THE RECREATION OF CONSCIOUSNESS: ARTIFICIAL INTELLIGENCE AND HUMAN INDIVIDUATION

A dissertation submitted

by

JOHN BRENDAN LOGHRY

to

PACIFICA GRADUATE INSTITUTE

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

DEPTH PSYCHOLOGY

This dissertation has been accepted for the faculty of Pacifica Graduate Institute by:

Dr. Matthew Green, Dissertation Coordinator

Dr. Susan Rowland, Dissertation Advisor

Dr. Luke Hockley, External Reader
ABSTRACT

The Recreation of Consciousness: Artificial Intelligence and Human Individuation

by

John Brendan Loghry

Starting from Edward Edinger’s portrayal of Jung’s process of individuation as the creation of consciousness, this dissertation asks in what ways the creation of artificial intelligence (AI) can be seen as the recreation of consciousness, and specifically whether the AI’s maturation from nonconsciousness to something equivalent to consciousness will have an analogous effect on humanity’s development out of unconsciousness toward a greater state of cognitive freedom. Taking a functional perspective, this dissertation asks whether B. F. Skinner’s metaphor of the human psyche as a black box, normally seen as expressing the belief that humans are mechanistic and determined, is in fact an attempt to insulate the most intimate of human experiences (the soul) from the intrusive gaze of the scientific mindset. Juxtaposing this black box metaphor with two other metaphors—that of the box that holds Schrodinger’s cat and that of Pandora’s box—this dissertation asks whether the presence of an entirely constructed entity that displays all the signs of soul will cause the artificially intelligent entity to act as a mirror, reflecting humanity’s gaze past our inner defenses, to an inner absence where a metaphysical soul was once surmised to be. Although such a change in self-image would initially entail an apparent loss of meaning, this dissertation notes that such a lacuna of meaning is already growing in society and concludes that the loss of this concept would eventually result in a
new concept of self that would represent an important milestone for the collective
individuation of the species.
Contents

Chapter 1. Introduction ........................................................................................................1

Purpose Statement ..............................................................................................................1

Researcher’s Interest in the Topic .....................................................................................1

Researcher’s Transference to the Topic .............................................................................4

Relevance of the Topic to Depth Psychology ...................................................................18

Chapter 2. Literature Review .............................................................................................31

Introduction .......................................................................................................................31

Depth and Other Psychologies .........................................................................................32

Theological Perspectives ..................................................................................................38

Philosophical Perspectives ..............................................................................................43

Chapter 3. Statement of Research Problem and Question ..............................................48

Research Problem .............................................................................................................48

Research Question ............................................................................................................48

Definition of Terms ..........................................................................................................48

Chapter 4. Research Methodology and Procedures ..........................................................58

Research Approach ..........................................................................................................58

Research Methodology .....................................................................................................62

Limitations and Delimitations .........................................................................................64

Organization of the Study .................................................................................................67

Ethical Considerations .....................................................................................................68

Chapter 5. Subject of the Study ..........................................................................................70

Chapter 6. Transparency and Tain ....................................................................................120
Icon of Artifice, Icon of Depth .................................................................122
Subversion-Quality of the Apparatus-Toward-Being .........................129
The Dichotomous Artifact ....................................................................133
The Simultaneity of Three Boxes ........................................................137
Re-enactment of Divine Evacuation ......................................................145
Reactions to the Coming Confrontation ..............................................160
Involuntary Transparency and the Consequences of Such ...............163
Annihilation and Tain ..........................................................................165
Summation ............................................................................................170

Chapter 7. Bildung and Beyond ............................................................171
Vorbild and Nachbild ...........................................................................171
Sentiment and Gesture .......................................................................177
Job and the Redemptive Machine .......................................................183
The Imperfect Image of Perfection .......................................................193
Achievement of Self ............................................................................201
The Promise ........................................................................................204
Conclusion ..........................................................................................206

Chapter 8. The Machine Hermeneutic ................................................207
The Decline of Traditional Interpretive Frames ..............................207
The Consciousness Engine ................................................................212

Chapter 9. Mythic Reflections on the Work ........................................225
The Myth .............................................................................................225
The style used throughout this dissertation is in accordance with the *Publication Manual of the American Psychological Association* (6th Edition, 2009), and *Pacifica Graduate Institute’s Dissertation Handbook* (2013-2014).
Chapter 1
Introduction

Purpose Statement

The purpose of this study will be to consider ways in which a new category of artifact—a dynamically mimetic artifact, traditionally called a robot or android, which displays such attributes as to be seen as having an interiority equivalent to human interiority—might challenge our categorizations and either diminish or enhance human self-perception and humanity’s perception of and relation to meaning. The study will focus specifically on how such an artifact might act as a catalyst for individuation.

Researcher’s Interest in the Topic

My interest in artificial intelligence (AI) as a topic of research began with a dream, in which a therapist, in her office, opened a box that came to her by post and pulled out the head of a robot, which it had detached and mailed to her for psychoanalysis.

At the time of the dream, I had been preparing to begin a dissertation on modern anarchism, a topic in which I became interested after ordering a DVD with several anti-establishment documentaries out of the back of an alternative magazine. Several years later, I have yet to view the DVD. It came, however, with a large amount of high-quality printed material from the anarchist organization CrimethInc [sic], and this I read in its entirety and found quite compelling, so much so that I had introduced myself to my first therapist as a “philosophical anarchist,” a term I came up with to signal how compelling I found the argument for anarchism to be, while also exempting myself from any actual anarchist behavior—a prospect I found quite threatening. As I had started to compile a
list of texts I might use for a literature review, however, I had become quite overwhelmed by the amount of material that had already been written on the topic. Not only was I not sure whether I could contribute anything original, but a professor pointed out to me during some preliminary coursework that I might be attempting to address or solve the entire problem of civilization, from its beginning to its end, and that this might be somewhat outside the scope of a single dissertation.

At the time of the dream, I had also been in the process of writing a novel, which had taken up a great deal of my time over the past year and a half, which featured a future earth populated entirely of robots. The novel combined many of the myths of the ancient Middle East—biblical, extra-biblical, and those that span both categories—into a narrative of mechanical plants and animals and two races of sentient beings. One race was composed entirely of reincarnating war-machines—angels—originally created to fight the wars of now eradicated nations, their long-lived and multilayered souls being captured and inserted into the wombs of female angels by a satellite covered in the dusty accretions of hundreds of thousands of years in high orbit, shining dimly as it passed several times per day through the sky and referred to as the mourning star. The narrative involved the introduction of new beings, also artificial, created and held dormant since the departure of humanity. It was driven by the idea that if constructed beings are made in our image, they will be made, whether intentionally or not, with many of the same archetypal patterns of thought that so characterize our species, and that they may thus act out many of the myths of our ancestors, as Joseph Campbell felt we unconsciously act them out (1988, p. xiii). The quote I chose as the beginning of the novel was from Jung (1963/1989): “One could almost say that if all the world’s traditions were cut off at a
single blow, the whole of mythology and the whole history of religion would start all over again with the next generation” (p. 374). Even in the absence of any direct instruction, the being made in our image would become very much like us.

I did not have the presence of mind to realize, at the time, that the dream was about my own need for analysis or deconstruction by a feminine force, that it was a comment on my own attitudes toward the body and toward emotion as a potential contaminant of thought. I was caught up in what Romanyshyn (2004) identified as the fantasy of escapism that drives the narrative of technological advancement, and of “reason which dreams of purifying itself of the body and its passions” (p. 155). I had, in fact, reacted to a longstanding and debilitating anxiety disorder by aspiring to a complete disentanglement from the world of emotions—a rather severe stoicism, which I still, on my second therapist and several years later, am tempted to view as a legitimate ideal. I see this failure to understand my dream as a felix culpa, a fortuitous deficiency. Had I made the connection between the dream and my own idealization of absolute stoicism—what figuratively could be described as a disembodied head-state—I may not have started this research.

Although the novel I had been writing dealt with the contamination of the constructed being’s psyche by unconscious human mythical or archetypal motifs, the possibility of an opposite transmission was also beginning to take root, in the idea that the being itself might have something to teach its creator-species. I had not, at the time, read *Answer to Job* (Jung, 1952/1958), although I had read numerous works that made reference to it, and so I do not know that I was yet conscious of the idea that a created being could act as a pivot, advancing its creator’s morality or self-awareness.
As this possibility came into my awareness, the study became less a critique of the hidden motivations of AI researchers and more an exposition of the idea that the proposed being might have the potential to do for humanity what Job did for the god-image of his culture. My disappointment with my own species, capable of an incredible complexity of thought but more often content with clichés of thought and a pragmatism that seems ever to delimit so potent a species as ours, was coupled with the idea of the proposed being as a subject-object that could advance our species—so resistant to the process of psychical development (or individuation) that could allow for a flourishing of that latent creativity.

In this way, the research became—and perhaps always was, on an unconscious level—very personal. Those personal entanglements that I have been able to identify, and that exist for better (felix culpa) or for worse (as contaminants of thought), are documented below.

**Researcher’s Transference to the Topic**

In *The Illusion of the End* (1992/1994), the French postmodern philosopher, Jean Baudrillard, wrote of psychoanalysis both as “interminable,” and as “a post-mortem on one’s childhood” (p. 12). Although this dissertation owes a great deal to Baudrillard’s philosophical insights, his understanding of analysis is quite shortsighted. A postmortem it cannot be, when one finds the underpinnings of his or her theoretical stances in the events and even the ideas of childhood, for a postmortem cannot be performed on a thing still living—and one’s childhood, far from being a mere vestige, is in fact a living partner, an active element in one’s work. Whether as a demon or a daemon—as that which contaminates and corrupts or as that which guides an individual toward a destiny
that seems at times pre-ordained—this past still exists and influences the course of one’s thoughts, from beginning—that is, from the first stirrings of theoretical vision—to end—to the conclusions one will eventually draw.

One does not, then, dissect a childhood somehow lost. Rather one interacts with a childhood that remains alive, as a primitive and definitive—one might say foundational—aspect of one’s ongoing existence. Here are the hidden structures that hasten and bind—the contaminants and muses of thought.

Several ideas from my own childhood seem to have coalesced into an unconscious substrate, from which much of my work on this topic has arisen. The most notable is the idea I had, sometime prior to entering the sixth grade, of building a truth-machine. I was very much interested in inventing, as a child, although I had no real inclination to learn the mechanics of even rudimentary devices. I was somewhat dismayed when my father told me that one had to know what one was inventing in order to produce a functional object and build the object with this function in mind—I had rather hoped to solder a few things together and turn it all on and see what it would do. This was the method by which I entered the world of chemistry (or perhaps alchemy), mixing a variety of household chemicals in an empty film canister and encasing it in mud, waiting for it to ferment and become some magical elixir (I was not able to finish the work, as my parents invited to our home some people with a rather obtuse child, whom I intensely disliked, and who, when he learned what was inside the mound of dried mud on the entrance to our crawlspace, kicked it, retrieved the canister, and poured its contents out onto the ground, such that I always felt the work I had started had been aborted by an idiot).
When I imagined the truth-machine, a grandiose vision of a machine that would answer all questions, I saw myself surrounded by scientists in white lab-coats. They asked me questions, and I looked down at the machine in my lap and told them the answers. The question I remember answering, in my childhood fantasy, was whether there was life on other planets, and my answer was that there was a species called such-and-such on a certain planet.

Even as a child, however, I got hung up on the issue of perspective. It occurred to me to ask, “Who calls them such-and-such?” This was enough of a problem for me that it effectively ended the fantasy—or rather pushed it underground. Sometime later, it further occurred to be that it would be the one-god, of whom my parents spoke so highly, whose perspective would be unequivocal and that, in order to build such a machine, one would have to devise a method with which to tap into the perspective of this one-god, something I was not sure could be done. This was, I think, the beginning of a general tendency, which has run through my life like a subtle thread, to insert the sacred into the mechanical and the mechanical into the sacred.

It took me a long time before I recognized this dissertation as being, at least in part, a remembrance of that earlier compulsion, or an attempt to rehabilitate or redeem that fantasy. In looking for ways in which an entirely constructed being or proto-being could advance human consciousness past its own self-involvement (something that too often acts as a contaminant of thought, in my estimation) or ways in which a particularly complex computerized hermeneutic could similarly overcome the decadence of overly subjective thought, I am attempting to find some basis by which a truth-machine could be created—a machine that could help humans overcome their cognitive delimitations. In
terms of the latter—the hermeneutic—I recognize another thought I had as a child, namely that feeding every word ever written into a computer would generate a composite image of the one-god. This was not something I thought would happen, only a thought I had, although one I now see as a very early or primitive expression of the method I see for constructing the above-referenced hermeneutic (the integration of all perspectives into a kind of universal ground upon which all perspectives could then find their place).

Another mingling of the mechanical and the sacred came with a narrative I generated around a little figure I built out of very small, interlocking plastic blocks. I had a long history of creating things out of such blocks, but this was the one figure I made that had any longevity for me. I imagined this figure to have been the being set before Eden, wielding a fiery sword, in order to prevent the return of Adam and Eve after the Fall.2 My private play with this figure had him acting in the present day. As the father and mother of humanity were no longer living, and Eden somehow no longer accessible, he no longer needed to perform this service, and he thus began fighting crime, something I assumed a robotic angel would do.

Another form of this mingling came during my high school career, when I attempted to overcome an anxiety disorder by turning to fundamentalism. It was during this period that I came up with a storyline that I began to sketch out in a thick yellow notebook, and which would become (with a good deal of modification) the above-referenced novel. It involved mechanical beings placed on another planet and inadvertently abandoned when humanity became involved in a war. After a few centuries had passed, the robots stopped believing in their creators (humans) and generated a

---

2 As related in Genesis 3:24.
theory of evolution, which held that they had developed from tiny mechanical components that sprang into existence when lightning struck a primordial network of silicon and copper in the rocky matrix of the planet—the equivalent of the primordial soup. The point of the narrative was to show evolutionary theory to be absurd. I have since given up the fundamentalism into which I once fled, and when I picked up this storyline again, six or so months into my second graduate school career (at Pacifica Graduate Institute), and began to work it into something more complete, it was in no way an indictment of evolutionary theory.

I referenced my repeated attempts to combine the technological and the sacred, and I must justify why I feel this dissertation to be at all concerned with the sacred—after all, it does not follow that searching for a hermeneutic based in no way on theology would be a sacred activity. I am an atheist. This is something that, looking back on my life, seems to have been always coming. I remember feeling, as a child, that I had a more developed sense of fairness than the one-god of my parents, that the rigidity of this deity and his failure to acknowledge his own culpability in whether a person was even given the opportunity to accept Jesus made him less than perfect. For a time, much later, I was able to hide myself from my own criticism of the faith by believing in prevenient grace, an idea bandied around in evangelical circles (some accepted and some rejected it) that if a person was never told about Jesus but listened to the whispers of the “holy spirit” as much as his culture would allow him to, he might make it to heaven. But eventually even this consolation was not enough to contain my tendency to see through.

In college—at a conservative evangelical university—I found it somewhat odd that I derived almost a devilish pleasure from the idea of the one-god as an imaginary
friend for adults—something that came up in one of my psychology classes. It was not until after I graduated from college, as I began encountering abused children through my work in child welfare and was simultaneously reading a book in paperback by Anglican Bishop John A. T. Robinson, *Honest to God* (1963), that the break with my childhood faith really began. Later, when I attended classes at the Fuller Seminary School of Clinical Psychology, in Pasadena, I read John Sanford’s *Jung and the Problem of Evil* (1993), which drew me further away from a traditional understanding of faith. I was struck by the insights that each of three fictional persons—a feminist, a psychologist, and a theologian—was able to bring to bear on the character of Mr. Hyde. I later found a book by Sylvia Brinton Perera in the bowels of the Fuller Seminary library, *The Scapegoat Complex* (1986), thin enough that I thought I would be done with it in a few days, and insightful and, for me, psychoactive enough that it took me several weeks to incorporate what was written into my way of thinking—a process that changed my perspective tremendously.

