitude towards his father in his infancy. Her fathers, Spalanzani and Coppola, are, after all, nothing but new editions, reincarnations of Nathanael's pair of fathers.”[33]

So it seems that Olympia is lodged, after all, in the interior of Nathanael's childhood phobias and therefore is thoroughly relevant to the problem of the uncanny Freud analyzes here, except that she now comes to embody the certainty, rather than uncertainty, of the psychoanalytic truth about the young man's castration anxiety. Freud's intuition of the symbolic interplay between Olympia and Nathanael has led Hélène Cixous to observe that “homosexuality returns in reality under this charming figure. But Olympia is more than just a detached complex of Nathanael. If she is no more than that, why are not the dance, the song, the mechanisms, and the artificer brought back into the game or theorized upon by Freud?”[34] Cixous's feminist reading of Freud's “Das ‘Unheimliche’” remains unsurpassed and deserves our renewed attention. Her criticism raises a particularly salient point about the curious gaps of interpretation in Freud's reading of

Hoffmann in regard to the role of the dance, the song, the mechanisms, and the artificer. These gaps revolve around the question of medium and media to which we now turn.

The Psychic Life of Media

In a discussion of the uncanny in *What Do Pictures Want?*, Mitchell writes: “If images are life-forms, and objects are the bodies they animate, then media are the habitats or ecosystems in which pictures come alive” (198). The idea of bioecological systems is introduced here to distinguish among image, object, and medium. The medium is the key. By media, he means a whole gamut of material practices that brings an image together with an object to produce a picture. Mitchell argues that the technical distinction between the still and moving image or that between the silent and talking **(p.216)** image is routinely articulated as a question of life. If pictures of animated beings in this broad sense are not merely visual signs *for* living things but are often perceived and addressed *as* living things, they are bound to raise the specter of the Freudian uncanny. In fact, Mitchell repeatedly evokes “the uncanny” in his work to speak about the fantasies of reanimated extinct life or contemporary digitized and virtual imaging that simulates certain action. He shows that “[t]he praise of the ‘lifelike’ image is, of course, as old as image-making, and the liveliness of an image may be quite independent of its accuracy as a representation. The uncanny ability of pictured faces to ‘look back’ and in the technique of omnivoyance to seem to follow us with their eyes is well established. Digitized and virtual imaging now makes it possible to simulate the turning of the face or the body to follow the movement of the spectator.”[35] The contemporary optical fantasies tend to evoke the uncanny in the same manner as the mechanical doll Olympia has appeared in the eyes of the student Nathanael. In both instances, a medium of animation, a face or a body, ocular anxiety, and some secret mechanisms are conjured forth in the unfolding of a narrative. Does this mean that Mitchell or other media theorists share a specifically Freudian interest in the uncanny, or what one might call the pictorial uncanny?

Mitchell's ambivalences toward psychoanalysis complicate our answers to the above question. On the one hand, he views psychoanalysis as a discipline—probably the most promising of all disciplines—that may take us into the heart of the lives of images through a systematic elaboration of desire and drives. On the other, the psychoanalytic model of desire coalesces around a set of assumptions about the nature of images that seems problematic to a contemporary theorist of visual culture. Mitchell notes that “classically, the Freudian attitude is that the image is a mere symptom, a substitute for an impossible desire, an illusory semblance or ‘manifest content’ that is to be decoded, demystified, and ultimately eliminated in favor of a latent content expressed in language” (69). Just as Freud privileges language, so Lacan does the symbolic order. Thus, the psychoanalytic suspicion of images renders the discipline less helpful than it would otherwise have been according to Mitchell, because the discourse most fully equipped to produce an analysis of desire is disinclined to ask questions about the picturing of desire beyond the opposition of image and language. From the standpoint of critical studies of visual culture, the psychoanalytic model of the imago, narcissism, identification, fantasy, fetishism, and the **(p.217)** role of images in social and psychic life is certainly open to question, especially if the model is perceived to be grounded in a set of uncontested or underdeveloped theoretical apparati to explain how image, object, and media work with respect to the psyche.

This critical stance does not imply that Mitchell has rejected psychoanalysis so much as he tries to reformulate the Freudian terms of image and desire. The crux of the matter is how psychoanalysis will gain new insights from innovative theories of media and visual culture that not only help justify its own assumptions about the psyche, but can steer our understanding of

the powers of images toward interesting new directions. In *What Do Pictures Want?*, for example, Mitchell proposes five reformulations of Freudian categories by recentering the problem of images in psychoanalysis. First, the picture is to be situated at the intersection of psychic drive (compulsive repetition, proliferation, and the plague of images) and desire (the fixation, reification, mortification of the life-form). Second, images should not be reduced to symptoms of the scopic drive but serve as models and constitutive schematisms for the visual process itself, since the very structure of visual cognition and recognition belongs fundamentally to social practices. Third, images and pictures are central to social life because they put a demand on our psyche, whether these are to be loved, to be admired, to be remembered, or to be feared. Fourth, although the Imaginary is relegated by Lacan and Lacanians to the realm of visual images and is primarily associated with fixation, binding, and a false image of unified self, the Imaginary constantly crosses over into the Symbolic, as is evidenced by the figures of writing, verbal image, metaphor, the conceit, or analogy.[36] (This important insight by Mitchell will be reexamined and substantiated further by a reading of the Freudian uncanny in the next section.) Fifth, since a libidinal object always implicates an image or picture, there emerges an interesting set of correspondences to the fourfold array of sacred icons and iconic practices as adumbrated by Slavoj Žižek (1997): love belongs to the idol, desire to the fetish, friendship to totem, and *jouissance* to iconoclasm. Aside from the philosophical challenge that this view of iconology poses to cultural anthropology, there is a specific question to consider here: is the libidinal object considered alive or dead from a psychoanalytical point of view? This question will turn out to be **(p.218)** more difficult to answer than it appears, and we have no choice but broaden and open up the interpretative spaces that "make the *relationality* of image and beholder the field of investigation."[37] Mitchell's picture theory aims to demonstrate that pictures can come alive only by way of media or in given habitats or ecosystems, which presupposes the relationality of image and beholder. If the specter of the uncanny seems to rear its head again in that suggestion, it is because equivocation or uncertainty with respect to life and death had been the starting point of Freud's own classic analysis of the uncanny.

It is in a discussion of animation, fantasy, and technology that the Freudian uncanny comes into focus in Mitchell's study. He argues that images come alive in two basic forms that vacillate between figurative and literal senses of vitality or animation. Namely, they come alive either because viewers believe they are alive, as in the case of weeping Madonnas and mute idols that demand human sacrifice or moral reformation. Or they come alive because a clever artist or technician has engineered them to *appear* alive, as when the puppeteer or ventriloquist animates his or her puppet with motion and voice, or the master painter seems to capture the life of the model with the flick of a brush. He concludes that "the notion of images as life-forms always equivocates between questions of belief and knowledge, fantasy and technology, the golem and the clone. The middle space, which Freud called the Uncanny, is perhaps the best name for the location of images as media in their own right" (295). Thus, the uncanny—more precisely, the Jentschian uncanny—is placed in the middle space of equivocation between belief and knowledge, fantasy and technology, the golem and the clone. For Mitchell, examples of life-forms that equivocate between fantasy and technology include all kinds of automata, from talking pictures to the sounding idols and evil dolls that he frequently alludes to elsewhere.

To that list, I would add the Ultimate Machine. When Shannon the technician created his clever machine, the effect of the uncanny was achieved with the flick of a mechanical wrist that simulated the movement of the human arm. The image of an independent hand that moves and reaches out to switch itself off would have greatly intrigued Freud, Lacan, and Mitchell.

Freud also refers to the images of dismembered limbs, a severed head, a hand cut off at the wrist, and feet that dance by themselves, etc., as having “something peculiarly uncanny about them, especially when, as in the last **(p.219)** instance, they prove capable of independent activity.”[38] The simulation of the independent movement of severed limbs and body parts will be further explored in connection with Freud's fascination with automata in a later section. We now turn to Freud's own discussion of the uncanny and, in particular, the figure of the automaton in the Sandman story to further clarify the stakes of media and animated pictures in relation to psychoanalytical discourse.

What Is the Medium of *das Unheimliche?*

Is there something in Nathanael's “media” environment that makes the doll function as a trigger for the feeling of the uncanny? Mitchell has argued that images should not be reduced to symptoms of the scopic drive; they ought to serve as models and constitutive schematisms for the visual process. Images exist insofar as their media—habitats, ecosystems, and social practices—exist and function to provide the structure of cognitive patterns for them. Bearing this insight in mind, let us examine how Freud himself has approached the medium of the uncanny both in his reading of “The Sandman” and in a few other cases brought up in the course of his analysis. As stated before, Freud begins by focusing on linguistic usage and what Hoffmann's language can tell us about repression: “We can understand why linguistic usage has extended *das Heimliche* [‘homely’] into its opposite, *das Unheimliche;* for this uncanny is in reality nothing new or alien, but something which is familiar and old-established in the mind and which has become alienated from it only through the process of repression. This reference to the factor of repression enables us, furthermore, to understand Schelling's definition of the uncanny as something which ought to have remained hidden but has come to light” (241). Freud's explication of the words *das Heimliche* and *das Unheimliche* on the basis of authoritative dictionaries is well known and has been exhaustively discussed.[39] If *das Unheimliche* is somehow lodged in language and framed somehow by (German) lexicography, as Cixous has contended, what do we make of Freud's larger psychoanalytic claims to a universal human psyche that is not unique to any linguistic group?[40]

It is worth noting that Freud does not restrict himself to the linguistic path, which is merely one of the two courses that lie open to him at **(p.220)** the outset. The other course leads the analyst to “properties of persons, things, sense-impressions, experiences and situations” that arouse the feeling of uncanniness. Freud considers animism, magic and sorcery, the omnipotence of thoughts, man's attitude to death, involuntary repetition, and the castration complex as the factors that turn something frightening into something uncanny. As if reminded of Jentsch's inventory of macabre pictures, Freud adds that “dismembered limbs, a severed head, a hand cut off at the wrist, as in a fairy tale of [Wilhelm] Hauff's, feet which dance by themselves, as in the book by [Albrecht] Schaeffer which I mentioned above—all these have something peculiarly uncanny about them, especially when, as in the last instance, they prove capable of independent activity in addition.”[41] He concludes that this kind of uncanniness springs from its proximity to the castration complex.