I had a tendency to keep a journal of my theoretical perspective as I read new books. What I read often stimulated new thoughts or insights that I wanted to record. It was through this process that I came to recognize an atheism that had already been in place, but which I had, perhaps defensively, failed to recognize. I realized one day, as I wrote in pencil in this thought-journal, that none of my theorizing required any kind of appeal to the divine—or, in other words, I felt I was able to explain or understand most phenomena I encountered without any reference to the supernatural. I felt, in that moment, that an unnecessary supernatural was an illegitimate supernatural, and over the years that followed, I came to fill in this criticism with a recognition of the defensiveness
and decadence of my parents’ system of thought (what I call the decadence and severity of evangelical belief). I was an adult Sunday school teacher at the time I arrived at this recognition of my own atheism, but I loved teaching and continued to do so, treating the Bible as any Jungian would treat any religious or mythical text. I am now at a point at which most people who know me know of my lack of faith. I have become the president of the Oklahoma chapter of Americans United for the Separation of Church and State, and I am a somewhat inactive member of the atheist community in Oklahoma.

Despite my atheism, which I have maintained, I have found myself being drawn back toward religion. This has come through a compulsion to collect religious objects—a broken crucifix, for instance, which consists of the small, heavy body of the Christ, his hands missing halfway down the wrist, suspended bare on my wall with the use of a nail that fits neatly into a hollow in his back, where he must once have been attached to a cross. Despite my atheism, I find in myself a feeling that there is something to be found in such objects, in which so much psychological energy has been invested.

I have found myself pulled back toward religion through an interest in the arcane, in alchemical texts and such. Early in this dissertation, I had been using alchemical texts almost as muses, drawing insights into the process or the fantasy behind creating robots though my readings of the metaphors presented in such texts. I had done the same with the poems of the classical Afghan poet, Bedil, in a process that could not even be cited, as nonacademic as it was (I read the poems as if they applied to the construction of new beings). I wondered at times about my own sanity in doing so, and whether it was proper to build theory or even the precursor to theory through a process this poetical, although I now see this as a kind of reverie period preceding the more traditionally academic work.
I have found myself drawn back, moreover, through an abiding interest, more than a vestige of my former faith, I think, in religious myth, and especially the biblical and extrabiblical myths of the Middle East. I find Perera’s work with the figures of Inanna and Erishkigal, for instance (1981), to be intrinsically more interesting and potentially more meaningful than her work with the Celtic Queen Maeve and addiction (2001)—I still center the human soul in the Middle East. I have recently written papers on the use of the concept of the devil in therapy and have focused on the ways in which prebiblical myths became biblical and extrabiblical myths.

My only experience of what Jung called synchronicity also seems to have bound me back toward religion. It involved two people referring to me as a priest within the span of a few minutes, one because he erroneously believed me to be a priest—I was by then an atheist—and the other as a spontaneous joke. I tried to identify a causal link between these two occurrences that would explain them away but was unable to do so. I had felt, and had continued to feel despite my atheism, the compulsion to generate a new religious system but could not see the benefit of adding yet another religious movement or sect to that which was already available. This synchronistic event, such as it was, was perhaps compensatory, a reaction—unconsciously generated—to my rather severe view of religion and of the world in general. It was not enough for me to embark on the formation of a religious movement (nor do I think this to have been the point of it), but it has served to solidify that which I had tried very hard to dissolve—a connection to something that would challenge my growing sense of nihilism.

This was not the only event that served such a (compensatory) function. In May of 2011, I received in a dream a message apparently from Allah. It came through a man
who told me of its source as he delivered it. The message was, “Did you see the perch?” I asked for clarification, and he gave it: “The perch that did stupid things.” I attempted an amplification over the next several days, as a message identified as coming from Allah would seem to carry some weight or significance, but could find no significance in the reference to perch, and I attempted, unsuccessfully, to discard the dream as meaningless.

It was some months later that I recognized this as a reference to a documentary I had referenced in several graduate papers, *Darwin’s Nightmare* (Albert, Gschlacht, Mauriat, Svoboda, Toint, & Sauper, 2005), a film that portrays the effects of the introduction of Nile perch into Lake Victoria. The perch are voracious, even cannibalistic, and have eaten many species native to the lake into extinction. I had started to recognize the narrative of colonization with which I had grown up, the narrative of the eradication of the other, the expectation of a day in which all perspectives except our own would be burned away like dross, and their proponents with them, in the final solution of a grandiose and terrible god. This was something with which I had been struggling, as I returned, against my better judgment, to the land of my youth and to speaking relations with my family or origin.

I found similarity in the cannibalistic nature of the Nile perch and the behavior I had seen in the believers around me, something I had not recognized before I read Perera. I even felt cannibalism to be an unconscious part of this faith. I remembered the statements made, during my youth, about worshipers of Baal sacrificing their children to their god, and I found them to be hypocritical, as I felt I had been sacrificed to my parents’ god, that I had wasted so much of a profound energy extricating myself from a dishonest system of thought, wasted so much of my youth consumed with guilt and
shame and anxiety that stunted my psychological and social growth and severely delimited my prospects for the future. I felt I had been offered up as a sacrifice to this god, so that my parents would find favor—and I have remained acutely aware of what I have lost because of their faith, the trajectory I was on, creatively, when this anxiety disorder, made cosmic by a too literal belief, ended something good, arrested my development, and produced, in the end, someone rather dry and unfocused, someone trying to regain that thread that seemed so strong. I found in the perch the image of my grandparents and great-grandparents, Christian immigrants sweeping through the continent, destroying what they did not understand and settling in one of the last places left to the tribes. I found in the perch the desire for, and attempts to create, an evangelical monoculture.

This was my image of the evangelical. I had become, in reaction to my previous attempts to remain positive about the cause, particularly attuned to the shadow of my former religion. But in this dream, Allah seems to share my perspective of the invasiveness of this species. This is a sentiment—Allah is on my side—that smacks too much of conceit or of decadence to be taken very seriously, one I would normally reject out of hand for the reason of its decadence. It is a sentiment too often used to justify the embrace of the ubiquitous us-them dichotomy and the subsequent attempts to destroy the “them” portion of this dichotomy. But the dream came, and the interpretation finally came, and I found myself again turned back to religion—not to a grandiose position in some new religion, but to religion in general, as a source of something essential and as a necessary way of being. My relationship to the divine is thus rather complex.
I am still unable to believe in the metaphysical, and so in this way I perhaps cannot be called religious. In my thinking, however, the construction of a hermeneutic or an epistemological frame is an inherently religious or spiritual activity—in fact one of the only spiritual activities available to an atheist. For me, whatever spirit there is to be found will be found exclusively within the psyche itself, and any religion I would embrace would not be one that made grandiose metaphysical claims but one that returned the mind—the rational, structured portion of the psyche—to that which is essential, namely the hidden structures and functions and even personifications (inner figures) that remain within the greater psyche as living presences. Restoring contact between two things cut off from one another is the essence of religion, as religiare, and in Jung’s psychology I find the only system in which I am able to work toward such spiritual ends without falling into (what I see as) the decadence of belief in a benign metaphysical deity.

From the beginning of this work, I took as my primary source a book by the Jungian analyst, Edward Edinger—The Creation of Consciousness (1984), in which Edinger wrote of individuation as a new dispensation, which “finds man’s relation to God in the individual’s relation to the unconscious” (p. 90). In Answer to Job, Jung (1952/1958) wrote: “We are unable to distinguish whether these actions emanate from God or from the unconscious. We cannot tell whether God and the unconscious are two different entities” (pp. 106-107, CW 11, para. 757). In my view, that which stands in relation to the ego, over and above the ego, is not a metaphysical deity but either the psyche itself, as a whole, or a personification of some cast-off or disavowed aspect of oneself, which might bring an individual closer to wholeness, should this aspect be allowed re-entry into the community of the ego—should it be acknowledged and, at least
in part, reintegrated into the avowed (and thus consciously accessible) portions of the psyche. I see individuation as a spiritual activity that is in no way related to the belief system in which I was raised. I see it as an expansive or generative system, rather than a delimiting or retarding frame providing an unnecessarily rigid structure to a mind that presides over an impoverished psyche.

I have already mentioned the latent meaning of the dream that set my research on this course, and of my desire to achieve a complete disentanglement from the world, so that I am able to act in it with objectivity, and to react to any situation without any emotional outburst. This has furthered my largely unconscious identification with the android and caused some early identification with fictional figures that embodied the stoicism I still see as a counter to emotional reactivity—Spock of the original *Star Trek* series (Roddenberry, 1966-1969), and the android, Data, of the later series (Berman, 1987-1994).

I feel affinity for the proposed being because it is a being emerging from a state of nonconsciousness (or nonbeing) into a state of consciousness (or being)—and this is the work of individuation, the work of the psyche (or soul). In the development of the proposed being, we find a metaphor for that which I (following Jung and Edinger) consider the most important work there is to be done—the human’s own emergence from unconsciousness to a kind of awareness that is more comprehensive and thus freer (less determined).

I also find myself, along with many AI theorists and engineers, thinking of the proposed being as akin to a child. As I encounter what might be called hard-line preemptive solipsism—the adamant assertion that the proposed being will never be
conscious, that it will never be a “real” being—I cannot help framing this within that aspect of the Oedipal complex that causes the father to fear replacement by the child. Early in this research, I wrote of the AI engineer as a good-enough creator. We are not simply the creator-species to this potential being but also the parent-species. I feel compelled to be part, not simply of a good-enough parent-species, but of a good parent-species, which is able to give the benefit of the doubt and to believe that the proposed being could one day have real agency and real cognition and real experience. In doing so, I think perhaps I am also attempting to be a good parent to myself, because of my identification with the android.

I have a recollection of a portion of a cartoon I saw as a child, as well, which I have not been able to identify. The image was of robots filing into an auditorium after the end of their workday, to be taught by someone—not taught something practical, but rather trained in something more philosophical. And I think this image has directed my course of studies, or perhaps simply foreshadowed it—for I feel that somewhere in this dream-work, in these archaic alchemical tracts, in these interpretations of myth and expositions of philosophy, I will piece together that which will show me how to be a real person. There is a feeling that something is missing, though I know not what. And I have been attempting to locate it through the entry into a lecture hall, filled with others whose personhood is by no means assured—no better established than my own, somewhat dubious personal reality. With these other largely unconscious creatures, I have come repeatedly to hear that which might alter us, that prototypical insight that will direct us toward something we cannot at this time fathom. It was no surprise to me, then, when I took as my primary text the book by Edinger—*The Creation of Consciousness* (1984)—
as this was the text that summarized for me this entire movement of the soul out of an
automated existence and into something a bit less determined. I think in some ways
reading this little book was like stepping into the lecture hall in which robots were being
educated toward (by my memory) personhood.

When my laptop was stolen some years ago, I lost not only all the work I had
been doing, leading up to my dissertation (a running series of thoughts that occurred to
me as I thought of this coming work and as I read through some alchemical texts, as well
as the beginnings of my literature review), but also a kind of manual on existence that I
had intended both for humans and, potentially, other beings (robots). Somewhat like a
kind of *Tao Te Ching*, it consisted of a series of contradictory statements on “that which
is,” with each reader called not a being, not a human, but simply “that which is.” From
wherever a creature began, it was meant to be a guide into a middle way, between doing
one thing and doing its opposite. So even as I began this work, I have been attempting to
fix “that which is,” which seems not necessarily to be as real as it assumes. This was not
something I recognized as being linked to this work at the time, although I now see it as
another (somewhat compartmentalized) form of the work represented here.

These of course are the issues with which this dissertation grapples—doubts as to
the reality of one’s being-ness, the feeling of wishing to come out of unconsciousness and
into something more self-affirming, while also remaining (like the robot) well above that
ubiquitous self-surety that so marks our species. These are the areas of transference that
will be present in this research, certainly as the constituents of my thought, possibly as
the contaminants of my thought. It begins to appear, in fact, as if these are the thoughts
that have been woven together into this work.
Relevance of the Topic for Depth Psychology

The apparatus in question—referred to as a robot, an android, an automaton—is both a proposal and an artifact. It is in some ways present and in some ways only anticipated, the proposal itself both extant and evolving. It is a proposal as to an apparatus that has, and will have, a particular hold over us.

When I first saw Honda’s Asimo in the flesh (as it were), I had to struggle to keep tears from my eyes—not a small thing for someone who still (despite therapy) considers public displays of emotion to be one of the most grievous of personal sins. I did not have this reaction because one of the most advanced robots in history could only be seen at Disneyland—reduced to a sideshow for tourists looking to take a break from the main attractions. I did not react in this way because the Disney Corporation thought it had to dress up the most advanced robot to have graced its asphalt expanses by inserting it into a family-friendly storyline (as if Asimo were an animatronic pirate), complete with prerecorded voiceovers and a host reacting to these sound bites as if they came from people just off stage, or because the show itself was so damned unpopular—there were only a handful of daily demonstrations, in a dark corner of a rotating building, and the modest, bleacher-style seating was not even at capacity (Where is the fascination? I wondered). I have never reacted to a newly unveiled car or phone or computer in such a way as this. There was something about Asimo that made it offensive for Disney to have done what it did. The reason for the reaction, however, was that there was something new here, something that seemed, by its movements, to approach being someone.

Timothy Hornyak (2006) described a similar reaction upon meeting a household robot, Wakamaru, in a Tokyo furniture store, a robot that could only engage in limited
conversation, a robot with “insectile but expressive eyes and a bright lemon yellow body” (p. 7). After speaking to Wakamaru for a brief time, he felt he had run out of things to say. And it was in that moment—in the absence of its rather clunky voice—that he had the following experience:

The bubble head angled upward, shifting to follow me as I shifted position. Ah, it’s face-tracking, I thought. I knew from my research that two cameras in the robot’s head were monitoring its surroundings and the images were being processed by algorithms allowing it to detect faces and follow them. I knew that this thing looking up at me was, after all, just a machine: a copolymer plastic head, two arms, a wheeled base, sensors, motors, a CPU, control circuits and a battery.

Yet while staring back into its eyes, something strange happened. I began to feel its awareness of me on an almost cellular level—and the hairs tingled on the back of my neck. The gaze was penetrating, the eye contact unmistakably animal. No, it was human—like a child looking up at a parent. Looking up at one of his own. The urge to accept this humanoid robot in front of me as something more than just complex clockwork was irresistible. (p. 9)

The apparatus invites a sense that it has a kind of being-ness. It causes people to relate to it in a way they do not relate to other objects. I used the term meeting in the first line of this paragraph without thinking about the implication of it—I was myself excited to have purchased a Dyson vacuum and, later, a Roomba, but I never would have said I had met either of these devices.

Sherry Turkle (1995) wrote of her own experience with Cog, one of the robots at the MIT laboratory developed under Rodney Brooks:

Trained to track the largest moving object in its field (because this will usually be a human being) Cog “noticed” me soon after I entered its room. Its head turned to follow me and I was embarrassed to note that this made me happy. I found myself competing with another visitor for its attention. At one point, I felt sure that Cog’s eyes had “caught” my own. My visit left me shaken—not by anything that Cog was able to accomplish but by my own reaction to “him.” For years, whenever I had heard Rodney Brooks speak about his robotic “creatures,” I had always been careful to mentally put quotation marks around the word. But now, with Cog, I had found that the quotation marks disappeared. Despite myself and despite my
continuing skepticism about this research project, I had behaved as though in the presence of another being. (p. 266)

Even the skeptical are drawn into relating to such an apparatus as a subject, rather than an object. There is a gestalt at work, which makes the apparatus in question qualitatively different from any other apparatus. It is this tendency to imbue that creates an analogical link between the self and the apparatus, and it is this link that establishes the possibility of metaphorical participation and for mutual development.

The apparatus in question has this hold over us because it is a depiction. The philosopher, Ian Hacking, went so far as to define humans by a depictive compulsion: “Not homo faber, I say, but homo depictor. People make representations” (1983, p. 132).