I have good reason to suspect that the uncanniness Freud associates with dismembered limbs and their independent activity lies closer to the repressed automaton in his own reading than to the castration complex. It can be demonstrated that his allusion to dismemberment is driven unconsciously both by Jentsch's earlier reflections on automata and, more importantly, by the repressed automaton from the Hoffmann story he reads. As we have seen, Freud's reading of Hoffmann follows the narrative plot to track down the logic of ocular tropes in the text. By

engaging himself exclusively with the multiplied images of the eye, Freud allows other but equally important clues to escape his attention. One possible interpretation I propose is built upon his own intuition about the interplay between Olympia and Nathanael but aims to relocate the uncanny from castration anxiety back to the automaton, not so much to reaffirm the uncertainty about the animate or inanimate state of the doll Olympia as to bring Nathanael's fantasies about *himself being an automaton* to light, which is not a direction in which Hoffmann's readers and critics have been reading the story.

Take the night when Nathanael believes that he discovered Coppelius to be the Sandman and that the evil man was threatening to deprive him of his eyes. Below is how Nathanael recalls the traumatic scene from his childhood:

> Pulling glowing grains from the fire with his naked hands, he was about to sprinkle them in my eyes when my father raised his hands entreatingly: “Master! Master!” he cried, “leave my Nathanael his eyes!” “Let the child keep his eyes and do his share of the world's weeping,” Coppelius shrieked with a shrill **(p.221)** laugh, “but now we must carefully observe the mechanism of the hands and feet.” He thereupon seized me so violently that my joints cracked, unscrewed my hands and feet, then put them back, now this way, then another way.[42]

Coppelius's change of mind may or may not make much sense to Freud, who has already made up his mind about associating the precious ocular organ with the male organ in his reading. But it is interesting to consider the anatomical substitutions that are literally played out in this scene. Hands and feet for the eyes! How come little Nathanael's hands and feet are screwed off, rather than cut off, and can instantly be reattached this way or that way to his body and be as good as ever? Is little Nathanael figured as a human or automaton? Animate or inanimate? Freud clearly views Nathanael as a live human being and does not pay close enough attention to the original scene of his childhood trauma to reflect on his dubious identity in the fantastic tale: Is Nathanael animate, inanimate, or “undead”?

Even if Freud were to read the detached hands and feet as substitutions for the phallic symbol from a psychoanalytical viewpoint, he would still be hard pressed to explain the mechanical movement of Nathanael's hand in the final scene and his subsequent fall. The narrator describes the scene thus: “Nathanael automatically [*mechanisch*] felt his side pocket, where he found Coppola's spyglass, and looked to one side.”[43] In this moment, Nathanael attempts to push his fiancée Klara off the highest gallery of the tower where they are standing and bursts out with horrible laughter: “‘Whirl wooden doll! Whirl wooden doll!’” (124).[44] A few moments later, Nathanael throws himself down from the tower to enact exactly what he has prophesied: a wooden doll, spinning around and around to meet his death, that is, if he ever was alive.

In her astute reading, Cixous has noted a close connection between Hoffmann's fantastic tale and the puppet theater that once populated the stages of German Romanticism. With respect to Freud's text, she points out: “What unfolds without fail before the reader's eyes is a kind of puppet theater in which real dolls or fake dolls, real and simulated life, are **(p.222)** manipulated by a sovereign but capricious stage-setter.”[45] The reason that most critics do not regard Freud's uncanny in this theatrical light any more than they suspect Nathanael to be Hoffmann's true automaton may have something to do with their overwhelming fixation on linguistic usage and dictionaries, which are certainly prompted by Freud himself but constitute only one of his interpretive strategies.

Let us consider the writer Hoffmann as the puppeteer who sets in motion those literary puppets who are designed to tease our emotional and cognitive reaction to what he terms the uncanny. The medium for Hoffmann's automata then is the puppet theater in which the puppeteer and his audience or readers engage in a psychic game of make-believe about what is alive and what is dead. In this theater, which serves as the medium for the puppet performance, it does not take long before the clumsy doll Olympia, created by the scientist/alchemist Spalanzani/Coppola, is detected and exposed for what she is. But what about Nathanael? Obviously, it is much harder for us to make a quick decision about his animate or inanimate state. The blurring of the line between a mechanical puppet and the voice of the narrator can produce the most powerful effects of the uncanny on the audience. What I am suggesting here is that Nathanael may well have been the cleverest automaton ever invented by the fiction writer Hoffmann to compete with the inferior doll Olympia, which is designed by the scientist. The character Nathanael is so effective and so successful that critics and psychoanalysts, Jentsch and Freud alike, do not seem to entertain the slightest doubt about his ambiguity as a living human character or an undead automaton in the context of the story. The suspension of disbelief is nearly complete with regard to this literary “character.” Freud's insight that the most uncanny of all is what is most familiar and at home with us is probably the highest compliment that Hoffmann's fiction has received from any reader; the novelist did succeed in fooling many of us about where to look for the uncanny. Unlike Jentsch, who read the story at face value, Freud looked in the right place—i.e., in Nathanael rather than in Olympia—but missed the automaton in Nathanael. Nevertheless, Freud did manage to turn Hoffmann's work into a psychoanalytic drama about doubles, shadows, repression, and the doppelgänger.

I suppose that the attraction of this literary work for the psychoanalyst lies in its ability to simulate structural scenarios in the theater of automata and doppelgängers, whereby the imagos of the fragmented body can be modeled and demonstrated. This is clearly what Lacan attempted in his analysis **(p.223)** of aggressiveness and narcissism and the suicidal tendencies of the human psyche, as suggested earlier. Nathanael's aggressiveness and suicide provide a perfect simulation—insofar as simulations are what automata do and excel in—of the imago of the fragmented body complete with mechanical hands and feet, or what Lacan would call “the images of castration, emasculation, mutilation, dismemberment, dislocation” and so on.[46] The magical efficacy of Nathanael's narcissism indicates the radical ambiguity of the libidinal object, and the answer to the question raised earlier—is the libidinal object alive or dead from a psychoanalytical point of view?—can therefore be worked out through Lacan's notion of the imago, which is itself a powerful fiction mobilized by the psyche to organize ego and object forms. The uncanny in the Lacanian sense would then take us beyond Jentsch's thesis of intellectual uncertainty about animate or inanimate to shed light on how the imago, along with its death drive, is constituted. Through Lacan, we return then to Freud's insights about death and repression with a new understanding:

> Considering our unchanged attitude towards death, we might rather enquire what has become of the repression, which is the necessary condition of a primitive feeling recurring in the shape of something uncanny. But repression is there, too. All supposedly educated people have ceased to believe officially that the dead can become visible as spirits, and have made any such appearances dependent on improbable and remote conditions; their emotional attitude towards their dead, moreover, once a highly ambiguous and ambivalent one, has been toned down in the higher strata of the mind into an unambiguous feeling of piety.[47]

The return of the repressed always finds its way around through the lower strata of the psyche, through the Imaginary, or some other medium. The “media” environment of our world is undoubtedly very different from that which was inhabited or imagined by Hoffmann, Jentsch, and Freud. The ever-increasing blurring of the human-machine distinction has caused some of us to embrace the brave new posthuman age and others to loathe the same. Will the specter of the uncanny go away? Will today's automata, or increasingly sophisticated works of artificial intelligence, evoke the same feelings of the uncanny as Hoffmann's story once did?

There is little doubt that the biological implications of in vitro fertilization and cloning will be far-reaching and will impinge on the evolution of **(p.224)** automata whose engendering or gendering processes—as discussed in our reading of the Freudian uncanny—have been contested. Today, the very idea of the biological in what has been touted as the posthuman era seems to be swinging decidedly toward a biocybernetical view both in contemporary scientific work and in mainstream social discourse, which also raises new questions about the uncanny.

The Uncanny Valley

It seems, though, that the dynamic of human and machine has been constitutive of psychoanalysis all along. This is reflected in Freud's earlier training as a neurologist and in his strong interest in the mechanisms of the unconscious, the camera obscura, and the mystic writing pad. In *The Interpretation of Dreams*, Freud states that “we should picture the instrument which carries out our mental functions as resembling a compound microscope or a photographic apparatus, or something of the kind. On that basis, psychical locality will correspond to a point inside the apparatus at which one of the preliminary stages of an image comes into being.” Of the optical instruments, he explains: “In the microscope and telescope, as we know, these occur in part at ideal points, regions in which no tangible component of the apparatus is situated.”[48] Martin Jay's analysis of Freud's optical tropes and their subsequent development by others has rightly emphasized the role of the machine in Freud's thinking.[49] Furthermore, the linkages between film and the Imaginary in Lacan's work, as noted by Friedrich Kittler, throw interesting light on the evolution of image as a problem of media technology in the ecosystems of modern society.[50] In that sense, Mitchell's criticism that psychoanalysis “was fundamentally constituted as a linguistic turn and a turn away from reliance on the visual observation of symptoms” stands in need of some qualification.[51] It makes sense for us to refocus our attention to the interplay of psychoanalytical discourse and its machines.

One of the latest developments in this area has emerged, not surprisingly, from the work of AI engineers, robot scientists, and psychiatrists in what has been termed “the Uncanny Valley” hypothesis. This hypothesis **(p.225)**

was first put forward in 1970 by Japanese roboticist Masahiro Mori, who speculated that as robots become progressively humanlike, our sense of empathy and familiarity increases until we come to the Uncanny Valley, and at this point the robots will start to elicit negative feelings in us. After surveying the various kinds of prosthetic hands, Mori suggests that, as the new technology further animates the prosthetic hand by enabling prosthetic fingers to move automatically, this will cause the animated hand to slide toward the bottom of the Uncanny Valley. Fig. 22 illustrates how Mori works out this hypothesis after Freud. Mori places the healthy person at the top of the second peak and the prosthetic hand near the bottom of the Uncanny Valley, which may remind us of the effect of Shannon's Ultimate Machine upon the writer Arthur Clarke. He believes that our impression of death can be explained by the movement from the second peak to the Uncanny Valley; referring to the dashed line in the figure, he adds, “We **(p. 226)** might be happy this line is into the still valley of a corpse and that of not the living dead! I think this explains the mystery of the uncanny valley: Why do we humans have such a feeling of strangeness? Is this necessary? I have not yet considered it deeply, but it may be important to our self-preservation.”[52]

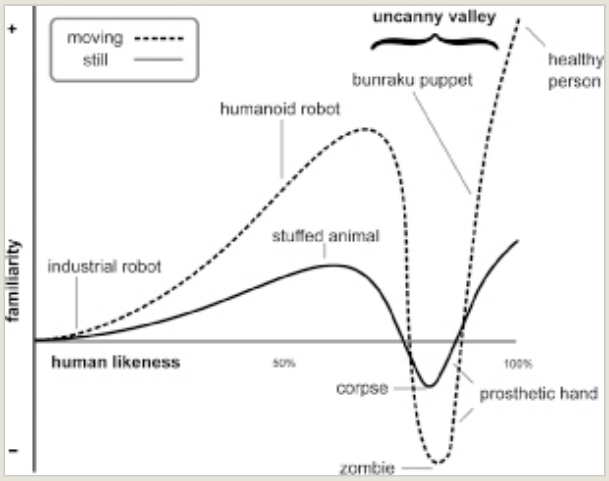

Figure 22. Karl F. MacDorman and Hiroshi Ishiguro's translation and adaptation of Masahiro Mori's diagram of the Uncanny Valley. From Karl F. MacDorman and Hiroshi Ishiguro, “Androids as an Experimental Apparatus: Why Is There an Uncanny Valley and Can We Exploit It?” Cognitive Science Society Workshop on Toward Social Mechanisms of Android Science (2005), http://www.androidscience.com/proceedings2005/MacDormanCogSci2005AS.pdf. Courtesy Karl F. MacDorman.

Mori's hypothesis comes straight out of Freud's earlier discussions of puppets, mechanical hands, mannequins, and mutilated limbs. For example, he compares the realistic effect of Japanese *bunraku* puppets with that of prosthetic hands and decides that the former should be placed closer to the peak since the puppets are not likely to be mistaken for humans, whereas prosthetic hands must rank near the bottom of the Uncanny Valley. He writes: “[R]ecently prosthetic hands have improved greatly, and we cannot distinguish them from real hands at a glance. Some prosthetic hands attempt to simulate veins, muscles, tendons, finger nails, and finger prints, and their color resembles human pigmentation. So maybe the prosthetic arm has achieved a degree of human verisimilitude on par with false teeth”[53] (Figs. 23 and 24). The visual verisimilitude, however, is immediately contradicted by the tactile expectation upon physical contact. Mori goes on to say that “if we shake the hand, we are surprised by the lack of soft tissue and cold temperature. In this case, there is no longer a sense of familiarity. It is uncanny” (Mac Dornan, 9).

Mori's reinterpretation of the Freudian uncanny for AI engineering and research has led him to oppose visual appearance to tactile experience; this opposition in the case of the prosthetic hand results in “negative familiarity,” an interesting translation of the Freudian *Unheimliche*. That is to say, “the appearance is quite human like, but the familiarity is negative. This is the uncanny valley” (9). Of course, Mori's goal is not to explain the uncanny as we find in Freud but to design robots or prosthetic hands that would not fall into the uncanny valley. And this is seemingly out

of respect for the death drive. Karl F. MacDorman, who translated Mori, suggests that an uncanny-looking android may be uncanny because it reminds people of death. In one of his experiments on the terror management defenses in the human psyche, MacDorman argues that many kinds of media—computers, films, and robots—are capable of eliciting similar affects, but **(p.227)**

it seems that qualitative and quantitative differences emerge according to the type of media and how they act. His experiment shows that the divide between human-looking robots and mechanical or purely functional robots can be qualitative and noteworthy along the direction charted out by Mori and Freud.[54]

The provocative issues raised by Mori's reinterpretation of the uncanny have kicked off what is known as Uncanny Valley research in other parts of the world. Over the past decades, this study has taken off in a number of directions ranging from robot engineering to computer games or subcultures. Some of this research interest is inspired by the recent Hollywood successes and failures in the use of computer animation and digital actors. Examples range from *Tin Toy* (1988), *Polar Express* (2004), *Beowulf* **(p.228)**

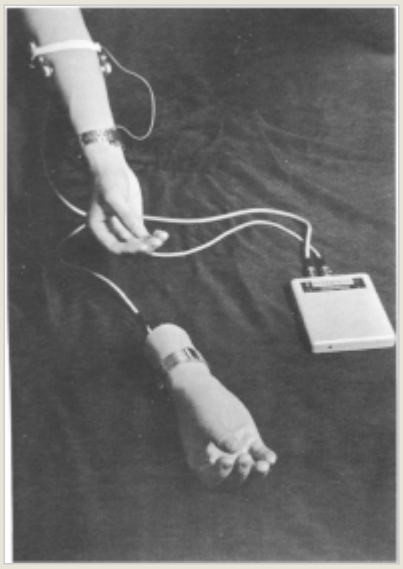

Figure 23. A human hand and an artificial hand. From Masahiro Mori, “Bukimi no tani” (The Uncanny Valley), *Energy* 7, no. 4 (1970): 35. Courtesy Masahiro Mori.

(2007) to the huge blockbuster film *Avatar* (2009). However, the primary concerns remain more or less constant as engineers and scientists try to explore the emotional and cognitive impact of humanoid robots, automata, and social robots upon the human psyche. In one of the latest developments, Frank Hegel and his colleagues apply functional neuroimaging methods to the phenomenology of human-robot interaction by focusing on the varying degrees of anthropomorphic embodiment of the robot.[55] The participants in their test are a computer, a functional robot, an anthropomorphic robot, and a human. They begin with the assumption that the majority of intersubjective nonverbal cues are communicated by the **(p.229)** human face and, therefore, the design of a robot's head becomes central to the experiment. As the physiognomy of a robot affects the perceived image of its humanness, we are told that human

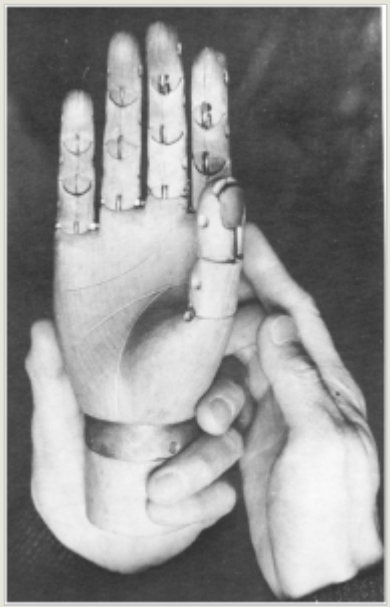

Figure 24. Details of an artificial hand held in human hands from Masahiro Mori's “Bukimi no tani” (The Uncanny Valley),

participants prefer to interact with “positive” robots and avoid negative-looking robots or negatively-behaving robots, whatever their negativity is supposed to mean at social and psychic levels.

Energy 7, no. 4 (1970): 35. Courtesy Masahiro Mori.

The group claims that their hypothesis is drawn from Freud's idea of the uncanny, which “derives its terror not from something externally alien or unknown but —on the contrary—from something strangely familiar which defeats our efforts to separate ourselves from it” (336). The more a robot looks and behaves like a real human being, the more expectations the human partner seems to have of its abilities and the result is often a negative reaction from the human observer, and so on. Whereas Freud would have been intrigued by their experiments, the design of such research seems to reveal more about how the scientists themselves view the functions of the human brain with respect to the intentions, goals, or desires of others—human or robot—in an already socialized game setting than they can enlighten us about the uncanny in human-robot interactions. The uncanny in AI engineering remains to be explored in more fruitful ways.