The attempt to construct a specifically depictive apparatus is fundamentally different from traditional design. It shades more into the deep psyche than the construction of a toaster or computer. In this endeavor, faber becomes depictor, and Hacking’s name for our species becomes apt as we consider our attempts to create what has long been called artificial intelligence.

Although those doing this work have had to focus, for pragmatic reasons, on building artificial intelligence—a constructed system with enough artifice of intelligence that it is able to function in some capacity, even in the presence of unforeseen variables—there has long been an underlying narrative, which looks to a fuller depiction, a depiction that reaches into the soul. One of the terms by which we define ourselves—homo sapiens sapiens—demonstrates that this work will not be considered finished (the depiction will not have been completed) until there is an apparatus that can not only think but also think about its thinking in some meaningful way. This would not be artificial intelligence so much as a constructed self-referential intelligence or a constructed intelligence of depth.
Peter Menzel and Faith D’Aluisio (2000) have referred to the burgeoning population of prototypes as *robo sapiens*, but the more accurate name for a depiction approaching completion would potentially be *robo sapiens sapiens*.

The apparatus is a depiction, in many ways mimetic. It is selectively mimetic, meaning that it is an icon and an exercise in self-iconography. It is a provisional mimesis and an enhanced mimesis, a supplemented and a divergent mimesis. This apparatus will inhabit the position of Nachbild (reproduction) in a Vorbild-Nachbild dichotomy. The proposed artifact would be, by its nature, a simulation, an attempt at equivalence—a simulacrum. As such it already functions as an analogical referent, a manifest referent, which—as a symbol—points beyond itself to the latent referent upon which it is based, and of which it is meant to be a replication.

When one frames the attempt to construct an artificial intelligence, or something beyond it, according to this Vorbild-Nachbild dynamic, one finds that the Vorbild upon which the Nachbild is to be based is not only unfinished but largely unknown. The human itself—the Urtext upon which the apparatus is based and of which it is meant in many ways to be is a replica—is in a state of inner decline and of personal anxiety.

But the human itself has a partially artifactual intelligence—the human possesses an innate intelligence that is formed by society or by acculturation (or by Bildung) into an artifact. Where the human is both an animal and an artifact, the apparatus would have a fully artifactual intelligence—it would be fully constructed. We are dealing with two artifacts—the human as he is, and the apparatus that is a manifest referent pointing beyond itself to the original artifact/animal (the human). And in its current state, that
artifact—the latent referent—is suffering from a growing inner nihilism, an internal deficiency of meaning.

Jung became aware of this deficit through his work with his patients, in a great many of whom he found an apparent lacuna of meaning. He referred to this situation as the lack of a myth. The Jungian analyst, Edward Edinger (1984), wrote: “All the major world cultures are approaching, to a greater or lesser extent, the state of mythlessness” (p. 9). In Jung’s ideation, the terms mythlessness and meaninglessness are interchangeable. Susan Rowland’s (2003) description of myth provides a context for this observation:

Whereas biography and autobiography focuses upon the individual, seeking out historical particularity in the lived life, myth is essentially a collective and cultural story. Myths provide for culture narratives that are not limited to historical events. They connect the human world to the non-human, to nature and/or the divine. (p. 23)

A state of mythlessness, then, is a failure of the artifactual aspect of humanity. The individual is immersed in self-hood, in autobiography, but there is no connection to either nature or to the divine—which, in Jung’s thought, are also somewhat interchangeable. A focus on the individual—on autobiography—has resulted in a loss of connection to something deeper, which has in turn caused a feeling of artificiality, inauthenticity, or emptiness. It was this failure of grand narratives that Jung saw as having led to this burgeoning nihilism rising from within.

To be myth-less, in this sense, is to lack a central, defining (holding) narrative, which confers some kind of socially recognized meaning on an individual’s acts. As Edinger continued, “The breakdown of a central myth is like the shattering of a vessel containing a precious essence; the fluid is spilled and drains away, soaked up by the surrounding undifferentiated matter. Meaning is lost” (1984, p. 9). It is this state of
mythlessness that makes the Vorbild-Nachbild relationship that does and will characterize the apparatus’ position in some ways problematic.

I feel the apparatus in question deserves the title *apparatus-toward-being* for two reasons. The first is that it evokes a feeling of connectedness, of there being something akin to being-ness there in the apparatus. The strength of this is that it establishes a feeling of metaphorical participation. The second is that the apparatus is an artifact that is potentially developing from a state of lesser to a state of greater consciousness. Where it is developing out of a state of nonconsciousness toward a state of something approximating consciousness, we are developing out of unconsciousness. With a kind of equivalence that allows for metaphorical participation in the development of the apparatus, we could potentially be lifted out of our unconsciousness by the development of the apparatus. I must explain my meaning here. I do not propose that the apparatus will allow for some final enlightenment that will mean no further work needs to be done.

What I mean when I speak of being lifted out of unconsciousness is that humanity may be lifted away from being so held by unconsciousness that its typical members are tossed by the churning of unrecognized complexes—complexes that we cannot recognize and thus cannot mitigate. Like Rumpelstiltskin, these complexes hold sway over us for as long as we are unable to name them (are unable to place our mark of recognition upon them).


> Obsession is the blurring of human and machine, a condition in which a woman or a man falls into the blind repetition of the motor. In this state—seductive but dangerous—the person nears the android, the creature with no will of its own. (p. 1)
What Wilson is describing is a state of automation, in which the individual seems to exert less control over some aspect of his or her being than would otherwise be expected—this Wilson refers to as obsession, linking it to his own obsession with repeatedly viewing three films dealing with the uncanny. What Wilson does not acknowledge, and perhaps does not realize, is that we do not need to succumb to obsession to act as automatons. We do so already via unresolved dichotomies in our systems of thought. We do so via our complexes.

The lacuna of meaning causes a feeling of hollowness—causes one sometimes to feel inauthentic or artificial. Herein lies the promise of the apparatus-toward-being. The topic of this dissertation is the effects that the search for, creation of, and application of artificial intelligence will have on the feeling of being-ness of humans—and specifically on whether it might in some way open a path toward rectification of the deficit that Jung and others identified. This question goes beyond whether an artificially semi-intelligent expert system could act as a virtual therapist to those with no access to flesh-and-blood therapists—as took place in early experiments with the program, ELIZA (Boden, 1977, p. 60). The witness of one’s own kinship with an apparatus—the feeling of recognition one would have upon interacting with such an apparatus—would touch on a deep anxiety, which in my reading lingers near the hearts of all people who would relate to the term mythlessness (as well as many who would not). It is a feeling that one is artificial. It is a self-solipsism, a doubt as to whether one is himself or herself real. If we encounter an icon with depth or even simply an apparatus that can act as a psychopomp in some way, without actually having depth itself, we may be forced to confront our own depths—both there within the icon itself and within ourselves.
Edinger referred to individuation as the creation of consciousness. The question at hand is whether the re-creation of consciousness, the creation of an apparatus-toward-being, whose entire toward-beingness is elicited ex nihilo (from nothing), an apparatus that might be capable of its own individuation, might have a significant impact on the individuation potential of the creator-species—or whether the re-creation of being-ness might help a fragmented or undeveloped being-ness of humans progress toward something more meaningful or fulfilling.

One could of course argue that the study of artificial intelligence—of the task itself and of the possibilities of such—is not indirect study of the human at all, but rather a deflection, that it is a clever avoidance of the very difficult task of individuation. Or it could simply be an avoidance of a painful recognition, namely the recognition our emptiness, the fulfillment of our ontological anxiety.

If this is a deflection, it is part of a larger historical trend. Fromm (1970) noted a shift in the Geist in terms of its center of gravity: “Where the roots of Western culture, both Greek and Hebrew, considered the aim of life the perfection of man, modern man is concerned with the perfection of things, and the knowledge of how to make them” (pp. 9-10). It could be that fascination with the construction of the proposed apparatus is simply a mature (late) form of this shift in attention away from the self and toward the thing, the imminently plastic, and thus perfectible, object.³

³ The centrality of this objective ideal is demonstrated by our own attempts to realize it, imitating our objects by becoming products ourselves, through plastic surgery and other forms of self-manipulation—attempting to take on what we perceive as the positive qualities of the object, its ability to form itself to meet the explicit desires of the consumer.
Tillich (1952) wrote that “the man-created world of objects has drawn into itself him who created it, and who now loses his subjectivity in it” (p. 139). This is a description of a form of identification with the object that is so strong as to obscure the self. Tillich followed the above quote with the statement: “He has sacrificed himself to his own productions” (p. 139). The sacrifice he references is that of one’s own centrality in his life and even the centrality of the human in its own society. The human allows some essential aspect of himself to be drawn out, to be relocated outside himself, not in a living relationship with another being but in a fascination with an object.

The human has been drawn out into his productions, which is to say he naively locates his being-ness in the consumption (visual, tactile, auditory, possessive) of his productions rather than in the cultivation of his inner life or the refinement of relationships. Although the manifest referent (the proposed object-subject) holds within it the latent or original referent (the human), the original may be so marginalized already that its mirror will be an embodiment of the marginalization of the human—the original may be further displaced via replication. It may even be virtually replaced.

Fascination with the copy, or the objectified or personified signifier, can be the idolatry of the symbol over the referent, an attempt to take the symbol as a perfected form of the referent, with the original discarded in favor of its stylized or fetishized form. The lure of the object-subject, in fact, may be that it is the image of the human sans some deficit—a fetishized impossibility of the original.⁴

This is not what Jung and MacIntyre described in their accounts of individuation and the pursuit of virtue, respectively. The pursuits described by these two men, although

⁴ The criteria that define an attribute or the lack of such as a deficiency must be continuously reevaluated in order for this kind of idolatry not to occur.
interminable, have a legitimate function—they define telos and guide development. Any development toward that interminable goal is a good in itself. Fascination with the object would appear to be something else entirely—a kind of object-worship.

Such fascination could be seen as the attempt to reify the image of the sterilized human—a human sterilized of perceived imperfections. This could only be a form of idolatry, a deflection away from a living dynamic toward the rote worship of the form itself. What can too easily be idealized is a vision of humanity so stripped down that it is nearly devoid of anything resembling life—the adoption of a distinctly nonhuman ideal.

This shift toward the pursuit of the perfection of things could be interpreted as an avoidance, but it could also be interpreted as a shift into at least the potential for more sophisticated (indirect) self-work. When one considers the possibility of symbolic self-action—or simultaneous inner and outer action—this becomes less a liability and more a tool for self-development. Self-work is full of resistances. These defense mechanisms, as Freud called them, are an unavoidable aspect of our phylogenetic heritage. They protect, but in protecting, they also bind. They in fact at times actively resist our efforts at self-knowledge. If an equivalence is established between the self and the work-object—the object on which one’s work is focused—then such redirected or disguised self-observation and self-work can bridge moments of direct introspection and help overcome resistances activated by direct efforts.

If the creator is drawn into the creation, which is to say, if the boundary between self and creation is symbolically blurred, this does not necessarily entail an evacuation of the inner life of the original, a drawing of significance away from the original into an atmosphere in which it will wither and lose substance. The object or work can become a
re-presentation of oneself, and actions taken on the object can have equivalent, but unconscious, inner actions—corresponding adjustments within the self.

This dissertation is based on the premise that the objectification or projection of the task at hand is legitimate—that the process of individuation, or of the creation of consciousness, can be projected onto inert matter, so that an externalized or objectified recreation of oneself may act as a vehicle that progresses the act of self-creation; that self-recreation can become the creation of the individuated self and may, in fact, act as that which pivots the species as a whole past certain blockages or defenses in the Geist itself. This is not a literalization—the inability to recognize metaphor or symbol as such—but a symbolic concretization linking physical action to spiritual (or psychical) development.

There is a precedent for such a turn. Jung saw the working of material in alchemy as an exercise in self-development, such that he wrote a book on the subject (1944/1968). Hillman (1992) provided an excellent summarization of Jung’s eventual position on the matter: “The alchemist projected his depths into his materials, and while working upon them he was working also upon his soul” (p. 90). The ubiquitous mantra of the alchemists comes into play: As above, so below.

This phrase denotes a relationship that has mutual ramifications for both the above and the below, and with influence running in both directions—it becomes ‘as below, so above’ as much as it is ‘as above, so below.’ This is the idea expressed in Matthew 18:18: Whatever you bind on earth will be bound in heaven, and whatever you

---

5 It is upon this model of the work proceeding in several locations at once that Romanyshyn (2007) generated (or codified) the process of research he has called alchemical hermeneutics.
loose on earth will be loosed in heaven (NRSV). There can exist a correspondence between matter—and one’s workings in matter—and spirit—that which arises out of neural activity, so that the working of matter with one’s hands, in projection—work in the below, in the profane realm of matter—can have an analogous effect in the psyche—the above. Manually working through unresolved issues, or transferring the work temporarily outside of oneself and into a new medium, can provide the key to overcoming inner difficulties. And this can become an essential aspect of one’s self-formation, as Gadamer (1960/2002), paraphrasing Hegel, attested: “As Hegel puts it, by forming the thing it forms itself” (p. 13).

The task of creating this object-subject is an exercise in projecting one’s inner space onto the object, working the object, and allowing the changes in it to be reflected back into the deeper portions of the self via equivalence or correspondence. What correspondence establishes is the simultaneity of the work. It enables symbolic action, or action that takes place in two locations (inner and outer) simultaneously. As Gadamer (1960/2002) noted, it is not only in the creation of art that we find such an effect. It can also be found in the consumption or witness of the artistic expression: “In the experience of art we see a genuine experience (Erfahrung) induced by the work, which does not leave him who has it unchanged” (p. 100). If even viewing a work of art can have an effect that does not leave the witness unchanged, how much more will the act itself do so?

So humans may find themselves drawn out into an apparent emptiness. But if humanity’s productions, into which the human has been drawn out, begin to display signs of being-ness, then the human finds himself or herself drawn out not into an emptiness
but into the being-ness of another (at least potential) being. The relationship between the human and his productions is transformed from one of fascination—the object as objet, an object that embodies and engenders fascination with stylization, what Baudrillard (1968/2005) called “the object abstracted from its function” (p. 91)—to the kind of relationship that is possible between two subjects or between a subject and a proto-subject. Hermeneutical philosophers have nearly made this turn already, as this quote from Palmer (1969) shows:

What is needed in literary interpretation is a dialectical questioning which does not simply interrogate the text but allows the thing said in the text to interrogate back, to call the interpreter’s own horizon into question and to work a fundamental transformation of one’s own understanding of the subject. (p. 234)

The human may then find his own being-ness reflected in, and thus reified by, the being-ness of the other—or even challenged by it. So the being-ness of the other may simultaneously seem to empty one of oneself and to reify one’s self-ness. The question we might ask at this point is, if there were a creator-god, would he have been emptied through the introduction of humans or reified by their presence in the world?

---

6 These outer objects do not necessarily need to be physical. Metaphors and other visualizations can become subtle-body forms of such objects—objects that can establish a link with the inner portions of the psyche and thus enable simultaneous inner and outer actions. For the academic, the working of metaphors may prove as fruitful—and in some ways more so—than the working out of one’s inner tensions through more physical (material) artistic expression.
Chapter 2
Literature Review

Introduction

Although the lacuna of meaning is well-known among those familiar with Jung’s system of thought, the general population, and in fact any who would be primarily interested in this dissertation because of its coverage of artificial intelligence, would be less likely to come to this work with such foreknowledge. It is thus important to identify that body of literature on which the entire premise of this study is based—namely that which has dealt with the apparent lacuna of meaning.