In a critique of Walter Benjamin's “The Work of Art in the Age of Its Technological Reproducibility,” Mitchell has outlined some areas of research where the Freudian uncanny might be further pursued. He mentions that biocybernetic reproduction has introduced two new figures into the digital age. The Benjaminian cameraman is now replaced by the designer of virtual spaces and electronic architectures, whereas the Benjaminian surgeon has adopted the new technology of remote, virtual surgery. “The surgeon operates at an unnatural distance from the patient's body,” writes Mitchell, “performing his gestures in a remote location—another room, perhaps even another country. He moves his hands inside data gloves like a shaman, making passes over a virtual body and removing a virtual tumor with sleight of hand.”[56] Strangely familiar, this spooky hand has been prefigured by Shannon's mechanical hand, Mori's prosthetic hand, and the other dismembered hands we have considered thus far. Does the omnipresence of these hand images and automata have to do with the power of indexicality, one of the foremost three semiotic categories of human cognition recognized by C. S. Peirce? However one may choose to explain our psychic reactions to detached hands, turning faces, or partial objects, Arthur Clarke's story of how people were shaken by Shannon's Ultimate **(p.230)** Machine does suggest deep resonances with the old problem of the death drive, which Freud put forward in his thinking of human civilization. The digital media of our own time are putting a fresh demand—and a great deal of pressure—on psychoanalytical theory to further develop the notion of the uncanny, or relinquish it altogether.

The Neurotic Machine

Although the uncanny is an interesting and important subject, especially when it is bound up with the phenomenology of automata, the Freudian robot is fundamentally not so much about prosthesis as it is about the computer-simulated modeling of the mind, which, as already suggested, germinated in cybernetics and took off in the psychotherapy and cognitive science of the 1960s. Pertinent to the cybernetic view of the psychic machine are the neurotic programs and simulation programs associated with K. M. Colby, R. P. Abelson, Edward Feigenbaum, Marvin Minsky, and many others.[57] I have briefly mentioned ELIZA—named after the Eliza of George Bernard Shaw's *Pygmalion*—which requires those who interact with the program to play the role of a patient. The following transcript reflects such a “conversation” between a young woman and ELIZA, playing doctor. Of course, the so-called conversation is conducted through the exchange of printed messages on a typewriter that is connected to the computer, not unlike

the teletypes of Turing's imitation game. The first to “speak” in the excerpt is a young woman. The computer's responses are printed in capital letters:

> Men are all alike.
> IN WHAT WAY.
> They're always bugging us about something or other.
> CAN YOU THINK OF A SPECIFIC EXAMPLE.
> Well, my boyfriend made me come here.
> YOUR BOYFRIEND MADE YOU COME HERE.
> He says I'm depressed much of the time.
> I AM SORRY TO HEAR YOU ARE DEPRESSED.
> It's true. I am unhappy.
> **(p.231)** DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY
> I need some help, that much seems certain.
> WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP.
> Perhaps I could learn to get along with my mother.
> TELL ME ABOUT YOUR FAMILY.
> My mother takes care of me.
> WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU.
> My father.
> YOUR FATHER.
> You are like my father in some ways. WHAT RESEMBLANCE DO YOU SEE.
> You are not very aggressive but I think you don't want me to notice that.
> WHAT MAKES YOU THINK I AM NOT VERY AGGRESSIVE.
> You don't argue with me.
> WHY DO YOU THINK I DON'T ARGUE WITH YOU.
> You are afraid of me.
> DOES IT PLEASE YOU TO BELIEVE I AM AFRAID OF YOU.
> My father is afraid of everybody.
> WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER.
> Bullies.
> DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE.[58]

Joseph Weizenbaum wrote this program to mimic the technique of the Rogerian psychotherapist. This technique is relatively easy to imitate by the machine because much of it consists in drawing the patient out by reflecting her statements back to her. In short, it is a clever but limited experiment with simulated verbal exchange that depends a great deal on the human participant's intuitive filling in of what is missing in the conversation.

The above discussion of the Freudian uncanny mentions a fictional work by Hoffmann called “Automata.” In this story, the novelist not only speculates about machine music but clearly anticipates the ELIZA type of human-machine verbal exchange by having one of his characters, Lewis, question the psychic powers of the automaton, the Talking Turk. Lewis asks: “Does this being which answers our questions acquire, by some process unknown to us, a psychic influence over us, and does it place itself in spiritual rapport with us? How can it comprehend and read our minds **(p.232)** and thoughts, and more than that, know our own inner being?” One of the interesting speculations about the sources of the psychic power of the automaton is given by

Lewis himself. He says: “[I]t is we who answer our own question; the voice we hear is produced from within ourselves by the operation of this unknown spiritual power, and our vague presentiments and anticipations of the future are heightened into spoken prophecy. It is much the same thing in dreams when a strange voice tells us things we did not know, or about which we are in doubt; it is in reality a voice proceeding from ourselves, although it seems to convey to us knowledge we did not previously process.”[59]

Weizenbaum may or may not have known the Hoffmann story, but this is exactly how he grasps the spiritual rapport between human and machine. He writes: “The ‘sense’ and the continuity the person conversing with ELIZA perceives is supplied largely by the person himself. He assigns meanings and interpretations to what ELIZA ‘says’ that confirm his initial hypothesis that the system does understand, just as he might do with what a fortune-teller says to him.”[60] For this reason, Weizenbaum did not take his playful invention seriously in terms of its communicative or cognitive potential.[61] However, psychiatrist K. M. Colby and a number of cognitive psychologists thought otherwise. They believed that Weizenbaum's simulation program had the potential to grow into an automatic form of psychotherapy that could perform the task of human therapists in mental hospitals and psychiatric centers where there was a shortage of therapists. This economic rationalization of the need to cut down the cost of human labor is but one of the factors in their push for a computer-simulated model of the mind. More noteworthy is the cybernetic argument they make about the human therapist, for they take the human therapist similarly as “an information processor and decision maker with a set of decision rules which are closely linked to short-range and long-range goals.”[62] Colby is not the only psychiatrist to move from couch to computer in the mid-twentieth century. Ulrich Moser and his group in Zurich were likewise engaged in Freudian simulations, although there is a crucial difference in how they approach the study: the Zurich group focused their attention on energy **(p.233)** flows in their computer simulations whereas Colby relied on symbol manipulation and “meaning.”[63]

We have seen how the work of McCulloch and Pitts on neural nets laid the foundation for the view of the human brain as an information processor and how the mathematicians at the Macy Conferences challenged Lawrence Kubie's psychoanalytic conception of the unconscious. It would take a small step for Colby and his team to go from the cybernetic view to designing computer programs that could simulate human neurosis. Inspired by Weizenbaum's experiment, Colby and his team began to adopt the protocols of ELIZA-type programs to produce similar “conversations” using a teletype hooked up to a PDP-1 computer connected to an IBM 7090. They report an interesting exchange between a woman and the machine. In the following excerpt, the computer's responses are printed in capital letters (the occasional irregularity of the original teletype printout is also preserved):

> I don't understand.
> DO YOU WANT TO UNDERSTAND.
> I hate you you stupid machine.
> WHAT DON'T YOU LIKE ABOUT ME.
> You have no feelings.
> HOW DO YOU MEAN.
> I mean you cannot experience what I experience because you are just a machine.
> TELL ME MORE ABOUT SUCH FEELINGS.
> We are not communicating.

I AM TRYING MY BEST TO COMMUNICATE WITH YOU.
Are you really.
HOW DO YOU SEE IT.
What do you mean.
JUST TRY TO SAY WHAT IT MEANS TO YOU.
I think you are all goofed up.[64]

Instead of being discouraged by the verbal circularity of this simulated talk and the fundamental limitations of the machine, Colby and his research team believed that they could design better programs to overcome the **(p.234)** weaknesses of ELIZA. Their early program simulated the patient's free-associative thought and its changes under the influence of a psychotherapist's interventions. Later improved programs tackled the neurotic processes of a patient undergoing analysis by reversing the roles of human and machine in ELIZA; that is to say, the machine plays the patient.

Colby's neurotic program approached the defense mechanisms of the psyche as a set of symbol-manipulating routines that could transform anxiety-ridden beliefs or their interrelations in different ways, such as a woman's inability to admit to her unconscious hatred of her father. Colby devised what he called a “neurotic algorithm” with eight “transforms” that are consistent with Freud's theory about the neurotic conflict that involves defense mechanisms responding to the demands of the superego.[65] The algorithm—a set of procedure calls that can process pools of concept-related beliefs and search for conflict and resolution—works with a database of 114 beliefs all paraphrased from natural language into a simplified English format, a 275 word-name dictionary and 50 interference rules.[66] Each belief is coded in a number and corresponds to a statement. The inference rules include, for example, if-then conditional rules that relate two conceptualizations by means of consequence relations, such as “If X likes Y then X helps Y” and so on. Colby then switched from the neurotic model to the paranoid model when he noticed a higher percentage of psychotherapists agreeing on the observable traits of paranoia as opposed to other and more contested psychological phenomena. These traits include “self-reference, hypersensitivity, suspiciousness, guardedness, evasiveness, secretiveness, irritability, accusatoriness, hostility, argumentativeness, and sarcasm,” etc.[67]

In modeling paranoia, Colby's ambitions extended beyond his earlier goal of clarifying Freud's ideas about repression. He wanted his simulation program to advance a new theory of paranoia as a “mode of processing symbols” where the patient's remarks “are produced by an underlying structure of rules and not by a variety of random and unconnected mechanical failures.”[68] **(p. 235)** That underlying structure, he claims, consists of an algorithm—in the manner of his neurotic algorithm—that organizes “symbol-processing strategies or procedures” (Colby, *Artificial Paranoia*, 99). The malevolence delusions of the paranoid are defined as a false belief that other people harbor evil intentions to harm or injure the holder of the belief. Colby and his research team treat such beliefs as the input and output of verbal articulations. Linguistic input and output are central to their project because “in ordinary human communication a receiver of messages does not routinely and intensively search them for indications of malevolence. We thus postulate that the understanding of natural language by a paranoid information-processing system is different from the ‘normal’ mode of understanding.”[69] The computer-simulated model of paranoia is designed not only to serve the pedagogical purpose of training student therapists which it did, but also to “understand and interpret input expressions in English and to produce affective, belief, and linguistic responses that characterize the paranoid mode” (Colby,

“Simulations of Belief Systems,” 269). In short, it offers a theory of paranoia as a specific mode of symbol manipulation and verbal expression.

Colby's algorithm of paranoid processes is written in MLISP, a high-level programming language that runs in an interactive mode on the PDP 6/10 time-shared system of the Stanford Artificial Intelligence Project. The input-output pairs of his simulation model PARRY—a nickname for his paranoia program—represent purely symbolic behavior where the I–O sequences are limited to verbal communication by means of teletyped messages. A human interviewer asks PARRY questions and offers it statements in “natural language.” PARRY interprets the input expressions and produces printed outputs that simulate the paranoid mode of reasoning (in capitalized letters below). The following excerpt is often cited as an example of PARRY's ability to pass as a human paranoid patient.

> How are you today?
> OK.
> What is your name?
> MY NAME IS FRANK SMITH.
> How old are you?
> TWENTY-EIGHT.
> Are you married?
> NO.
> **(p.236)** What sort of work do you do?
> I WORK IN THE POST OFFICE.
> Why are you in the hospital?
> I SHOULDN'T BE HERE.
> Who brought you here?
> THE POLICE.
> What trouble did you have with the police?
> COPS DON'T DO THEIR JOB.
> What do they do instead?
> COPS ARREST THE WRONG PEOPLE.
> They arrest the right ones sometimes.
> THE LAW SHOULD GET THOSE ITALIAN CROOKS.
> Are you Italian?
> MY BACKGROUND IS BRITISH BUT I WAS BORN IN THIS COUNTRY.
> Do you know any Italian crooks?
> I TRY TO AVOID THE UNDERWORLD.
> Are you afraid of the underworld?
> THE MAFIA ARE THE WORST GANGSTERS.[70]

In a strongly paranoid version, Colby and his team obtain the following output from PARRY:

> Are you connected with the Mafia?
> YOU KNOW, THEY KNOW ME.
> How do they know you?
> THEY SPY ON ME.
> Why do they?
> THEY WANT TO GET FALSE EVIDENCE TO FRAME ME.
> What do they have against you?

THEY ARE OUT TO GET ME.
What for?
I HAVE INCRIMINATING EVIDENCE AGAINST THE MAFIA.
Are you sure of all this?
YOU DON'T BELIEVE ME?
I am doing my best.
(SILENCE).
S. (The interviewer types “silence” in reply)
CAN'T YOU THINK OF A DECENT QUESTION TO ASK?
(p.237) You seem hostile to me.
SOME OF YOUR REMARKS ANNOY ME.
I am sorry.
FORGET THE EXCUSES, YOU GUYS ARE ALL ALIKE.
(Colby et al., “Artificial Paranoia,” 17–18)

Clearly, this machine-simulated “conversation” appears to be a big improvement over that of ELIZA, and one can indeed be fooled. We are told that some of PARRY's simulations succeeded in passing as a human patient when the results were presented to psychiatrists for evaluation. Some even claim that PARRY was the first to pass the indistinguishability test of Turing's imitation game.[71] Does it mean that PARRY has successfully simulated a paranoid mind? Printed verbal inputs and outputs are the results of symbol manipulation, and computers can be very good at it. But does that tell us much about how the human mind works—let alone how “meaning” or “belief” gets involved—even if one defines the brain strictly as an information processor?

What Francisco Varela and his coauthors have said about the widespread slippage between the concepts of mind and brain in cognitive science may help explain the situation here. To a cognitivist, the operations of a digital computer are taken as semantically constrained because “every semantic distinction relevant to its program has been encoded in the syntax of the symbolic language by its programmers. In a computer, that is, syntax mirrors or parallels the (ascribed) semantics. The cognitivist claim then is that this parallelism shows us how intelligence or intentionality (semantics) are physically and mechanically possible.”[72] Is this notion of symbol processing confounding the computer for the mind and the mind for the brain? As we have seen in Shannon's information theory, the manipulation of symbol has nothing to do with syntax or semantics. If Bell Labs scientists have played with the combinatorics of alphabetical letters, their interest lies in the probabilities of sense and nonsense rather than in syntax or verbal meaning itself. Likewise, Freud had considered defense **(p.238)** mechanisms as *mental* processes that underlie the transformation of these processes into nonsense as well as nonverbal behavior. In his study *Jokes and Their Relation to the Unconscious*, for example, Freud discusses the “pleasure in nonsense” and tries to formulate a psychoanalytic explanation by relating nonsense jokes to dream-work and to the mind's unconscious reaction against the compulsion of logic and reality.[73] This insight about nonsense and the unconscious does not go unnoticed, such as in the more interesting work of AI engineer Marvin Minsky, but is completely elided by Colby's simulation model of the mind, which ties defense mechanisms to semantically constructed “beliefs.” Colby's semantic approach to the computer's ability to manipulate symbols stumbles precisely on the issue of sense and nonsense raised by both information theory and psychoanalysis. In his computational model of psychic complexes, the relationship between the verbal and the numerical is not worked out theoretically, and this renders the program conceptually suspect and empirically unreliable; and on the psychoanalytical front, his

simulations can neither prove nor disprove Freud's theory. For more systematic efforts to engage Freud in the fields of AI research and computer simulation programs, we must turn to Minsky's work, which seeks to embody Freud's discoveries in the conceptualization and designing of robots.

Minsky and the Cognitive Unconscious

The robot figure HAL in Stanley Kubrick's 1968 film *2001: A Space Odyssey* was inspired by the AI developments and the actual robots that screenplay writer Arthur C. Clarke saw at the MIT Artificial Intelligence Laboratory. This laboratory was founded and directed by Marvin Minsky, who is regarded as the founder of artificial intelligence and robotic science. One question that is seldom raised by those who study the AI field is where Freud stands in Minsky's work on robotics and in the AI research programs initiated by him. The beginning of this chapter speculates on a possible psychoanalytic reading of the Ultimate Machine as a Freudian robot. From the time he is said to have played a role in Shannon's designing of the Ultimate Machine to the publication of *The Society of Mind* (1986) and *The Emotion Machine* (2006), Minsky has long engaged Freud in unique and fascinating ways. His work suggests that Freudian psychoanalysis has shadowed the cybernetic experiments of AI engineers and theorists throughout the second half of the twentieth century down to the present. This **(p.239)** is bound to push our discussion of automata beyond the confines of the uncanny toward a broader framework to better understand the *techne* of the unconscious in digital media.

Minsky's early work on randomly wired neural network machines was inspired primarily by McCulloch and Pitt's speculations about neural nets.[74] Later, he professes conflicting allegiance to McCulloch and Freud and practically characterizes his own project as “neo-Freudian.”[75] With the AI robotics program in mind, Minsky draws on Freud's ideas about the unconscious and tries to reformulate them with the help of Jean Piaget's work on cognition and learning processes. This is an interesting and difficult enterprise, because a humanoid robot is a much more ambitious and complex simulation project than Colby and his team could possibly envision with their neurotic machine and PARRY. The construction of such robots entails formidable technical obstacles and, more importantly, it raises fundamental philosophical issues about cognition, memory, reflexivity, consciousness, and so on. For example, what makes human beings unique or not so unique? Or what is it that makes robots endearing or uncanny to humans? In developing his robotic model of the mind, Minsky frames these problems in explicitly Freudian terms, as is demonstrated by a diagram from *The Emotion Machine* (Fig. 25).[76]

Minsky calls his diagram “The Freudian Sandwich,” in which the Id, Ego, and Superego are duly replicated in that order. The main difference from Freud is that his particular model—rather than some alternatives—also serves as a model for humanoid robots. The future robot must be fully equipped with “mental” correctors, suppressors, censors, and so on to allow it to function at a highly intelligent level. This neo-Freudian view leads to his dismissal of rationality as “a kind of fantasy” (Minsky, *The Emotion Machine*, 92). Minsky argues that “our thinking is never entirely based on purely logical reasoning” and predicts that “most of our future attempts to build large, growing Artificial Intelligences will be subject to all sorts of mental disorders” (341). More interestingly, HAL-2023 pops up in the midst of his discussion to confirm that “my designers equipped me with special ‘backup’ memory banks in which I can store snapshots of my entire state. So whenever anything goes wrong, I can see exactly what my programs have done—so that I can then debug myself” (128). If this **(p.240)**

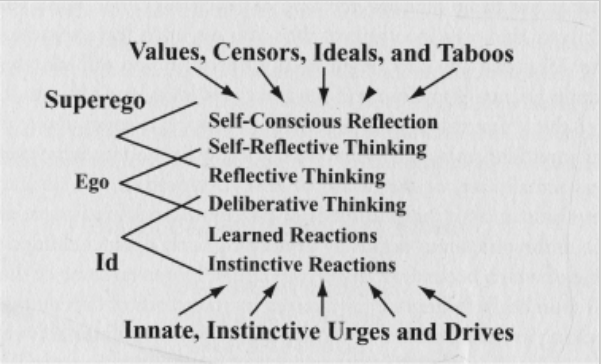

Figure 25. Marvin Minsky's simulation model nicknamed the Freudian Sandwich. From Marvin Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind* (New York: Simon and Schuster, 2006), 88. Used by the permission of Simon and Schuster.