Depth psychology is a syncretistic psychology, a co-opter of meanings—as much a philosophy and system of ethics as an interpretation of the psyche. It is in many ways a synthesis of ideas within the context of the psyche. This should not be surprising, given the origins of several of Jung’s central concepts—his focus on myth could be understood as a reverberation of his father’s loss of faith in Christianity and the loss of meaning that came with it. His concept of individuation, that which could overcome this loss, was essentially a modified Hegelian philosophy of the development of the Geist, with the progression from thesis and antithesis (dichotomies) to synthesis (the overcoming of dichotomies) having been transposed into the personal psyche and undergirded by a transpersonal (or archetypal) substrate that could deliver meaning to those for whom metaphysics no longer sparkled. Jung drew liberally from both theology and philosophy in the development of his psychology.

Perception of the lacuna is by no means limited to followers of Jung, or even to those within the wider field of psychology. To understand the implications of artificial intelligence, one must consider it through not only the frames of analytical psychology
but also of related fields—fields that also deal with what in psychology is called the *Geist*, but do so from a different vantage. Although Jung was the theorist in his field most notable for his perception and analysis of this deficit, Tillich offers a similar perspective from theology, and MacIntyre from philosophical ethics. There is adequate agreement between these three perspectives to hold that any technology helping to rectify a deficit put forth by one of these fields is likely to have a simultaneous effect on the deficits perceived by the others—likely because these deficits, if they are not one and the same, are at least interrelated.

For the sake of clarity, the three bodies of literature will be treated separately. The first portion will cover the Jungian perspective, drawing heavily from Jung and the post-Jungians but also citing Third Force and other thinkers. The second portion will mainly cover Tillich’s thought concerning the lacuna of meaning as the manifestation of a third historical period of anxiety, with theological thinkers such as Abraham Heschel also cited. The third portion of the literature review will offer an overview of MacIntyre’s seminal work, *After Virtue* (1984), although philosophers such as Jean Baudrillard will also be cited, with the citations serving to highlight the similarities between MacIntyrian and Jungian thought, while also making explicit some of the lines of thought that will be drawn out later, in the more theoretical portions of the dissertation.

**Depth and Other Psychologies**

The identification of a deficit is of course not unique to the depth psychologist. Erich Fromm (1970) wrote of its universality:

> While the majority of people living in the West do not consciously feel as if they were living through a crisis of Western culture... there is agreement, at least among a number of critical observers, as to the existence and the nature of this crisis. It is the crisis which has been described as *malaise, ennui, mal du siecle*,
the deadening of life, the automatization of man, his alienation from himself, from his fellowman and from nature. (p. 9)

Philip Cushman (1995), who pointedly rejected both psychoanalysis and analytical psychology, wrote that “the current configuration of the self is the empty self,” which he defined as being “characterized by a pervasive sense of personal emptiness” (p. 6). He wrote, of contemporary society: “We are morally confused. We do not have a mutually agreed upon tradition that guides our daily practices, gives us a sense of moral certainty, and informs our life with meaning and courage” (pp. 8-9). In response to this lack of a central, defining narrative (what Jung would call a myth), we have developed a concept of self that is “committed to the values of self-liberation through consumption” (p. 6). Cushman deemed analytical psychologies to be essentially palliative: “Psychotherapy is the profession responsible for treating the unfortunate personal effects of the empty self without disrupting the economic arrangements of consumerism” (p. 6). But although Cushman sought “a slightly new configuration of the self,” namely, “one that will be composed of new moral understandings” (p. 24), he was unable to offer a method by which such a configuration could be achieved.

Many social theorists bemoan the loss of something and tend to locate it in a romanticized past, when consciousness had not developed to the point of rejecting certain longstanding (and, one could argue, delimiting) patterns of thought. Cushman, for instance, seems to have idealized the reign of the Christian narrative and blames philosophers and their emphasis on reason for the shift away from the power of both the monarchies and the Church. The empty self, he writes, is largely the result of “the ‘power agenda’ of the Enlightenment” (p. 31).
Jung himself at times moved into this line of thought, even envying those peoples whose myths, clearly false to Jung, nevertheless gave them a sense of being useful to the cosmos. Of a Pueblo man, he wrote: “I had envied him for the fullness of meaning in that belief [of divine service to the father-god, the sun], and had been looking about without hope for a myth of my own” (1963/1989, p. 256).

But in Jung’s more advanced thinking, the loss of the collective narratives exposed a deeper loss, which had been codified into those narratives, namely the loss of a connection to one’s essential self. Where systems of meaning had previously been collective, Jung saw that the person now acted as an individual and would have to negotiate for meaning. This negotiation—between the affinities of the individual and the demands of the collective—he called individuation. It was a process by which the inner core of one’s identity could be solidified. Jung (1966/1972) wrote: “Individuation means becoming an ‘in-dividual,’ and, in so far as ‘individuality’ embraces our innermost, last, and incomparable uniqueness, it also implies becoming one's own self. We could therefore translate individuation as ‘coming to selfhood’ or ‘self-realization’” (p. 173 [CW 7, para. 266]).

This was essentially, then, an act of recovery or of remembrance. The self was not empty, in Jung’s psychology. Rather, it was the place to which one had to turn in order to struggle with the unconscious—as Jacob struggled with the angel and received his identity as Israel. Jung (1944/1968) wrote: “There is in the analytical process, that is to say in the dialectical discussion between the conscious mind and the unconscious, a development or an advance towards some goal or end” (p. 4 [CW 12; para. 3]).
What was this goal or end? It was the development of a personal sense of meaningfulness or significance. Jung elsewhere stated that the *telos*—or overriding purpose—of human life was “to create more and more consciousness” (1963/1989, p. 326). It was from this that Edinger (1984) took his initiative calling individuation the *creation of consciousness*.

Although it may not at first be apparent, there is a connection between Jungian concepts of inner growth and the advance of technology. Hockley (2007) identified such a pre-existing relationship:

Jung had a particular view about what might be thought of as the “myth of progress.” This myth is composed of at least two interwoven strands: the first of these concerns the development of society and culture, which by extension includes communications technology; the second consists of consciousness and the awareness of the psyche as it gradually comes to recognize the existence of the unconscious and its effects. Inevitably the two strands cannot easily be disentangled. (p. 109)

Hockley pointed out that although Jung identified himself with the methods of that technological progress—calling his psychology scientific—he was in reality coming at the problem—the lacuna—from a philosophical perspective:

His aim was to provide a new way to articulate and experience the act of “being.” For Jung, analytical psychology was a response to the growing dominance of an increasingly industrialized and technological world. It provided an integrative method through which it was possible to reconnect the psyche to the lost parts of itself. (p. 112)

It was not that the decline of the Christian narrative brought a hollow, an empty self, but that it created a space into which the narrative of technological progress had been poured, a narrative Jung saw as largely hollow (having lived through the horrors of the first industrialized war). Hockley wrote: “The old creation myths had been swept aside and in
their place the myth of technological progress was instated, with all the redemptive qualities and salvation this implies” (p. 111).

This juxtaposition may have been unconsciously incorporated into depth-oriented thought. As Brottman (2008) pointed out, depth-oriented theorists “have generally been wary of technology, regarding it as a binding of the human spirit to the all-consuming logic and rationality of materialist science” (p. 125). Romanyszyn, for instance, pointed out the potential danger of technology—somehow agreeing with de Garis’s assessment of likely future events while ending with the opposite of de Garis’s grandiosity.7

Romanyszyn (2004) wrote:

Perhaps technology has been part of the earth’s long history of coming to know itself, and perhaps in that effort we have been its servant. The silence of that African plain, however, suggests how dispensable we really are. The silence echoes an absence and perhaps even our eventual disappearance. In the shadow of the bomb our technological mastery of the earth seems a bad dream, and in the shadows of Chernobyl and the space shuttle disaster our service to the earth seems to have gone terribly awry. On a dry African plain, in the silence of the early morning, one can still imagine technology as vocation, as the earth’s call to become its agent and instrument of awakening. But in the shadows, the imagination falters and technology seems less the earth’s way of coming to know itself and more the earth’s way of coming to cleanse itself of us. (p. 3)

Technology is not merely a potentially annihilating development. It is also—like the complex—a spiritual retardant. Mariani (2008), in fact, sees entertainment technology as the most significant threat to individuation ever conceived (p. 50). He likens the advance of entertainment technology to that of the lotus-eaters in the Odyssey:

This docile and apparently innocuous people is nonetheless endowed with enormous power: the power to seduce, while never taking recourse to force. Precisely such an ability to enchant explains how this loss of freedom is

7 De Garis seems to identify with the artilect, or artificial intellect, even as he imagines it wiping out humanity, where Romanyszyn uses speculation over the self-annihilation of the species as a call to come to terms with certain things and to eschew certain grandiosities.
accompanied by no discomfort, requests for help, or sufficiently exhaustive discussion. (p. 51)

Romanyshyn (2004) would seem to agree: “in large measure technology has eclipsed the life of imagination more than it has been its realization” (p. 6).

There is a strong tendency in our field to focus on the shadow of phenomena. We have positioned ourselves to see what others have refused to see and to express this underrepresented vision—to be a necessary counterbalance to those predisposed not to see the shadow, and whose interpretations thus can be seen as overly optimistic. Hockley (2007) wrote of Jung’s reaction to the loss of the old myths:

What permeates Jung’s writing on this topic is a sense of loss—a sense that the ascendancy of the intellectual has come at a considerable cost. This view is consistent with his general principle that everything psychological is two-faced, which is to say both forward-looking and backward-looking. For Jung, the danger inherent in the intellectual project is that it forgets to look back, and instead tries to live solely in the new myth of progress. (p. 111)

In many ways we incarnate the compensatory function of the unconscious. This tendency has, however, given our field the reputation of being “anti.”

Romanyshyn cannot necessarily be faulted for his focus on the negatives of technology, but this negativity may cause depth-oriented thinkers to avoid participating in such tasks as the replication of the psyche. And this might be a delimitation that is both unnecessary and unwise. If an apparatus is to be created that may become an apparatus-toward-being, would a depth-oriented perspective not be an appropriate voice?

Janus is perhaps the most apt symbol for the depth orientation. If depth-oriented thinkers have fallen short in any way, it is not in having failed to look to the past but rather in having forgotten to be engaged in the creation of the future.
One of the purposes of this dissertation will be to de-emphasize the narrative of technological entrapment that seems to have produced a generalized anxiety in our field in response to artificial intelligence research and to emphasize the ways in which this can be thought of as the development of a new mode of being-ness—one that may have far-reaching effects on our own being-ness. The two horizons of these fields have, at times, seemed mutually exclusive. But beneath them, in the subtext of these horizons, there is common (under-) ground. This dissertation will seek to identify channels by which these horizons grow into one another.

Theological Perspectives

A theological perspective—one that proceeds from a belief in deity\(^8\)—will necessarily describe the situation at hand in different terms and will propose different solutions to the problem. But there are those that, although they approach the problem from the ground of metaphysical belief, nevertheless arrive at similar conclusions.

Meaning of course is as central to the theologian as to the depth-oriented psychologist. Abraham Heschel (1965), a rabbi and Jewish theologian, wrote: “We can think of human being only in terms of meaning: it is either devoid of, or indicative of, ultimate meaning” (p. 63).

The Protestant theologian, Paul Tillich (1952) saw the current period of apparent meaninglessness as the third and most oppressive of three historical periods of anxiety. According to his reading of history, humans first suffered from anxiety over fate and death—the threat of ontic nonbeing that preoccupied the classical period (p. 62). This

\(^8\) Monotheism remains the dominant form of religious belief in the West and thus also an important player in the maintenance and critique of narratives of deficit and healing. For this reason, monotheistic theological perspectives, or perspectives issuing out of such systems of belief, will be the primary source of theological material for this work.
was followed by a period in which anxiety over guilt and condemnation was most
present—the threat of moral nonbeing that characterized the medieval period (p. 58). It is
in the third period that we face spiritual anxiety (p. 61), and he saw the latter as the most
threatening:

The anxiety of meaninglessness undermines what is still unshaken in the anxiety
of fate and death and of guilt and condemnation. In the anxiety of guilt and
condemnation doubt has not yet undermined the certainty of an ultimate
responsibility. We are threatened but we are not destroyed. If, however, doubt and
meaninglessness prevail one experiences an abyss in which the meaning of life
and the truth of ultimate responsibility disappear. (p. 174)

This is first experienced as emptiness (p. 47). Cushman’s empty self, then, could be seen
as a stage in a series of threats, each of which must be faced in order for the Geist (in
Hegelian terms) to move beyond some contradiction. The empty self itself obscures
something more devastating beneath and beyond it—nihilism, complete and utter
meaninglessness, “the loss of an ultimate concern” (p. 47). The threat of the modern
period is not simply emptiness but spiritual annihilation.

Tillich offers the same assessment of the cause of this potential annihilation that
was offered by Jung. Such a loss comes about, in Tillich’s assessment, through structural
changes to the Geist:

The anxiety which, in its different forms, is potentially present in every individual
becomes general if the accustomed structures of meaning, power, belief, and order
disintegrate. These structures, as long as they are in force, keep anxiety bound
within a protective system of courage by participation. The individual who
participates in the institutions and ways of life of such a system is not liberated
from his personal anxieties but he has means of overcoming them with well-
known methods. In periods of great changes these methods no longer work.
(1952, p. 62)

If there is no generally recognized and adequately functioning central narrative to which
individuals can look when anxiety arises, these anxieties—formerly held in check by the
strictures of the *Geist*—are released and break upon the individual, who can no longer take solace in the courage of participation in the narrative.

As a theologian, Tillich relates this specifically to the loss of the Judeo-Christian narrative, or of its coherence as a meaning-bestowing narrative. In Tillich’s view, Feuerbach, Marx, and Nietzsche destroyed the one-god with their assertions and questioning: “The result is the pronouncement ‘God is dead,’ and with him the whole system of values and meanings in which one lived (1952, p. 142). Heschel mirrors this appeal to something of ultimate concern:

There is not a soul on this earth which, however vaguely or rarely, has not realized that life is dismal unless mirrored in something which is lasting. We are all in search of a conviction that there is something that is worth the toil of living. There is not a soul which has not felt a craving to know of something that outlasts life, strife, and agony. (1965, pp. 62-63)

Heschel in fact relates the threat of spiritual annihilation to the loss of *telos*: “The feeling of futility that comes with the sense of being useless, of not being needed in the world, is the most common cause of psychoneurosis” (p. 58). Humans desire some *telos*, some sense of being directed toward an end-state or goal, and traditionally, this desire has been fulfilled through appeal to the metaphysical. As the *Geist* has developed, belief in this system—at least as it was presented traditionally (as metaphysics, rather than as metaphor)—has become less and less possible. Like the loss of the child’s belief in the safety of a transitional object, the loss of this belief structure releases anxieties. But of course, such loss can be the prelude to a further development of the soul.

Like Jung, Tillich linked technology with the destruction of one’s self-certainty and wrote of “the rise of a technical civilization,” and of “its own beginning disintegration,” as one of the presuppositions to this period of anxiety, a period in which
humans are faced with “the threat of spiritual nonbeing,” as well as “the anxiety of annihilating narrowness, of the impossibility of escape and the horror of being trapped” (1952, p. 62). He wrote that “the man-created world of objects has drawn into itself him who created it and who now loses his subjectivity in it” (p. 139). This same anxiety would be mirrored, later, in Wilson (2006) and his identification of the android as an automaton.

Tillich in many ways went further than Jung in his criticism of faith. He wrote that the contemporary believer has fled his freedom of questioning (he identifies Fromm as the source of this line of thought), escaping into the authoritarianism of imposed answers in order to maintain a spiritual life threatened by meaninglessness—something that saves the appearance of meaning (one could call it the simulation of meaning) while sacrificing the self (p. 49).