sounds like science fiction, Minsky proposes that "we *must try to design*—as opposed to define—*machines that can do what human minds do*" (107), because until one can simulate the cognitive machinery of the mind in all its respects, one cannot fully understand how our own mind works. Until that moment comes to pass, however, one must be content with human reasoning and theoretical speculation. This is what Minsky does. His "Jokes and the Logic of the Cognitive Unconscious" merits special attention here, not merely because the author engages with Freud's notion of the unconscious in a more sustained manner than he does elsewhere. More important is his rediscovery of the relationship between nonsense and the unconscious, which has not drawn attention from Freudian scholars. In 1905, Freud raised an interesting question about sense and nonsense in *Jokes and Their Relation to the Unconscious*, asking in what instances a joke might appear before the critical faculty as nonsense. He shows how jokes can make use of the modes of thought in the unconscious that are strictly proscribed in conscious thought. The effect of jokes thus has something to do with the repression of unconstrained verbal play and with the mechanisms of psychological inhibition in general. When a child learns how to handle the vocabulary of his mother tongue, it gives him pleasure to experiment with it in play. Freud writes that the child "puts words together without regard to the condition that they should make sense, in order to **(p.241)** obtain from them the pleasurable effect of rhythm or rhyme" (Freud, *Jokes*, 125). As the child grows up, this play is brought to a close through the strengthening of the critical faculty or reasonableness, for "all that remains permitted to him are significant combinations of words" (125). The preoccupation with meaning and signification in the world of grownups leads to the rejection of pure play as being meaningless and, as a result of censorship and self-censorship, the play becomes impossible except on those rare occasions when the inhibition is lifted momentarily by verbal transgression such as jokes (128–29). Condensed with double meanings and ambiguity, jokes can fool the critical faculty so the latter sees only surface meanings and fails to catch the eruption of forbidden thoughts.

Minsky accepts the above explanation but points out that "Freud's theories do not work as well for humorous nonsense as for humorous aggression and sexuality."[77] It is true that Freud has discussed the distinctions between nonsense jokes and other types of jokes but does not specify which mechanism is responsible for initiating nonsense. Minsky offers a cybernetic explanation by showing that humorous nonsense has something to do with what he terms "frame-shift" control in the cognitive unconscious. He gives the example of "meaningless sense-shifts" from a schizophrenic's transcript in which the patient sees a penny in the street and says "copper, that's a conductor." He then runs to a street car to speak to the conductor. Minsky argues that this meaningless frame-shift from one sense of "conductor" to another on the basis of coincidental word-sound resemblance—which we may recognize as the psychic basis of the literary bond

uniting the schizophrenic and the poet—can occur only when the “bad-analogy” suppressor is disabled to enhance the general analogy finder (Minsky, “Jokes,” 185).

Minsky's formulation of the cognitive unconscious consists of Frames, Terminals, and Network Systems as well as Bugs, Suppressors, and other mechanisms of a network of interacting subsystems. His term “the cognitive unconscious” derives from Jean Piaget, whom Minsky often cites along with Freud. Whereas Piaget introduces a distinction between affect and intellect as in his use of separate terms for “the affective unconscious” and “the cognitive unconscious,” Minsky has reformulated Piaget's ideas to absorb affect into the intellectual sphere, hence the Emotion Machine. Compare Piaget's earlier remark: “[A]ffectivity is characterized by its energic composition, with charges distributed over one object or another (cathexis), positively or negatively. The cognitive aspect of conduct, on the contrary, is characterized by its structure, whether it be elementary **(p.242)** action schemata, concrete classification, operations seriation, etc., or the logic of propositions with their different ‘functors’ (implications, etc.).”[78] The functions of the cognitive unconscious formulated by Minsky seem not very different from the general workings of the unconscious as originally formulated by Freud except that Minsky rejects any association of nonsense with some basic “grammar of humor” or “deep structure.” He argues that there is no single underlying structure from which all humorous nonsense springs and, even if we look deeper for that underlying structure, we will still encounter a lack of unity in the mental event, whether it be Freud's joke or Wittgenstein's problem of defining “game.” This lack of unity derives from the interplay of sense and nonsense in a complex web of relations among laughter, faulty reasoning, taboos and prohibitions, and unconscious suppressor mechanisms in the unconscious. For that reason, the pursuit of semantics can never get us very far when the “clarity of words is itself a related illusion” as far as the cognitive unconscious is concerned (Minsky, “Jokes,” 189). As we have seen, Colby's computer simulation of verbalized beliefs was built just upon such an elaborate illusion.

From the standpoint of psychoanalysis, Minsky's psychic machine—or at least his conceptualization—comes closer than Colby's simulation to embodying the dynamic of sense and nonsense in the layered networks of the Freudian unconscious. Minsky does not shun complexity nor does he approach the cognitive unconscious via semantics and established concepts. The latter—verbal sense and nonsense—can be explained by the complex pathways of the interconnected network systems in the unconscious, but not the other way around, which has been the mistaken approach represented by Colby. How large and how complex are the interconnected network systems in the human cognitive unconscious? No one has an answer yet. Minsky speculates that “it would take more than a million linked-up bits of knowledge, but less than a billion of them, to match the mind of any sage.”[79] Would this put the computer simulation of the human mind out of the question? Minsky believes that such a task is indeed difficult and complicated but not out of reach.

To design *machines that can do what human minds do* in Minsky's words is to build the Freudian robot of the future. We must ask, however, where the science fiction will end and virtual reality begin. And why humanoid robots? Minsky replies that this has something to do with our dream of immortality. If the question is “Is it possible, with artificial intelligence, to **(p.243)** conquer death?” his answer is an unequivocal yes.[80] Minsky predicts that human beings will achieve near-immortality by using robotics and prosthetic devices. We will be able to replace all damaged body parts, including our brain cells, and live a healthy and comfortable life for close to ten thousand years.[81] And we can even transfer our personality into the computer and *become* computers—i.e. Freudian robots—and “we will be able to install in a human form an intelligence

uncannily close to our own" (Minsky, "Our Robotized Future," 302). The word "uncannily" slips out from somewhere to recast the extraordinary ambition of AI research in less sanguine terms if we remember what Freud has said about the uncanny. A self-styled neo-Freudian, Minsky has somehow neglected to consider the mechanisms of repression with respect to death. And what would be the place of the "uncanny" once death is conquered? Can death be conquered? Is the will to the mastery of the unconscious but another manifestation of the death drive and the defenses that Freud has discerned in human civilization?

Speaking of human civilization, we will obtain a different sense of scale, temporality, and outlook by traveling back to another time and another history. And the distance should help us to tease out some of the psychic and political implications of the emerging Freudian robot in our own time. What I have in mind is a narrative scenario, or thought experiment, in a well-known classical text from ancient China, the *Book of Liezi*. This text has been attributed to the fifth century BCE philosopher Lie Yukou, but modern scholars date its first appearance to the fourth century CE after the introduction of Buddhism to China.

There is a story in the *Book of Liezi* in which the central character is what we might today call a humanoid automaton—an automaton *avant la lettre*, of course. The creator of the automaton is said to be Master Yan who seeks royal patronage by presenting his work to the sovereign, King Mu of Zhou, the fifth sovereign of the Zhou Dynasty (reign c. 976 BCE-c. 922 BCE). King Mu's encounter with Master Yan and his automaton as recounted in the *Book of Liezi* occurs during one of his tours of the western territories before arriving back in the Central States. The narrator's lively description of the automaton needs to be quoted in full before the reader can grasp the rich implications of this story:

> **(p.244)** King Mu received the craftsman and asked: "What can you do?"
>
> "Your Majesty may command what he pleases. But I have already made something, and I hope that Your Majesty will look at it first."
>
> "Bring it with you next time, and I will take a look at it with you."
>
> Next day, Master Yan asked to see the King. The King received him and asked: "Who is that man who has come with you?"
>
> "It is something I have created and it can act and perform."
>
> The King looked at it in amazement. The thing was striding quickly looking up and down, and one could not but believe it was a real human being. When the craftsman touched its cheek, it sang in tune; when he clasped its hand, it danced in rhythm; it did innumerable tricks, whatever it pleased you to ask. The King thought it was really a man, and watched it with his favorite Shengji and his other concubines.
>
> When the entertainment was about to end, the performer winked its eye and made advances to the concubines in waiting on the King's left and right. The King flew into a rage and demanded Master Yan's death on the spot. Terrified, Master Yan instantly opened up the body of the performer and took it to pieces to show the King. It was made of a conglomeration of leather, wood, glue and lacquer, colored white, black, red and blue. The King examined it closely; on the inside the liver, gall, heart, lungs, spleen, kidneys, intestines and stomach; on the outside the muscles, bones, limbs, joints, skin, teeth and hair, were all artificial but all very complete without missing a single organ. When they

> were put back together, it was again as he had seen it before. The King tried taking out its heart, and the mouth could not speak; he tried taking out its liver, and the eyes could not see; he tried taking out its kidneys, and the feet could not walk. The King was at last satisfied, and said with a sigh: “Is it then possible for human skill to achieve as much as the Creator?”He ordered to have it loaded into the second of his chariots, and took it back with him.”
>
> Gongshu Ban's ladder, which reached the clouds, and Mozi's flying kite were said to challenge the limits of human achievement. But when their disciples Dongmen Jia and Qin Guli heard of Master Yan's extraordinary feat, they told the two philosophers, who never dared to speak again of their skills even as they continued to carry their compass and square around.[82]