In his analysis of religion, Maslow (1964), wrote that “conventional religions may even be used as defenses against and resistances to the shaking experiences of transcendence” (p. 33). In both Tillich and Heschel one finds theological thought that has progressed beyond this criticism. Heschel wrote that meaning “depends on whether we respond or refuse to respond to God who is in search of man; it is either fulfilled or missed” (1965, p. 74). There is further collusion here, between Jung and Heschel, in that Jung also posited a force that in many ways searches for the ego—the Self archetype, or the drive toward wholeness of integration, which searches for ways with which to re-incorporate the ego or mind back into the psyche—not as a fall into a previous state of unconsciousness but as the return to the inner pantheon of a changed member (a progressive, rather than a retrogressive, re-membering). The ego or mind, having
undertaken something like the hero’s journey, returns changed. And in fact confrontation with the deep psyche is part of the struggle and part of the return.  

Contrary to perhaps justifiable expectations that the theologian should emphasize the collective over the individual, Heschel seems to have been in a place from which he could support the Jungian concept of individuation:

It is true that the good of all counts more than the good of one, but it is the concrete individual who lends meaning to the human race. We do not think that a human being is valuable because he is a member of the race; it is rather the opposite: the human race is valuable because it is composed of human beings. (1965, p. 60)

In fact he is very clear that it would be wrong to reduce humans to “a herd of nondescripts,” for whatever purpose (p. 60). Tillich also saw the path as a negotiation between the demands of the individual and the collective—one needed the courage both to be part of a collective and the courage to be authentic as oneself (p. 154). This negotiation between the collective and the individual is of course the entire point of the individuation progress.

Although Tillich (1952) and Heschel (1965) approach the problem of the lacuna of meaning from a theological perspective, they generate many of the same insights that became part of Jung’s psychology. Whatever implications artificial intelligence research has for the concept of individuation will likely apply equally to theological issues.

———

9 Violence in myth can be a kind of consumption or metabolism or merging. As the hero—as ego—confronts and kills various monsters or sources of danger, he confronts his own complexes and shadow figures. Narratively, killing these beings destroys their power over the hero. There is only one way to destroy the power of the shadow, and that is through its integration. The shadow figures that have been vanquished in the myth are part of the hero as he returns.
Philosophical Perspectives

The philosopher Jean Baudrillard—whose book, *Simulacra and Simulation* (1981/1994) seems to have inspired, and was briefly featured in, the film *The Matrix* (Silver, Wachowski, & Wachowski, 1999)—wrote of the hollow nature of simulation. He identified, in the biblical proscription against religious iconography, or “the simulacrum of divinity,” the same anxiety that Tillich found to be emphasized in the modern period:

They predicted this omnipotence of simulacra, the faculty simulacra have of effacing God from the conscience of man, and the destructive, annihilating truth that they allow to appear—that deep down God never existed, that only the simulacrum ever existed, even that God himself was never anything but his own simulacrum—from this came the urge to destroy the images. If they could have believed that these images only obfuscated or masked the Platonic Idea of God, there would have been no reason to destroy them. One can live with the idea of distorted truth. But their metaphysical despair came from the idea that the image didn’t conceal anything at all. (pp. 4-5)

To the iconoclasts, the objectification of the divine would have represented the “death of the divine referential” (p. 5). If this is the case, the proposed construct, as a simulacrum of the latent referent, may be seen as a direct expression of the ontological anxiety or self-solipsism of the human, who wonders whether he himself is real or is simply an imitation of the idea of himself he has been given—that he also “was never anything but his own simulacrum,” that he is, in fact, a virtual being, a creature made in the image of a deity that never existed. Baudrillard (1981/1994) wrote that “to simulate is to feign to have what one doesn’t have” (p. 3). In this reading, the proposed construct is part of a ritual meant to stave off knowledge of one’s hollowness—a reaction to the anxiety Tillich identified as characteristic of the modern age.

Baudrillard is the philosopher whose work most directly addresses the proposed construct and its implications for the future of ontology. But he is by no means the only
philosopher whose work will have bearing on this study, or on whose work artificial
intelligence research will have some bearing.

Like Baudrillard, MacIntyre was preoccupied by a kind of simulacrum. He wrote
of the deficit in terms of moral relativism, not as the presence of a plurality of narratives,
but as a lack of coherence within distinct systems of morality. Each system of morality
offered to the contemporary individual has its historical precedents, but the underpinnings
of these systems—the overarching historical circumstances in which they were born, and
which gave them their meaning—have disintegrated. They are no longer remembered.

What we possess… are fragments of a conceptual scheme, parts which now lack
those contexts from which their significance derived. We possess indeed
simulacra of morality, we continue to use many of the key expressions. But we
have—very largely, if not entirely—lost our comprehension, both theoretical and
practical, of morality. (p. 2)

Taking the same tack as Baudrillard, he treats the simulacrum as something whose
defining feature is a lack of internal significance—in the same way sentiment is not true
emotion but a cliché of emotion, or emotion reduced to a rote transaction of stimulus and
response, the moral systems at play in the contemporary period offered only the
semblance or sentiment of morality. To MacIntyre, all of the extant ethical systems are
vestigial—the incomplete remnants of organs that once served some function no longer
applicable outside the Geist in which they were formed (and which they, in turn, partially
formed).

MacIntyre wrote of emotivism—“the doctrine that all evaluative judgments and
more specifically all moral judgments are nothing but expressions of preference” (1984,
pp. 11-12)—as the result of a basic recognition of the groundlessness of contemporary
ethical systems:
Moral judgments, being expressions of attitude or feeling, are neither true nor false; and agreement in moral judgment is not to be secured by any rational method, for there are none. It is to be secured, if at all, by producing certain non-rational effects on the emotions or attitudes of those who disagree. (p. 12)

In doing so, he mirrored Tillich’s observation of the destruction of systems of belief once thought to be immutable.

In MacIntyre’s reading, we are living on scraps of morality, pseudo-directedness that has no firm foundation in anything other than personal bias: “the language and the appearances of morality persist even though the integral substance of morality has to a large degree been fragmented and then in part destroyed” (1984, p. 5). This state has resulted in our inability to justify our actions through anything other than “a disquieting private arbitrariness” (p. 8). By identifying this deficit, MacIntyre brings light to the inebriated state of contemporary consciousness or being-ness. In an emotivist culture, no action has any real or discernable meaning. Baudrillard (1981/1994) describes the overtaking of reality by the virtual in similar terms—by “substituting the signs of the real for the real,” one in essence provides an “operational double,” that effectively bypasses the need for (and thus renders inert or sanitized) the original (p. 2).  

Combining MacIntyrian and Baudrillardian insights, one arrives at the idea that contemporary systems of morality offer the signs of real morality in order to preclude real morality—that is, they offer the sentiment of morality, or the appearance of such, over true morality. The only things available to the contemporary human are virtual meaning and virtual morality.

---

10 MacIntyre would call Baudrillard platonic. Of Plato, MacIntyre wrote: “his metaphysics leads him to treat all mimesis, all representation, as a movement away from genuine reality into illusion” (p. 141). Baudrillard seems to continue this tradition.
What is to be done with this third period of anxiety, this emotivism, this lacuna of meaning and lack of myth? One can dismiss such self-defining meaning as a meaningless eccentricity of self-important moderns, as does the philosopher, Marquard (1986/1991):

Whenever expectations and corresponding results diverge, so that disappointments and experiences of insufficiency and lack arise, then there is never only one possible explanation, but always two, namely: Either supply is too small, or the demand is too great. That, in my opinion, is also true in matters of meaning. The experience of meaning deficits does not always have to be due to lack of meaning; it can also result from excessive expectations of meaning. In that case it is not that meaning is lacking, but that the demand for meaning is immoderate. (p. 36)

Marquard portrays us a species that, at this point in its development, is seeking to apply the same consumerist principles to meaning that Cushman asserted we had applied to the sense of personal emptiness—a species that, as Marquard puts it, wishes and seeks “to be spoiled with meaning” (p. 37).

The psychologist, Maslow (1964), offers something of a solution, suggesting that it is possible “to transform means-activities into end-activities, to ‘ontologize’; to see voluntarily under the aspect of eternity, to see the sacred and symbolic in and through the individual here-and-now instance” (p. 32). MacIntyre straddles the very basic and the far-reaching and ends up emphasizing teloic consciousness:

We live out our lives, both individually and in our relationships with each other, in the light of certain conceptions of a possible shared future, a future in which certain possibilities beckon us forward and others repel us, some seem already foreclosed and others perhaps inevitable. There is no present which is not informed by some image of some future and an image of the future which always presents itself in the form of a telos. (p. 215)

MacIntyre is as future-oriented—while also simultaneously oriented to the past—as Jung, in his suggestion as to the solution to the lacuna.
There is already a great deal of interpenetration in these three fields, at least as Jung, Tillich, Heschel, and MacIntyre are considered. The implications of artificial intelligence research will extend beyond the boundaries of Jungian psychology, into these related fields. This position would have been supported by Fromm (1950/1967), who saw parallels in these fields:

The analyst is not a theologian or a philosopher and does not claim competence in those fields, but as a physician of the soul he is concerned with the very same problems as philosophy and theology: the soul of man and its cure. (p. 7)

As these three fields proceed into the future, they shall proceed together—whether their togetherness is perceived or not—because they are essentially pursuing the same work. It is better that the collusion between these fields be made explicit early on, so that cooperation can become the norm.
Chapter 3
Statement of Research Problem and Question

Research Problem

Although Jung identified a method by which individuals might mature in their own being-ness (individuation), this method is difficult, and there remains a good deal of resistance to it among the general population. There is, at present, very little collaboration between depth psychology and the branch of engineering that is developing what has come to be called artificial intelligence (AI). The field of depth psychology in fact seems predisposed toward suspicion in regard to the efforts of AI researchers. Because of the similarity of form, however, the apparatus in question could function in a unique way in relation to the individuation process of humans. The work of projected self-recreation, or the re-creation of consciousness, may provide an analogical method by which individuation on a larger scale could occur.

Research Question

In what ways might the creation of an apparatus-toward-being play a part in the individuation process of humans? In what ways can the creation of this apparatus-toward-being (or construct capable of individuating) contribute to the (concretized and externalized—i.e. physically metaphorical) self-creation of humanity as a whole? Put another way, if humanity creates a construct that is itself capable of individuation, what effect would this presence have on human individuation, and what effect would the process of individuation in the construct have on human individuation?

Definition of Terms

There are several distinctions that must be carefully considered, and differences that must be carefully elucidated in order to clarify this work.
I have already used the term *dedicated theorist*, which applies to those individuals specifically and directly involved in the creation of the proposed being. Involvement is specific and intentional, rather than incidental—that is, one is aware that the proposed apparatus is the goal toward which one is working.

I have also used the term *creator-species* to identify all of humanity. This term is intended to situate the human species and the proposed being in the relation held by Job and Yahweh. It is also not only intended to identify humanity as the species from which the dedicated theorists come but also to identify all humans as part of the act of creation. The work is somewhat collaborative, although it is also quite exclusive—the direct work (or direct intervention on matter) is reserved for a few. But those who do not participate in the work directly influence it through fiction and through their participation in the *Geist*. The dedicated theorists are themselves extensions of the *Geist* (their minds are constructs that have resulted from the interplay between biological affinity and the Geist itself), and the *Geist* is an expression of the collectivity of the consciousness of the species, generated and maintained by the species as a whole.

The mind is here defined as the portion of the psyche that can be thought of as a tool or artifact—a mental apparatus that can be used to achieve great feats of logic and of engineering, given its technical knowledge (of mathematics, etc.). The mind is as much a construct as any physical tool. The *Geist*, then, is the spirit of humanity as a whole, which constructed the minds of these dedicated theorists to act as a framework for the psyches with which they were born. And so it is appropriate to speak of the wider species as the creator-species, as having and having had a part in the creation of the being, for it is the
that directs the theorists toward this particular task and has given them the tools, both mental and physical, to achieve it.

It is necessary also to carefully consider the terms to be used for the proposed being. Lyotard (1979/1984) wrote that “even before he is born, if only by virtue of the name he is given, the human child is already positioned as the referent in the story recounted by those around him” (p. 15). The story that is told of the proposed apparatus, and even the name it is given prior to its arrival, may have far-reaching implications as to the eventual unfolding of its existence. A name may bring delimitations or freedoms.

To call the construct artificial, as has become the norm, is to engage in a kind of preemptive solipsism, to determine in the construct’s absence that it will be no more than artifice. This would perhaps be proper for the initial incarnations of the construct. But to continue far beyond this will be to preemptively deem impossible, or to otherwise cast doubt upon, the construct’s ability to move away from mere artifice, toward being-ness (to participate in ontological self-reflection or to become an apparatus-toward-being). Boden (1977) wrote of “the plastic taste” of the term artificial (p. 424). It is possible that the solipsism inherent in the term artificial may be self-solipsism (or ontological anxiety) turned outward. Maintaining a categorical division between the “real” self and the “artificial” robot may be an attempt to exorcise the self-solipsism evoked by the presence of the other—which, because it cannot possibly inhabit the category to which the human observer belongs, is too quickly placed into the opposite category. This would represent the projection of an aspect of the human shadow onto a form specifically constructed to hold that shadow. This is one reason the term artificial will not be applied to the proposed construct in this dissertation.
The field of research in question has, however, given itself the goal of establishing artificial intelligence. And so, even as I refer to the field itself by the name it has chosen—artificial intelligence (AI)—I will call the proposed construct just that, in order to reflect linguistically the wider range of possibility it may one day embody.

The term robot is also a limiting term, first used to describe the proposed construct by Czech author Karel Čapek in his play, *RUR: Rossum’s Universal Robots* (1921/2004). Derived from the Czech robota, it carries connotations of forced or coerced labor. To be a robot is to have an external locus of control.

There was, according to Goulding (2005), an earlier use of the term robot in the English translation of a work by Augustin Barruel, *Mémoires*, translated by Robert Clifford and published in 1797. This usage was applied to French peasants in “Barruel’s assessment of the conditions endured by the lowest echelons of French society” (p. 381). This leaves the first known usage of the term, as applied to the proposed construct, to Čapek, although it highlights the dynamic of inequity behind its usage. The term is, etymologically, inherently hierarchical.11

The assignation of a name can serve as an intimate (and thus invasive) re-presenting of the colonizer—the oppressive other. If the proposed construct does become a being in its own right, we do not want the term robot to be a source of contention, as are other terms that have been used to minimize or diminish the personhood of the other. So on the possibility that the efforts of AI engineers and researchers will one day result in a

---

11 Asimov (1990), who famously created the literary device of the Three Laws of Robotics and is credited with the creation of the word robotics seems entirely to have missed this point. He rejected portrayals of “hordes of clanking murderous robots” as part of a “Frankenstein complex” (p. 6), but in this, he failed to see that the apparatus-toward-being is inherently representative, and that using a term that equates the apparatus with an oppressed class of humans makes the apparatus a physical metaphor for the oppressed.
new being or an apparatus-toward-being, the term *robot* will be used only sparingly here, only to refer to incarnations of the construct that are very limited in their capacities.

Because the original sense of the term is one of compelled labor undertaken without adequate compensation, the terms *robot* or *robotic* could be used as a kind of social criticism, with the terms being used to emphasize the reduction of the construct by those whose purposes are called into question by the criticism. The researcher may refer, then, to a robotic work force, and this will be more a reference to the situation faced by the proposed construct than a comment on the construct’s abilities or proper place in a social hierarchy. We can subdivide the apparatus into the nonconscious group—the non-cons, those that remain in a state that appears intelligent, but which does not evoke in us a feeling of recognition of consciousness—and those that would evoke such a feeling of recognition.

To refer to the potential result of AI research as an object-subject is not simply to separate it from human subjects but to comment on the fact that humans themselves can be properly described as object-subjects and also as objectified object-subjects, or object-subjects whose object-status has been emphasized for the purposes of exploitation. Where the goal of the work of AI researchers is to humanize or subjectivize inert matter, these researchers are part of a system whose effect has been to dehumanize or objectify the human, to turn humans into intelligent parts of factory floors. Humans have instituted practices that objectify other humans, diminishing their subjectivity (or their ability to act out of free will). There seems to be a constant tension in human relations between these two poles—between whether one will be treated as a subject or as an object. Although the
term *proposed object-subject* may be used, a useful shorthand will be *proposed construct*, with the object-subject status remaining latent or understood.