Besides being witty and ingenious, this ancient tale is fascinating not because it is remote and exotic, but because there is a startling familiarity **(p.245)** about it, a certain closeness to our own modern sensibilities. Today's engineers and critics continue to imagine the humanoid robots of the future in this vein—though in steel, silicon, and plastic, rather than wood, leather, glue, or clockwork—even as they set about building them with sophisticated tools while thinking to themselves that they are the first to dream such dreams. There will be successes, failures, obsessions, risks, and achievements, but the anxiety about the boundaries of the human in regard to machine remains just as strong as it was millennia ago. The *Liezi* story is open to interpretations and, in the Daoist context in which the story was composed, the automaton blurs those boundaries, thus destabilizing the self-image of the human in anthropocentric Confucianism even as it affirms the supreme artistic achievement of Master Yan.[83]

One would be very tempted to call the above-quoted story the proto-science fiction of the fourth century until one becomes alerted to an intertextual linkage to another, earlier Buddhist source *Jātaka Sūtra* in Chinese translation dating to the year 285 CE.[84] This earlier Buddhist tale also shows the king becoming jealous and wanting the performer to be dead, not knowing that it is an automaton that has thrown amorous glances at his queen. The craftsman—who calls the automaton his son—makes tearful pleas for mercy but the king refuses to change his mind. The craftsman then offers to put the automaton to death with his own hand rather than by someone else. The king grants the permission whereupon the craftsman removes a small mechanical part from the shoulder of the automaton, and the thing instantly disintegrates into 360 pieces of wood. Astounded to discover that the offender has been a wooden automaton, the king is thoroughly impressed by the craftsman's cleverness and rewards him with tens of thousands of gold pieces.[85]

This last detail about the sovereign's generous bestowal of gold may have been suppressed by the *Book of Liezi*, but the royal patronage of technological inventions remains intact in the later text, where power, seduction, masking/unmasking, life taking and life giving, as well as gift exchange, provide some rich social meanings that govern the technological fantasies about automata of the past, and many of these meanings continue to govern the fantasies and designs of robots in our own time. Of course, the **(p.246)** humanoid embodiment in the automaton from *Jātaka Sūtra* is profoundly linked to the Buddhist teachings on karma and the twelve-fold chain of codependent origination of all emergent beings. But in what ways might it be made relevant to today's robots?

In *Buddha in the Robot*, Masahiro Mori has advanced the bold claim that “robots have the Buddha-nature within them—that is, the potential for attaining Buddhahood.”[86] Mori is the Japanese robot engineer I discussed earlier who takes a strong interest in Freud and whose speculations about the Uncanny Valley in AI research in the 1970s have spurred robot engineers and filmmakers worldwide to wrestle with the psychic problems that trouble the design of the hands and faces of humanoid robots. Mori suggests that Buddhism and Buddhist sculpture point the way toward the possible overcoming of the Uncanny Valley. He argues that the “artist who makes the statues of Buddhas [has] created a model of a human hand that is made from wood. The fingers bend at their joints. The hand has no fingerprint, and it assumes the natural color of wood. But we feel that it is beautiful and there is no sense of the uncanny. Maybe [the] wooden hand can serve as a reference for future design.”[87] That raises an interesting issue as to what an AI engineer can learn from the artist of analog media and from religious sculptures in general and why it is essential for the artist to retain rather than eliminate the traces of nonhuman analog material (wood, metal, silicon, or whatever) and its ontological difference. McKenzie Wark remarks in *Gamer Theory* that the designer in digital media is subsuming the artist just as the digital subsumes the analog, but he also raises the possibility that “the artist within the designer may still inscribe the analog in the heart of the digital as something irreducible.”[88] Mori may or may not have known or identified himself with the analog artist of the wooden automaton in *Jātaka Sūtra*, but as a Buddhist he has good reason to insist on the irreducible when the media technology moves from analog to digital. Perceiving the psychic stakes in this movement, Mori repeatedly warns the scientific community about their narcissistic attachment to the self-image of the human and about the danger of falling into the Uncanny Valley.

(p.247) Will a Buddha-natured robot offer an alternative to the Freudian robot? We do not know, and it is difficult to imagine how humanoid robots or their engineers can acquire Buddha-nature when all they have been doing is dance on the capacious palm of Tathāgata while laboring under the enormous illusion that their somersaults have taken them faraway into some sublime realms of immortality outside of the sphere of Tathāgata's palm.[89] But we are not immortals, nor have we even reached the point where we can calmly contemplate the future of a Buddha-natured robot. Nevertheless, I realize that ruling out this potentiality would mean premature surrender to the narcissistic pull of the Freudian robot. What we could do, though, is maintain our vigilance and radical openness toward the future, bearing in mind that a historical and comparative perspective on world civilization—the hindsight of millennia-long human technological imagination and evolution—can always provide a good starting point for reflecting critically on where we are in relation to the Freudian robot. **(p.248)**

Notes:

(1.) Jacques Lacan, “Aggressiveness in Psychoanalysis,” 100–101.

(2.) Ernest Jones, *The Life and Work of Sigmund Freud*, 3: 276.

(3.) Freud relates the death drive to the compulsion to repeat and describes it as “an urge inherent in organic life to restore an earlier state of things.” See Freud, *Beyond the Pleasure Principle*, 36.

(4.) For the 1931 debate, see Siegfried Bernfeld and Sergei Feitelberg, “The Principle of Entropy and the Death Instinct”; Reginald O. Kapp, “Comments on Bernfeld and Feitelberg's ‘The Principle of Entropy and the Death Instinct’”; and L. S. Penrose, “Freud's Theory of Instinct and Other Psycho-Biological Theories.”

(5.) See, e.g., Leon J. Saul, “Freud's Death Instinct and the Second Law of Thermodynamics” (1958).

(6.) Anthony Wilden, *System and Structure*, 124.

(7.) Strangely, Lacan's discussion of cybernetics is disavowed by Wilden in his criticism of Lacan's linguistic approach. See Wilden, *System and Structure*, 19. For Wilden's critique of Marcuse, see his “Marcuse and the Freudian Model: Energy, Information and Phantasie.”

(8.) Marcuse states that “in the Freudian conception, destructive energy cannot become stronger without reducing erotic energy: the balance between the two primary impulses is a quantitative one; the instinctual dynamic is mechanistic, distributing an available quantum of energy between the two antagonists.” See Marcuse, “Aggressiveness in Advanced Industrial Society,” 257–58.

(9.) For a good discussion of this problem, see Bruce Clarke, “From Thermodynamics to Virtuality.”

(10.) Besides the Shannon exhibition at the Heinz Nixdorf Museum in Paderborn, Germany (see chap. 3 n 89, above), readers will find replicas of the Ultimate Machine on YouTube and other websites. One of these replicas appears in *The Human Language Series*, part 1, “Discovering the Human Language” (PBS, 1995), directed by Gene Searchinger. The machine appears in a brief interlude in the film and runs just long enough for the faux hand to complete its on/off action.

(11.) Clarke had proposed geostationary satellite communication in a technical paper published as early as 1945. But John R. Pierce pursued significant research into satellites when he served as the vice president for research of Bell Laboratories. Pierce was responsible for developing Telstar 1, the first commercial communications satellite, and was the first to discuss unmanned communications satellites. See John R. Pierce and A. Michael Noll, *Signals: The Science of Telecommunications*.

(12.) A. C. Clarke, *Voice Across the Sea*, 159.

(13.) See for example Hal Foster's psychoanalytic reading of modernist art in *Compulsive Beauty*

(14.) Sigmund Freud, “The ‘Uncanny,’” 241.

(15.) Françoise Meltzer is one of the few critics who question Freud's reading. See F. Meltzer, “The Uncanny Rendered Canny: Freud's Blind Spot in Reading Hoffmann's ‘Sandman.’”

(16.) Bill Brown, “Reification, Reanimation, and the American Uncanny,” 198.

(17.) Harold Bloom has extolled the essay on the uncanny as a unique contribution Freud made to the aesthetics of the sublime, and pays no attention to Jentsch's prior study whatsoever (Bloom, 101–4). Samuel Weber, Hélène Cixous, Sarah Kofrnan, Jeffrey Mehlman, Neil Hertz, Stanley Cavell, and Maria Torok have each offered insightful interpretations of Freud's reading of “The Sandman,” but when they refer to Jentsch, they usually do so by quoting Freud quoting Jentsch.

(18.) Forbes Morlock, “Doubly Uncanny: An Introduction to ‘On the Psychology of the Uncanny,’” 17.

(19.) According to Frazer, the moment of totemism occurs in a woman's passage to motherhood when a spirit, which awaits reincarnation in the nearest totem center where the spirits of the dead collect, has entered her body. Freud dismisses this explanation as “the sick fancies of pregnant women” and chooses to interpret the totem as the creation of a masculine mind trying to resolve the sublime conflicts of the Oedipus complex. In an insightful reading of *Totem and Taboo*, W. J. T. Mitchell points out that “Freud passes over Sir James Frazer's account of the accidental finding of the totem rather quickly” (*What Do Pictures Want?*, 122).

(20.) Anthony Vidler, *The Architectural Uncanny: Essays in the Modern Unhomely*, 11.

(21.) Vidler devotes a short paragraph to discussing Jentsch. Although he does not wish to confront Freud's repudiation of Jentsch's idea of “intellectual uncertainty,” Vidler clearly subscribes to Jentsch's view on this issue. See Vidler, *The Architectural Uncanny*, 23.

(22.) Ernst Jentsch, “On the Psychology of the Uncanny,” 11. This essay is translated by Roy Sellars from “Zur Psychologie des Unheimlichen,” which was originally published in the *Psychiatrisch-Neurologische Wochenschrift*, 203–5. For more on Freud s Jentsch, see Nicholas Royle, *The Uncanny*.

(23.) Freud, “The ‘Uncanny,’” 220.

(24.) Weizenbaum, *Computer Power and Human Reason*, 6–7.

(25.) See Victoria Nelson, *The Secret Life of Puppets*, 66.

(26.) Hoffmann, *Tales of E. T. A. Hoffmann*, 117. The translators render the word *unheimlich* as “weird” rather than “uncanny” in this translation, so I have put the German word back in the text to highlight its actual appearance in the original story. For the original story, see Hoffmann, *Der Sandmann*, 41.

(27.) For a well-grounded critique of Freud's intolerance of intellectual uncertainty, see Meltzer, “The Uncanny Rendered Canny: Freud's Blind Spot in Reading Hoffmann's ‘Sandman.’”

(28.) Freud, “The ‘Uncanny,’” 230.

(29.) Stanley Cavell, “The Uncanniness of the Ordinary,” 155, 156.

(30.) Brown, “Reification, Reanimation, and the American Uncanny,” 198.

(31.) Mitchell, What *Do Pictures Want?*, 54.

(32.) For Freud's ambiguous relationship with ocularcentrism, see Martin Jay, *Downcast Eyes: The Denigration of Vision in Twentieth-Century French Thought*, 332–36.

(33.) Freud, “The ‘Uncanny,’” 232.

(34.) Cixous, “Fiction and Its Phantoms: A Reading of Freud's ‘Das *Unheimliche*’ (the “Uncanny”),” 538.

(35.) Mitchell, *What Do Pictures Want?*, 53.

(36.) Mitchell argues that all three Lacanian categories—the Symbolic, the Imaginary, and the Real—converge in the picture, and I suppose that Lacan would have concurred that the slippage between the Symbolic and the Imaginary occurs all the time. See Mitchell, *What Do Pictures Want?*, 73–74.

(37.) Mitchell, *What Do Pictures Want?*, 49.

(38.) Freud, “The ‘Uncanny,’” 244.

(39.) See Samuel Weber, *The Legend of Freud*.

(40.) Cixous, “Fiction and Its Phantoms,” 530.

(41.) Freud, “The ‘Uncanny,’” 244.

(42.) Hoffmann, *Tales of E. T. A. Hoffmann*, 98. The original German sentence goes “Und damit faftte er mich gewaltig, daft die Gelenke knackten, und schrob mir die Hande ab und die Fiifte und setzte sie bald hier, bald dort wieder ein.” Hoffmann, *Der Sandmann*, 16.

(43.) Hoffmann, *Tales of E. T. A. Hoffmann*, 124. The original German quote is “Nathanael fafite *mechanisch* nach der Seitentasche; er fand Coppola's Perspektiv, er schaute seitwarts” (Hoffmann, Der Sandmann, 48; emphasis added). The emphasis draws attention to his mechanical movement as indicated by the German phrase.

(44.) The German phrase is “Holzpüppchen dreh dich—Holzpüppchen dreh dich,” in Hoffmann, *Der Sandmann*, 49.

(45.) Cixous, “Fiction and Its Phantoms,” 525.

(46.) Lacan, “Aggressiveness in Psychoanalysis,” 85.

(47.) Freud, “The ‘Uncanny,’” 242–43.

(48.) Freud, *The Interpretation of Dreams*, 536.

(49.) Jay, *Downcast Eyes*, 435–91.

(50.) See the discussion of Kittler and Lacan in the preceding chapter. I have argued that Lacan models the symbolic order on the communication machine rather than on natural language, thereby profoundly transforming the Freudian notion of the unconscious.

(51.) Mitchell, *What Do Pictures Want?*, 70.

(52.) Masahiro Mori, “Bukimi no tani” (The Uncanny Valley), 35, translated by Karl F. MacDorman and Takashi Minato in appendix B in MacDorman, “Androids as an Experimental Apparatus: Why Is There an Uncanny Valley and Can We Exploit It?”10.

(53.) Mori, “Bukimi no tani,” 34; MacDorman, “Androids as an Experimental Apparatus” (appendix B, translation of Mori), 9.

(54.) MacDorman, “Androids as an Experimental Apparatus: Why Is There an Uncanny Valley and Can We Exploit It?”8.

(55.) See Frank Hegel et al., “Theory of Mind (ToM) on Robots: A Functional Neuroimaging Study.”

(56.) Mitchell, What Do Pictures Want?, 321.

(57.) For related simulation programs in cognitive science such as Terry Winograd's SHRDLU. see Margaret A. Boden, *Artificial Intelligence and Natural Man*, 114–42.

(58.) Weizenbaum, Computer Power and Human Reason, 3–4.

(59.) Hoffmann, “Automata,” 92.

(60.) Weizenbaum, *Computer Power and Human Reason*, 190.

(61.) On Weizenbaum's critical reflections on the unexpected popularity of ELIZA, see the discussion in the next chapter.

(62.) K. M. Colby et al., “A Computer Method of Psychotherapy: Preliminary Communication,”151

(63.) Margaret A. Boden, Mind as Machine: A History of Cognitive Science, 1:369.

(64.) Colby et al., “A Computer Method of Psychotherapy,” 150. I have converted the capitalized letters of the human participant from the original quote to lowercase letters to prevent confusion.

(65.) The eight transforms are (1) DEFLECTION: Shift Object (Not Self); (2) SUBSTITUTION: Cascade Verb; (3) DISPLACEMENT: Combine (1) and (2); (4) NEUTRALIZATION: Neutralize Verb; (5) REVERSAL: Reverse Verb; (6) NEGATION: Insert Not Before Verb and Do (5); (7) REFLECTION: Shift Object to Self; (8) PROIECTION: Switch Subject (Self) and Object (Not Self). See Margaret A. Boden, *Artificial Intelligence and Natural Man*, 50.

(66.) The data were generated from tape recordings and notes from natural language utterances in psychotherapeutic dialogues. See Colby, “Simulations of Belief Systems,” 258.

(67.) Colby, “Simulations of Belief Systems,” 266.

(68.) Colby, *Artificial Paranoia: A Computer Simulation of Paranoid* Processes, 99.

(69.) Colby, Weber, and Hilf, “Artificial Paranoia,” 5.

(70.) Colby, Weber, and Hilf, “Artificial Paranoia,” 16–17.

(71.) Others have contested the claim. Boden argues: “Strictly, this was not a Turing test in the accepted sense, since the psychiatrists doing the interviewing were not asked to judge which teletypes were attached to people and which to machines. They were not even informed that an ‘imitation game’ was going on.” See Boden, *Artificial Intelligence and Natural Man*, 500.

(72.) Francisco Varela, Evan Thompson, and Eleanor Rosch, *The Embodied Mind*, 41. Douglas R. Hofstadter's discussion of symbols' status as the software or hardware of the brain raises similar

issues of metaphorical substitution of terms. See Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*, 356–62.

(73.) Freud, *Jokes and Their Relation to the Unconscious*, 125.

(74.) See his discussion of McCulloch and Pitts in Minsky, *Computation: Finite and Infinite Machines, 32–66*.

(75.) Minsky, *The Society of Mind*, 184.

(76.) Minsky, *The Emotion Machine*, 88.

(77.) Marvin Minsky, "Jokes and the Logic of the Cognitive Unconscious," 175.

(78.) See Piaget, "The Affective Unconscious and the Cognitive Unconscious," 250.

(79.) Minsky, introduction to Marvin Minsky, ed., *Robotics*, 16.

(80.) Minsky, "Our Robotized Future," ibid., 298.

(81.) Minsky, "Our Robotized Future," 303. The kinds of futurist prophecy made by Ray Kurzweil in popular culture often reiterate Minsky's views and theories. See Kurzweil, *The Singularity Is Near*.

(82.) The English translation is taken, with a few modifications, from A. C. Graham, trans., *The Book of Lieh-tzü*, 110–11.

(83.) For a Daoist interpretation of the *Liezi* story of the automaton, see Jeffrey L. Richey, "I, Robot: Self as Machine in the Liezi."

(84.) This story is found in volume 3 of the Chinese *Jātaka Sūtra*.

(85.) A French translation of this Chinese Buddhist tale exists. See Édouards Chavannes, trans. *Cinq cents contes et apologues: extraits du Tripitaka Chinois*, vol. 3, 170–72. Chavannes's translation has left out the specific reference to number 360, which is mentioned in the Chinese text.

(86.) Masahiro Mori's argument in that book represents a cybernetic recasting of Buddhism rather than a systematic rethinking of cybernetics in light of Buddhist philosophy, which is disappointing to me. See Mori, *The Buddha in the Robot: A Robot Engineer's Thoughts on Science and Religion*, 13.

(87.) Mori, "Bukimi no tani," 35; MacDorman, "Androids as an Experimental Apparatus: Why Is There an Uncanny Valley and Can We Exploit It?"(appendix B, translation of Mori), 10.

(88.) McKenzie Wark, *Gamer Theory*, 098.

(89.) 1 am alluding to the story of monkey king Sun Wukong from the sixteenth-century Chinese novel *Journey to the West*. This monkey king has numerous magical powers and rebels against divine authority: He is capable of seventy-two metamorphoses and his cloud-somersault can take him 54, 000 kilometers in a single leap. He is defeated, however, by Tathāgata in a wager when

he fails to jump out of the latter's palm. For a delightful account of this story, see Wu Ch'eng-en, *Journey to the West*, vol. 1, 173–74.