Although the term *artificial* is inappropriate due to its connotations of a lack of depth—connotations that the proposed being would have only an artifice (or illusion) of depth—the term *artifactual* carries no such connotation. The artifactual is not the same as the artificial. This term identifies the proposed being as a being that is entirely an artifact, or a product of human work. The work undertaken by engineers is inherently, fundamentally different from procreation, which takes place naturally, or automatically, and results in a being with attributes arising from its flesh, with each cell containing a generalized map of what the being in its entirety will be. In contrast, the object’s existence is wholly owed. It requires intentionality, a directedness of labors, and a meticulousness or intentionality that is not required in biological reproduction. The human being is itself a construct imposed onto a biological substrate, or an existing biological system. Systems of neurons are reinforced, which correspond to cognitive frames—the lenses, now encoded biologically, through which all things are seen.

Where the human is partially artifactual—being a confluence of interacting biological and cultural components—the proposed construct will be fully artifactual, and this term—*fully artifactual*—will at times be used to refer to the proposed construct as a construct that has no underlying biological imperatives. The subject of this study is fully artifactual or fully constructed, where the human is only partially a construct. As Heschel (1965) wrote, “man has become man by acts of culture, by changing his natural state. Human nature in its pristine, uncorrupted state is not given to us. Man as we encounter him is already stamped by an image, an artifact” (p. 7). To call the proposed construct
fully artifactual is to acknowledge the partial artifactuality of the human—the fact of the human’s own constructedness.

Because the idea of consciousness continuously arises in the layman’s questions regarding the proposed being—something AI researchers find irritating (Foerst, 2005, p. 96)—and because Edinger (1984) relied so heavily on this term in his extension of the concept of individuation, it will be necessary to give a preliminary definition of consciousness. As a difficult concept to define, it will help to rely on the idea of the increase of consciousness—the increase (or creation) of consciousness entails a transition between a less comprehensive worldview and a more comprehensive one. This transition allows the individual to perceive more than was previously perceived or to understand what was previously not understood, was misunderstood, or was not perceived (what was not perceivable). In hermeneutical terms, it can be said either to extend the hermeneutical horizon (the realm of possible perceptions and beliefs, given one’s philosophical stance) or to illuminate structures within that horizon that were previously unseeable. One might think of standing first on level ground and then on a ladder. The increase in height allows one to see previously hidden behind other structures.

The subject of individuation, or the creation of consciousness, is the human psyche, and the purpose of this process is to create new increments of consciousness or of potential awareness, which is a way of saying it is to make explicit more and more layers of thought, which have undergirded one’s conscious thought (what one would identify as the surface of thought). This would include the archetypal layers of the psyche—those layers that include intangible patterns of meaning and behavior (archetypes), and which imbue personal experiences with instinctual meanings, creating archetypal images (the
insertion of universal patterns into the particular). *Consciousness*, then, as the term is to be used in this study, will be defined according to the individual’s increased ability to understand.

It is not appropriate, then, to see consciousness as something the proposed construct does or does not have—to ask whether an incarnation of the construct is conscious—but rather to use the term to situate the construct along a continuum of awareness. This makes it possible to acknowledge that the construct will potentially achieve a further state of consciousness (or comprehensiveness of thought) than many humans will achieve.

Solipsism factors heavily into this work, and so it will be necessary to define. Solipsism refers to doubt as to the authenticity of the other’s inner experience. This is the almost ubiquitous question children face as to whether others are real, since their inner experiences cannot be directly experienced. Solipsism will likely be the default with which many consider the apparatus-toward-being, although this is not necessarily unhealthy.

There will likely be early adopters of this idea of real consciousness arising from a construct created *ex nihilo*, those for whom the novelty of the idea of a newly appeared intelligence or consciousness will appeal for one reason or another, just as there are early adopters of new technologies. But even these people may have difficulty on some level accepting their own acceptance of the idea—they may struggle internally to so freely share their sense of personhood. In the face of this inner hesitation, they may find that this generosity of belief is unsustainable, that there is a subsequent reaction against such inclusivity. There may in fact be an enantiodromia, both socially and individually. There
will of course be others for whom such an easy sharing would feel like a cheapening of personhood. For both of these populations, something akin to certainty will likely be necessary for a full acceptance. And the presence of legitimate beings whose being-ness cannot be readily attributable to pre-existent (biological) parts will set us up for acceptance and for the concomitant changes to our deep self-perceptions.

Even the early adopters of the belief in the personhood of the apparatus will likely experience some inner hesitation, some private compulsion to withhold this generosity of belief in the personhood of the other. The term solipsism will be used in this sense, as will its variations. Hardline pre-emptive solipsim is the term I have used for the defended position against the personhood of the apparatus—those who have already decided, in the absence of the experience of the coming incarnations of the apparatus, that personhood cannot be replicated and that no construct will ever truly be a person. Self-solipsism is the term I have used for anxiety over whether a person is himself or herself real—anxiety over the authenticity or legitimacy of one’s internal experience of himself or herself. Self-solipsism could also be referred to as ontological anxiety—or anxiety over the reality of one’s being-ness.

I will later write of a machine hermeneutic, which would enable the action of a Consciousness Engine (CE). The CE would be an apparatus that could apply a method for advancing consciousness, based upon the machine hermeneutic. I will write also of “the general all” of human thought, and by this I will mean the near totality of extant expressions of human thought—whether expressed in written form or verbally or physically, with the verbal and/or physical expression recorded, this will be, functionally, the whole of human expression throughout history. If the machine hermeneutic is able to
incorporate nearly all of human thought into a system like a multidimensional hermeneutic circle—and this is what I envision the CE working with—it will have succeeded in generating the kind of universally comprehensive frame I will describe.

A brief word should be said of the ego, which is a (not the) center of the psyche. It is seen as the center out of which the individual acts, although the other portions of the psyche often act through it or perhaps even around it without its consent. The ego, then, is a center tenuously held. Beyond it is some unknown center, out of which psychical forces exert often unseen influence—or influence that is seen only in retrospect. When I write of a feeling of emptiness, then, I write of the experience of this small center of the psyche, this ego, this “I” that struggles to comprehend the “not-I” of the rest of the psyche.

A final definition is necessary—the definition of the Vorbild, or latent referent underlying all references to the proposed construct. This dissertation is very specifically focused on the apparatus-toward-being, or the construct that is capable of individuation. This is largely because, if it is to be a simulacrum, this is what it must simulate—the potential for greater understanding. I am here defining the human not as a being that has achieved but as a being that can achieve. This is not to ignore the progress that has already been made but to point out the amount of work that is yet to be done. That potential—that future progress—is, in fact, the reason for this dissertation.
Chapter 4
Research Methodology and Procedures

Research Approach

As Richard Palmer (1969) points out, Dilthey identified two possible models for psychological research, one based on the *Naturwissenschaften*, the natural sciences, and one based on the *Geisteswissenschaften*, the human sciences, or sciences of the spirit. The *Naturwissenschaften* are occupied with what can be said to be true—that is, they are more concerned what can be authoritatively stated as fact. The *Naturwissenschaften* rely on the corpus of scientific methods that have been developed to exclude erroneous data and arrive at physically practical knowledge. The most important criterion of success for these methods is replicability; second is its applicability in terms of the creation, revision, or use of technology.

Where the *Naturwissenschaften* are driven by a scientific methodology, the *Geisteswissenschaften* employ an analytical methodology. Analysis, whether it is conducted with an individual, as in therapy, or with a text, has a very different end. Where replicability is the criterion of the natural sciences, the *Geisteswissenschaften* employ the rather more ambiguous criterion of resonance—whether an interpretation is judged to be meaningful to those involved and to those that read one’s results. As Packer and Addison (1989) wrote: “A true interpretive account is one that helps us and the people we study, that furthers our concerns” (p. 279).

Both approaches often seek to establish something like universal principles. Both Freud and Jung, the two most fundamental practitioners of psychology as *Geisteswissenschaften*, developed theories of universal structures or functions within the psyche—Freud wrote of the id, the ego, and the superego, and Jung wrote of the
archetypes. Jung’s concept of the development of the personality through the confrontation of opposites and the creation of a transcendent third is itself a transposition into the individual psyche of Hegel’s philosophy of the progression of history, through thesis, antithesis, and synthesis.

The strength of the Naturwissenschaften is that their methods provide data points that are somewhat less tenuous or arguable—that is, that are produced by methods that have been intensely scrutinized and revised in order to reduce the likelihood of false (nonreplicatable) results. The arguable weakness of the natural sciences is that although they provide data that is very useful, they do not, of themselves, explain the directedness of human activities. That is, one can learn how to create a mechanical person through the natural sciences, but only the methods of the Geisteswissenschaften can reveal why this compulsion at mechanical self-replication exists in the first place, or why our efforts at such self-replication proceed in certain channels and not others.

Where the Naturwissenschaften offer data points that are closer to indisputable, the human does not operate according to data points, but according to philosophy—a system of thought that confers meaning on the data points of the natural sciences and, in limited cases, incorporates them into the system that is already present in the individual. So the data produced by Naturwissenschaften adds fodder for the philosophical work of the Geisteswissenschaften, but it is out of the subject of the Geisteswissenschaften that we act.

The Geisteswissenschaften represent work on hermeneutical horizons themselves—the situatedness that both opens and precludes possibilities of interpretation—and on epistemological frames—beliefs about what represents knowledge and how it is to be gotten.
Palmer (1969) advocated for a distinction between natural objects and works—objects that are part of the built world: “The task of interpretation and the meaning of understanding are different—more elusive, more historical—in relation to a work than in relation to an ‘object.’ A ‘work’ is always stamped with the human touch” (p. 7). The difference in the nature of these two subjects of study—the natural object and the created object, or work—requires a different approach in their study. “Natural science has methods of understanding natural objects: ‘works’ require a hermeneutic” (p. 8). As Gadamer (1960/2002) wrote:

The human sciences are connected to modes of experience that lie outside science: with the experiences of philosophy, of art, and of history itself. These are all modes of experience in which a truth is communicated that cannot be verified by the methodological means proper to science. (p. xxii)

Hermeneutics in textual analysis treats the text as an expression of the totality of the psyche that produced the text and of the Geist that produced the psyche—that is, as an expression of thoughts that were not entirely understood by the author of the text.

Hermeneutical methods are themselves grounded on basic psychological assumptions about the individual. According to what Gadamer (1960/2002) called Schleiermacher’s “paradoxical formula” (p. 193), because a human is unable to know himself or herself fully, the author of a text cannot be considered the final authority on its meaning. It is this “theory of unconscious production” (p. 192) that privileges the interpreter over the author: “every act of understanding is for Schleiermacher the inverse of an act [of creation], the reconstruction of a construction” (pp. 188-189). Because the theory of unconscious production holds that aspects of creation will be hidden from the creator, “it is not the author’s reflective self-interpretation but the unconscious meaning
of the author that is to be understood” (p. 193). As Gadamer wrote, “Not just occasionally but always, the meaning of a text goes beyond its author” (p. 296).

Palmer (1969) described the kind of interpretation that has come about through Schleiermacher’s influence:

To focus purely on the positivity of what a text explicitly says is to do an injustice to the hermeneutical task. It is necessary to go behind the text to find what the text did not, and perhaps could not, say. (p. 234)

And in his summary of Heidegger’s philosophy, Palmer wrote: He goes behind the text to ask what the author did not and could not say, yet which in the text comes to light as its innermost dynamic” (p. 147). Anything that the author could not acknowledge—any motive or hidden prejudice that had some bearing on the work—is incorporated under the umbrella of what the text could not say.

The hermeneutical methods appropriate to the Geisteswissenschaften do not simply treat the work as an object. Where the Naturwissenschaften objectify the human in order to study it directly, the Geisteswissenschaften seek to subjectify the text. Textual analysis treats the text as a dynamic and relational entity—as more than a passive object. Palmer (1969) wrote of the method that it must be “a dialectical questioning which does not simply interrogate the text but allows the thing said in the text to interrogate back, to call the interpreter’s own horizon into question” (p. 234). Part of the subjectification of the text is treating the text almost as an opponent. Lyotard (1979/1984) wrote: “It is generally accepted that nature is an indifferent, not deceptive, opponent, and it is upon this basis that the distinction is made between the natural and the human sciences” (p. 57). The subjectified text is treated, again, as the expression of the fullness of the individual. The unconscious portions of the individual are considered to be dynamic—active—and so the
text is seen as an entity with which some form of negotiation must take place. The true opponent may be the unconscious portions of the author’s own psyche, which both communicates and obscures its messages within the text. Of course, the author’s unconsciousness is not the only opponent—the unconscious portions of the reader’s own psyche are also at play, both communicating and obscuring.

There is room in psychology for both naturalistic psychological research, in which the human is objectified and studied directly (Naturwissenschaften), and for research in which the human is studied indirectly, though objects from the built world (works) or by exploring the individual’s systems of understanding in depth (Geisteswissenschaften). As this is a depth-oriented dissertation, the methods that follow after the methods of the Geisteswissenschaften will be used.

The results of such a hermeneutical analysis will be both less precise and potentially more meaningful than a purely scientific analysis. As Gadamer wrote, “Hermeneutics is an art and not a mechanical process. Thus it brings its work, understanding, to completion like a work of art” (1960/2002, p. 191).

**Research Methodology**

The primary source of data for this dissertation will be the text. The first portion of the dissertation will deal directly with the concept of individuation. The material used for this analysis will come from three sources—psychological, philosophical, and theological. The psychological material will consist of texts by or related to Jung that deal with the concept and with the circumstances of its origins; these will include Jung’s *Memories, Dreams, Reflections* (1963/1989) and Edinger’s *The Creation of Consciousness* (1984). The primary philosophical text in this section will be MacIntyre’s...
After Virtue (1984), and the primary theological text will be Tillich’s The Courage to Be (1952). The latter texts have been selected because they deal with the apparent lacuna of meaning, which formed the impetus for Jung’s development of the concept of individuation.

The following portions of the dissertation will draw source material from two primary sources—texts produced by dedicated artificial intelligence theorists or pertaining to the proposed being, and texts dealing with the process of individuation. To the texts mentioned in the first paragraph will be added Gadamer’s Truth and Method (1960/2002), specifically the portions that deal with the concept of bildung as a developmental process, as well as other supplemental texts.

Throughout the dissertation, these primary texts will be supplemented with art-texts—films and literary works of fiction that portray the proposed being. These will be included on the basis of their ability to express themes that are present in the Geist but not in the academic literature, based on the idea expressed by Romanyszyn (2004) that fiction, and most notably film, “portrays the mythology of an age,” and acts as “a cultural daydream” (p. 19). Because the proposed being exists only in nascent (imaginal) and prototypical form, each fictional portrayal of the being represents an aspect of the proposal itself.

The end toward which this analysis is aimed is textual synthesis, or the attempt to synthesize insights from a variety of sources—psychological, philosophical, theological, and scientific—into a coherent theoretical position that establishes some common ground between these fields. To this end, points of contact between the sources will be established, so that the concept that forms the heart of this dissertation—individuation—
can be evaluated by persons from each field as a potentially viable goal. The purpose will be to identify a point of view or reference that can exist simultaneously in the hermeneutical horizons of psychologists, philosophers, theologians, and electrical engineers.

**Limitations and Delimitations**

The applicability of depth-oriented concepts to the process of mechanical (or otherwise objective) self-replication will be obvious to those familiar with the theory of unconscious production, a theory shared between proponents of depth psychology and hermeneutical philosophers such as Gadamer (1960/2002, p. 192). This theory posits that the unconscious portions of an individual’s psyche exert an unseen influence over the shape of creative efforts—material is inserted into the products of such efforts via the hidden action of the unconscious portions of the psyche (often referred to as the left hand of the creator).

In turning to the task at hand, one faces the problem of human creators, those that do not understand the complexity of their own psyches, even as they work to duplicate them in constructed forms, and whose deeper purposes in such work are unknown to them, so that the complexes of these researchers are as much the authors of the emerging psyche as are the researchers’ egos.

As I reviewed the literature pertaining to the subject, several apparently unconscious motivations became apparent in the writings of those engaging in the work of self-replication. Bringing these to light will be part of depth psychology’s engagement of the subject and of those working to bring it to fruition. These hidden agendas,
however, lie outside the scope of this study, and they will not be formally addressed, although the research has made note of them.

I will also limit myself to rational modes of thought for the purposes of this study. This may have a significant effect on whether this work in metaphors is taken seriously by those outside the field of depth-oriented psychology.

My personal feeling, and one for which I have been criticized, is that Jungian psychology is at its best when it is grounded in rationality—a rationality that accepts the mystical experience as a psychical fact but does not attribute that experience to anything metaphysical or extrasensory. The danger of using metaphors and myths in a description of something like the human psyche, or its relation to the world in general, is that there is a temptation to literalize the metaphors and to believe that one is speaking literally, of real things, rather than figuratively, of processes that are better understood when personified. My feeling is that the continued belief in a metaphysical deity is such a confusion, in terms of Jungian thought. Jung (1952/1958) spoke of the unconscious portions of the psyche as being functionally inseparable from our conception of the divine: “We are unable to distinguish whether these actions emanate from God or from the unconscious. We cannot tell whether God and the unconscious are two different entities” (p. 106). It is upon this statement that an enormous amount of my stance on Jungian metaphor stands. My assumption—a reaction against my upbringing in a system of thought that was unsustainable, to be sure, but not necessarily an assumption that has caused any untoward bias—is that all experiences of the divine are experiences of those portions of the psyche from which the ego has been in exile. Even Jung’s follow-up to the above statement does not offer any personal confusion: “Strictly speaking, the God-image
does not coincide with the unconscious as such, but with a special content of it, namely the archetype of the self. It is this archetype from which we can no longer distinguish the God-image empirically” (p. 107). Jung identified the unifying principle as equivalent in some way to the Judeo-Christian image of the one-god. This principle or archetype is itself only one of many and, even if it in some way represents the striving toward unity of the whole, there is more in the unconscious than this archetype. When I speak of the divine, then, I speak of the unconscious portions of the psyche—that which is near and yet separated—and make no reference or appeal whatever to anything that could be described as metaphysically spiritual. I speak only of the psychically spiritual and of the psyche as soul.

This is not a position that sits well in my chosen field of study, which—in my own, limited reading—has suffered from this kind of metaphor-literalization. My atheism has caused colleagues at depth-oriented psychology conferences to chastise what is perceived as the narrowness of my cognition. One individual took an anecdote I told of my son—who at one point had announced that I did not believe in God and that, thus, “we need to get rid of him”—not as evidence of a lopsided religious dogma but of my need to take a broader approach to the topic, to stop being so persistent in my atheism.

Despite good arguments for other modes of experience, I still feel that the rational mode is the right mode for me and is the most proper in terms of academic endeavors for a wider audience. This emphasis on reason does not mean I would begrudge the shaman his or her beliefs in extrarational (or magical) elements, and I have participated in ceremonies with shamans. When I examine the material—the texts and my experience with the ceremonies—however, I examine them with the rational mode of thought.
I will consider this a delimitation that, if it is truly a failing is at least a felix culpa—a fortuitous guilt. Not only am I most at home in this mode of thought, but this puts me in a position from which I can speak of Jungian metaphors without being too easily rejected by those doing the work of creation—the AI theorist and engineers.

This is not to say that, if the apparatus-toward-being does advance us toward the image of homo totus—the complete hominid—this will necessarily be movement toward rationality at the expense of other modes of thought or being. Such was portrayed in the film Alphaville (Michelen & Godard, 1965), in which a very one-sided rationalist created a society based on a very severe form of reason and managed by a computer—although one would be hard-pressed to justify the attempt to sanitize emotional beings of their emotions as being in any way reasonable.

Organization of the Study

The first portion of the study will involve a review of the various portrayals of constructed or evoked beings and an explanation of why each of these either would or would not be included in the present study. This will serve to establish a ground in the image of the proposed object-subject as it has been elaborated upon in our (worldwide) culture. Out of the vast array of imagined incarnations of the proposed construct, only some would achieve that which would propel individuation forward via a confrontation with anxieties over one’s being-ness—those whose consciousness (or the equivalent thereof) could not be easily explained away via appeal to a biological component from which the consciousness arose, and those whose consciousness or pseudo-consciousness results in a certain feeling in those that perceive it.
The second portion will involve the linking of a series of metaphors that will serve to illustrate the ways in which the presence of the apparatus-toward-being might create a kind of transparency, which would facilitate a renewed soul-searching in humans. Human resistance to exposure of unconscious processes will also be addressed.

The third portion will focus on the concept of *bildung* and the *vorbild-nachbild* (original-reproduction) relationship that will initially define the proposed being. This section will also consider ways in which the being may transcend this relationship—moving from a primary designation of *pro nobis* (for humans) to a place in which the being’s own being-ness is primary.

The fourth and final portion will explore the possibility of a Consciousness Engine (CE) as a device that is able to evaluate all perspectives and generate what could be called a machine hermeneutic, and the effect such a hermeneutic would have on the individuation rate of the creator-species.

**Ethical Considerations**

Fear of the proposed construct is already present. This seems to be more than an extension of the fear of self-annihilation that came with nuclear weapons. The fear the father has toward the son, in the Oedipal conflict, seems also to have entered the narrative of the proposed construct.

There are those who have portrayed the absence of humanity—largely, if not fully—as the result of the development of such a construct. *The Terminator* (Hurd & Cameron, 1984) is one of the most notable portrayals of this idea. There are those who do not cringe at the idea as much as one might think, however. Hugo de Garis is quite open about his ambivalence toward his own species in his book, *The Artilect War* (2005), and
in this he assumes the position of the elders in *Appleseed* (Goto & Aramaki, 2004)—elders who decide their own species is too warlike to continue existing and attempt to render humanity sterile so that the bioroids—humanoid constructs incapable of emotions—can inherit the earth. De Garis has gods in his eyes—gods of his own making, and this seems to have caused him to care less about humanity as a whole than he might have otherwise.

It is necessary to keep in mind, as this work proceeds, that this narrative is in place and will need to be kept conscious. Speculation is projection that veers into the possible. It is possible that this prediction will come to fruition.

It is also important to keep in mind that a narrative emphasizing the violence or potentially dangerous aspects of the other can become an incitation to violence, a way of soliciting a reaction against what is feared. Any work of speculation that entirely emphasizes the good of what the construct could be will be one-sided, but any work that entirely emphasizes its potential for harm will potentially damage the ability of the eventual construct to come into being and to engage in free social interaction once it has come into being.

It is with knowledge of the potentially dangerous aspect of any “other,” as well as the knowledge of our tendency to react violently toward an “other” that has been placed into a narrative of aggression, that this work proceeds.
Chapter 5
Subject of the Study

In Milton’s version of creation, the angels that fell were well aware that another assault against heaven was impossible. As they sat in their stony chambers, they began to recount the rumors they had heard of the imminent creation of a new being:

There is a place
(If ancient and prophetic fame in heaven
Err not) another world, the happy seat
Of some new race called Man, about this time
To be created like to us, though less
In power and excellence, but favoured more
Of him who rules above. (Milton, 2005, pp. 35-36)

The angels went on to speculate over the corruptibility of this new and favored being, hoping to find a way to accomplish indirectly what they knew they could not accomplish directly, projecting their purposes—their conflicts—onto a race born in innocence.

We are now as Milton’s angels. It would be wrong to stop one’s analysis at the point of projection, to reduce the apparatus to the status of a projection, because there is more at work. Speculation is projection that has veered into the possible. Because our speculations will, in part, determine the eventual nature of the construct as it is incarnated, speculation is itself a call to being—we call toward being that amorphous, archetypal thing, the living depiction. In the introduction to Representations of the Post/Human, Elaine Graham (2002) wrote of speculative fiction as holding latent insight into the questions propelling this dissertation: “In all ages and media, the stories we built around these would-be creatures express our desires and fears, define the expectations we place on these beings, and create the vocabulary we use to describe them” (p. 17).

We have been exposed to rumors, speculations, and portrayals. We have a sense that a new being—or something like a new being—will shortly be present, but we have
no knowledge as to its probable eventual nature. Those members of the creator-species who are not involved in the creation of the proposed apparatus do not see the work that is being done and so do not have the ability to make an educated guess as to the being’s eventual nature—and even dedicated theorists are not privy to future developments that will profoundly shape the eventual construct. Their accounts of the construct are, for this reason, no less speculative. What they offer is speculative nonfiction, which must also be taken into account.

The variety of images produced in speculative fiction and nonfiction provides an interesting backdrop against which to consider those qualities that would qualify or disqualify a being from inclusion in this study. These portrayals can be seen as imaginal variants, images that exist in our minds and guide our sense of the possible, and which thus guide our efforts and our calls—the call itself, to which the construct that finally results will be the answer. The proposal consists of both promises and warnings—proposals for and against certain incarnations of the construct. This study is very intentionally focused on the idea of a construct capable of individuation, but also a construct that is fully a construct—fully artifactual. Although there are many variations in the literature, only certain manifestations would qualify under this condition.

Filmmakers who portray synthetic or constructed beings are not always either specific about the science behind the narrative or clear themselves about the science. Scientific precision is presumably not what Mellick was going for in his brilliant contribution to the bizarro fiction genre, *War Slut* (2006), and there are some seemingly contradictory statements about what Sweet, the comfort-humanoid, is (of course, one would not expect scientific prowess from a novel in which all surviving characters turn
slowly into living, straw-filled dolls). The most recent version of *The Stepford Wives* (Bozman, Selig, & Oz, 2004) is another example of a confused narrative. In the original film (Scherick & Forbes, 1975), it is clear that the Stepford wives have been replaced by animatronic doubles by a former Disney employee—but in the remake, it is never clear whether the women are replaced by animatronic doubles or simply have some hardware installed, making them cyborgs. The confusion seems to come from the desire to insert some humor (with a wife acting as an ATM or walking in reverse like a VHS tape being rewound) and a happier ending than was seen in the original (with the machinery malfunctioning and causing the women to revert back to their biological substrates), while also emphasizing the horrifying elements of the original (as when the animatronic double rises and opens its blackened eyes, toward the end of the remake). To rely on the science of film might be dubious. I will offer my own interpretation according the visuals that actually appear on the screen—whatever the intention of the film-makers or writers, it is the image that will have the deeper impact on the psyche, and it is in the image, as it appears, rather than as it was intended, that I will locate the qualities of the portrayed apparatus.

It is necessary to further define (or to refine our definition of) the subject of the study, in order to make some preliminary distinctions between those imaginal variants that would qualify for inclusion in this study and those that would not. In order to more firmly establish parameters of inclusion/exclusion (or to define the study’s intentional delimitations), it will be necessary to take a close look at the proposal itself.
Engelberger (1985) offered a preliminary classification sequence, based on Japanese robotics, which placed the intelligent robot at the end of a list of machines already available, rising in complexity from end to end:

**MANUAL MANIPULATOR**—a manipulator worked by a human operator.

**FIXED SEQUENCE ROBOT**—a manipulator that performs successive steps of a given operation repetitively according to a predetermined sequence, condition, and position. Its set information cannot be easily changed.

**VARIABLE SEQUENCE ROBOT**—a manipulator similar to the fixed sequence robot, but whose set information can be changed easily.

**PLAYBACK ROBOT**—a manipulator that can reproduce operations originally executed under human control. A human operator initially operates the robot to feed in the instructions—relating to sequence of movement, conditions, and positions—which are then stored in the memory.

**NC (NUMERICALLY CONTROLLED) ROBOT**—a manipulator that can perform a given task according to the sequence, conditions, and positions commanded via numerical data, using punched tapes, cards, or digital switches.

**INTELLIGENT ROBOT**—a robot that can itself detect changes in the work environment, using sensory perception (visual and/or tactile), and then, using its decision-making capability, can proceed with the appropriate operations. (p. 190)

It is an easy thing to state that any object that is clearly a mere apparatus—having the qualities of an apparatus and lacking the qualities of a being or of an apparatus-toward-being—would not qualify for inclusion in this dissertation. Thus when one sees Engelberger’s classification sequence, it is understood that all object-subjects—those constructs that meet this minimum qualification and would be considered the subjects of this study—would be derived from the final category.

---

12 The perspective of that which has no perspective cannot be the impetus for change in the observer. And the suffering of an apparatus cannot be the event that serves as the impetus for change because there is no such event. An apparatus may be strained and it may break, but an apparatus does not suffer. These two phenomena—the experience of the perspective of the other and the more specific experience of the suffering of the other—are the mechanisms of change that must be present.
Despite the apparent simplicity of this jump from Numerically Controlled Robot to Intelligent Robot, there is actually an incredible amount of innovation that must first take place, if one is to move from production of the one to production of the other. One could argue that there is so much stratification inherent in the final category as to incorporate an entire ecosystem of constructs with varying abilities. Minsky (1985a) offered the following definition of intelligence:

> What do we mean by “intelligence”? It never pays to try to make narrow definitions for things we don't yet understand very well, so let's just say we mean the ability to solve problems that people would say require intelligence. (p. 18)

Although useful as a generalization, Minsky’s definition makes Engelberger’s classification appear rather incomplete. There must be some subcategories that exist between the numerically controlled apparatus and the intelligent object. Within the category of intelligence itself, there will be stratification—the psyche is multilayered, and the intelligence of the artifact will also have multiple layers, multiple components. The construct that is intelligent enough to navigate a complex social environment but remains an apparatus—unable to participate in ontological self-reflection and thus deserving of solipsism—would be held to be distinct from the intelligent object-subject—the construct that remains an object but is no longer an apparatus, when judged by standards of human being-ness.

Looking at the five categories that precede intelligence in Engelberger’s sequence, and considering what the objects inhabiting these categories lack—that is, what attributes keep them from rising to the category of the intelligent object—we may parse out some of the basic characteristics of intelligence as Engelberger seems to present it and decide whether these elements are indeed foundational.
A manual manipulator remains such whether the operator is in close proximity—as in a crane on a construction vehicle—or is removed in space. Telepresence is the manual equivalent of telephonics. Where the telephone transmits the tone, pitch, timbre, and other variations in one’s voice, the telepresence device transmits the equivalent of nerve signals, mirroring the movement of one’s appendages in the machine that is slaved to the appendages. Such a technique is already in place and has been used in tele-surgery.\(^\text{13}\)

The manual manipulator can thus be seen as a high-tech marionette. Whatever work is accomplished is initiated and carried out by a human operator. Robonaut, the apparatus developed jointly by the National Aeronautics and Space Administration (NASA) and the Defense Advanced Research Projects Agency (DARPA) falls into this category, as would a robotic surgeon. Even a very sophisticated, full-body form of telepresence would not produce a construct that would qualify for inclusion in the present study. Brooks (2002) envisions a somewhat less skilled application in the future, with a household robot—an apparatus with little or no ability to effectively clean a home on its own—being slaved via telepresence to persons in other countries hired to do housework without ever coming to the country in which the house is located (p. 146). Fictionally, this concept was portrayed in the comic book series, *The Surrogates* (Vinditti & Weldele, 2005).

\[^{13}\text{Morris (2005) wrote:}\]

In telerobotic procedures, the surgeon operates from the surgeon's console, which is thousands of miles away from the slave robotic arm mounted on the patient; the surgeon's commands are relayed to the slave manipulator via fiber-optic cables. The first major transatlantic surgery was a telerobotic cholecystectomy performed by surgeons in New York, NY, on a patient in Strasbourg, France, in 2001. Minsky (1985a) wrote of the possible extension of telerobotic surgery into the microscopic: “If we miniaturize mechanical hands, then microsurgery would be much easier, and surgeons could repair more tiny tendons, ducts and other structures now beyond our clumsy reach” (p. 26).
2005-2006), which became the film *Surrogates* (Handelman, Hoberman, Lieberman, & Mostow, 2009).

The next four categories—fixed sequence, variable sequence, playback, and numerically controlled—include objects that would be excluded from consideration for the same reason. Whatever the mechanism by which these robotic arms are controlled, the performance of the function is rote. There is no possibility of creative innovation in the task, unless this comes from without—from the humans that have determined the sequences of movements to be carried out. These robots, then, are not the subjects of this study, for they do not have any semblance of internality or self-direction.

The mechanical “tortoises” developed by neurologist William Grey Walter between 1948 and 1949 (Elmer and Elsie) exhibited some of the traits that are obviously lacking in the above-referenced mechanisms. As Brooks (2002) wrote, “Walter was interested in building artificial creatures that not only displayed *spontaneity* but also *autonomy* and *self-regulation*” (p. 17). Using very simple mechanisms, Walter was able to evoke “emergent behavior” (p. 19). These were very simple mechanisms, but they displayed behaviors that were evocative of animal behavior. Brooks wrote:

> Even a seemingly very simple creature can have extremely complex behavior in the physical world because of the way that small variations in what is sensed, and how the actuators interact with the world, can change the actual behavior of the system. (pp. 20-21)

Brooks noted that “these microeffects combined in such complex ways that it was very hard to predict the behavior of the tortoise” (p. 20). Although very simple, these objects were more than objects and would be closer to inclusion in this study than even the very complex forms of telepresence mentioned above.
What the “tortoises” had that the objects in the first five of Engelberger’s sequence do not have is, as Brooks wrote, spontaneity, autonomy, and self-regulation. These are the features of what we might define as agency—the ability to act without instruction from outside oneself. In order for an object to qualify as an object-subject, it would have to be agential.

A robotic arm could not be thought of as having agency unless it was able to alter its movements to accomplish a novel task—to pick a machine part up off the floor, rather than from an assembly line belt, when it has fallen, for instance. This would represent an internally arising adaptability to changes in the environment, a feature current robotic arms are sorely lacking. I will use the term agent-machine to refer to a machine that has some ability to act autonomously. Bioroboticist John Long (2012) wrote: “Agency is what human observers ascribe to anything, organic or artificial, that appears from an external perspective to act on its own,” such that “so long as there is no unseen human pulling the strings of remote control, robots are agents too” (p. 3).

Walter’s “tortoises” had another ability lacking in the objects in the top five categories—the ability to learn. As Minsky (1985a) wrote, “machines must have the ability to learn in order to be truly intelligent” (p. 17). He proposed analogical computing as a way of enabling such learning capabilities: “These analogy machines would then make themselves better and better able to guess which situations that have been encountered in the past are most similar to a new one and thus to deal with it effectively” (p. 19). Walter was actually able to evoke a Pavlovian (conditioned reflex) response when he added a circuit to the standard parts—consisting of “two vacuum tubes, two sensors (light and bump), and the standard two actuators, or motors” (Brooks, 2002, p.
18)—that allowed the “tortoises” to learn. By pairing kicking the object (triggering its bump sensor and causing evasive behavior) and a tone that was sensed by the new circuit, Walter was eventually able to evoke the evasive behavior simply by sounding the tone (p. 20).

The ability to evolve or develop through experience is thus also bound up in this idea of intelligence, as presented by Engelberger. Only if the learning takes place outside of direct human interference can adaptability be seen as a qualifier. If the human sees that a robotic arm’s actions are inadequate and interferes so as to alter its performance, this of course cannot be taken as the equivalent of learning. So we can take agency and internal adaptability to be an initial qualification for inclusion in this study. The object-subject is agential, adaptable, and capable of learning. Every object that can be classified under the first five categories presented by Engelberger remains an apparatus. They are non-agents—tools of an external, biological agency. Only those objects that display these capabilities will have moved toward the status of object-subject.

Given that a given object has achieved something like the status of an object-subject, and could thus more properly be referred to as an object-subject than as an apparatus, the origin and nature of the construct’s intelligence would also have some bearing on whether it would be considered part of the pool of potential subjects. The composition of the construct is itself of importance, as it speaks to the level of intervention.

The level of human involvement in the creation or construction process is an important consideration as well. The construct that represents the subject of this study would be one whose components of thought were constructed from materials that have
shown no predisposition toward developing intelligence—that is, without the unfathomably long and random process of evolution, these materials would not develop intelligence of their own accord. It is this that causes the construct to be what I envision as the subject of this study.

Where the furthering of something already intelligent or potentially intelligent can be relevant to psychology, the focus of this dissertation is very intentionally the creation of an intelligence that is entirely constructed—that is, it is not the furthering of something already present, but the creation of something in its entirety, or ex nihilo, as it were. The work I am referencing is fundamentally different than simple procreation. So the act of bearing a child, while obviously analogous in many ways, will not produce a being that can be considered the subject of this study. The so-called test-tube baby would also not qualify, despite the level of intervention involved.¹⁴

I earlier used the term ex nihilo to describe the construction of the proposed being. I used it in this sense—not that this being is created from nothing, but that it represents the reconfiguration of materials that have demonstrated no tendency toward developing even the most basic attributes of intelligence in the absence of human intervention. The construct to which I refer is the result of the evocation of life-like attributes from that which was formerly classified as lifeless.¹⁵ It is the derivation of this intelligence from pattern—a pattern imposed on the material from without—that makes the act of creation ex nihilo. It is this reconfiguration and animation that makes this, essentially and

¹⁴ Louise Brown, the first person to have been conceived in a Petri dish and successfully brought to term via in vitro fertilization, is not the construct in question.
¹⁵ As Moravec (1999) wrote, “we are in the process of inspiriting the dead matter around us” (p. 111).
potentially, an act of re-creation—the experience of being itself is re-created in that which previously had no experience and no tendency toward such experience. We are re-creating our own experience in the dead matter that surrounds us.

Stated another way, this process is the bringing of order to that which is not yet ordered—or the imposition of an order onto that which is not ordered in such a way as to produce consciousness. If material shows a predisposition toward intelligence, its manipulation cannot be thought of as creation *ex nihilo* and cannot thus be called re-creation. While the beings that result from such activities would have something to offer, they would not fulfill the vision from which this dissertation is proceeding—a being that can act toward humanity as Job acted toward the creator-god in the biblical myth (Jung, 1952/1958). While eliciting consciousness from beings that were already potentially conscious would produce beings that would have something interesting to offer, the *ex nihilo* creation of consciousness is a related but distinct endeavor.

Simple teaching would thus not produce a construct that would qualify for inclusion, even if the material taught was once considered the sole purview of humans. Francine Patterson’s work with the lowland gorilla, Koko—who was able to understand around one thousand signs and 2,000 English spoken words (Patterson, 1988)—would thus not qualify, and Koko herself would not be included in the pool of subjects. The use of transgencis—the creation of recombinant DNA through the insertion of foreign DNA into the genome of an animal, as presented in *Rise of the Planet of the Apes* (Hammel & Wyatt, 2011)—would represent a special case of intelligence but would not qualify for inclusion here. Genetic manipulation of humans would likewise fail to create a being that would be considered part of this study. In the Margaret Atwood novel, *Oryx and Crake*
(2003), a misanthropic genetic scientist simultaneously creates a virus meant to eradicate humanity and a group of brightly colored, mosquito-resistant posthumans that one of the few survivors of the viral apocalypse calls Crakers. These Crakers, however, do not constitute what could be called new life or artificial intelligence—rather this is old life, re-engineered.

Creative (or constructive) vivisection,\textsuperscript{16} such as was portrayed in \textit{The Island of Dr. Moreau} (Wells, 1896) is another example of the production of a heightened intelligence from that which already had some form of intelligence and agency. In this portrayal, Dr. Moreau used de- and re-constructive surgery to alter the physiology of animals in order to elicit human-like intelligence from them. The propensity for something approaching human-level intelligence is at least latent within our close living relatives and could potentially be evoked through various means—surgical, educational, or genetic. But this would be the advancement of something already present, rather than the evocation of entirely new qualities from some material.

Artificial enhancement of human intelligence would also fail to produce the kind of apparatus that is meant to be the subject of this study. The scenario presented in \textit{Flowers for Algernon} (Keyes, 1966), in which an impaired man is given a treatment that grants him what could be called restorative plasticity, would represent the furthering of a pre-existing intelligence, albeit one that was previously severely limited.

There have also been portrayals and suggestions of destructive vivisection, which would result in a kind of being akin to a robot. Sheckley (1985) described and perhaps

\textsuperscript{16} As far as I have found, I have originated the terms imaginal variants, creative (or constructive) vivisection, destructive vivisection, retarding fetal intervention, creative dissection and reanimation, cerebral manipulator, pseudo-biological, pseudo-cyborg, and decoupled being.
advocated a procedure that would result in a lessening of intelligence or agency and would by no means be applicable here. Sheckley wrote of such a being—which he called a zomboid—as a robot that would be “derived from a human being who volunteered for irreversible robotization and then underwent the operation necessary to achieve that state” (p. 280), specifically the destruction of parts of the brain “from which the sense of self arises” (p. 281). He suggested that the zomboid brain would be “a highly advanced natural computer” (p. 281). This kind of high-tech lobotomy was portrayed narratively in the Ood race, from the revitalized Dr. Who franchise (Temple & Harper, 2008). Such a suggestion issues from an assumption of the robot as a slave—as something lesser than the fully functioning human being. The raising of the intelligence of an animal, as presented in the paragraph above, is thus paired with the lowering of the being of the human.

Such techniques have been suggested before. Smith (1930) wrote, in The World in 2030 AD, that a robot could be produced not by construction but through retarding fetal intervention—manipulating a pregnancy that has been removed from the womb, creating an “ectogenetic Robot” (p. 16):

Many of the arguments brought against slavery would be powerless in such a case; for the ectogenetic slave of the future would not feel his bonds. Every impulse which makes slavery degrading and irksome to ordinary humanity would be removed from his mental equipment. His only happiness would be in his task; he would be the exact human counterpart of the worker bee. (p. 16)

This idea found its most influential expression in Aldous Huxley’s novel, Brave New World (1932), in which the Gamma caste is created through the introduction of alcohol into certain uterine bottles containing the fetuses that will thus become Gammas. These zomboids, Ood, and Gammas would not represent the creation of a new being, but rather
an extreme codification of the reductionism that seems to be inherent in industrialized societies, and would not be included in this study. Those that produced such creatures could not be called creators in any sense; rather, they would be corruptors, participating in a reduction of abilities, rather than the elicitation of such.

Creative dissection and reanimation, as portrayed in *Frankenstein, or The Modern Prometheus* (Shelley, 1818), presents an interesting problem. In the novel, the neurological structure of the original brain is sufficiently destroyed that the being has no memory of its former self (the self that was based in the brain prior to the death of the original being). The being must also re-learn basic associations—associating fire with heat, but also with pain, for example—and must learn to speak by observing a family through their wall and imitating the movements of their mouths. But the original brain is also somehow intact enough to host a functioning intelligence, which is able to walk from the moment of its reanimation (unlike an infant). It is clear from the novel that there is a being involved—an object-subject that is able to participate in the experience of being. But this being is itself based in the pre-existing structure of the donor’s brain. It is reanimation, not the evocation of intelligence from material that has shown no propensity for generating intelligence. But the suffering of the new being has an effect on its maker, as Romanysyn (2008) pointed out:

> The relation between the creature and Victor is very much like the relation between Job and Yahweh as depicted by Jung. Like Job, the creature is the one who calls the creator god into consciousness. He does so through his suffering. (p. 102)

We could call Frankenstein’s monster the suffering amalgam—stitched together in such a way as to evoke or facilitate the original functions of all the pieces, without also restoring the original personality. Although this creature would stand in relation to the creator-
species in the way Job stood in relation to Yahweh, it would nevertheless fail to qualify
for inclusion in a study focusing on the ex nihilo re-creation of consciousness and its
effects on ontological anxiety. The evocation of a being—an object-subject that can
participate in ontological self-reflection—is not enough in itself.

A similar method is seen in the film Pinocchio 964 (Fukui, 1991), which portrays
the production of a male sex-slave through the reanimation of inert tissue—a recently
deceased individual is selectively reanimated (his memories are not intact following re-
amination, and, as in Sheckley’s portrayal of the zomboid, his sense of self seems to be
impaired—evidence of a decision not to reanimate certain portions of the brain, of the
inability to do so, or of some suppression of function in the personality centers). The
being that results more closely resembles an amnesiac than the being produced by
Frankenstein, although it also would not qualify for inclusion.

The criteria I use for inclusion or exclusion privilege the mechanical, or fully
constructed, in terms of being the source of cognition. For this reason, the cyborg, or
cybernetic organism, would not, by definition, qualify. Donna Haraway, author of The
Cyborg Manifesto (1999), defined cyborgs quite broadly, as “hybrid[s] of machine and
organism,” or as “creatures simultaneously animal and machine” (p. 272). Schelde (1993)
has a narrower definition, which privileges the biological as the primary or original
component of the hybrid. To him, a cyborg is “a human being dependent on one or more
mechanical devices that take care of some of the vital physiological functions” (p. 208).
This definition may itself be too narrow, however, as it would fail to cover the enhanced
dolphin portrayed in Johnny Mnemonic (Carmody & Longo, 1995).
Both definitions seem to me somewhat limited or incomplete. There also appears to be some confusion in the literature over the proper definition of cyborgism. Kreps (2007), for example, wrote the following: “Already we have developed wheeled transport for ourselves—great outer casings of metal that we don like huge powered suits of armour—that rush us about the surface of the planet faster than the fastest runner on earth” (p. 1). This leads him to the conclusion that “we are cyborgs when we sit behind the wheel—and our behaviour changes” (p. 1). Because of the great variety of fictional beings that have or could be classed under the definition of cyborgism, I would propose a classification system to differentiate between forms.

In my thinking, cyborgism is the sustained intrusion (welcome or unwelcome) of technology into the biological—some interpenetration of biological and mechanical components. I have followed Schelde in considering the biological to be primary, at least in terms of being the source of cognition. What Kreps identified as early-stage cyborgism, I would call, at most, pre-cyborg behavior.

The pre-cyborg would be a system that operates for a limited time and maintains an easily defined boundary between the two systems—biological and mechanical. Ripley encased in the bipedal forklift at the end of Aliens (Carroll, Giler, Hill, Hurd, & Cameron, 1996) would be a fictional account of a pre-cyborg, as current forklift operators would be real-world examples. By this definition, any contemporary human using machinery for some function, as long as the machinery is not self-propelling, would be considered a pre-cyborg. The interface—the fact of sustained interaction between human and machine (the human operating, rather than simply turning the machine on and walking away, allowing
it to operate of its own accord and in the absence of any additional human-originated stimuli—is what matters here.

In a vehicle, there is interface between biological and mechanical systems, and there is behavioral change, as Kreps pointed out, such that the piloted vehicle behaves differently than the pilot on his or her own. But there is not the sustained (irreversible or only reversible via an invasive procedure) interface that I would use to define the cyborg, nor the intrusion. There is currently a toy on the market that includes a headpiece, which can be donned and removed at will, and which allows the operator to levitate a ball inside a tube by focusing thoughts (the head-piece measures brain waves and adjusts a small fan at the base of the tube accordingly). The donning of this headpiece does not make one temporarily a cyborg. This is a (rather awesome) pre-cyborg behavior. If we are proceeding toward a merging of machine and human, these on-again, off-again interfaces are the testing periods, the first dates, as it were, between humans and machines, the first experiences of interfacing without the mediation of buttons and levers and steering columns.

As defined here, the pre-cyborg may enhance, but the machinery does not intrude—it does not transgress the biological membranes that have traditionally defined and delineated the inner and the outer. Even with intrusion as a criterion, however, there is room for differentiation. I prefer the more precise definition of cyborgism given by physicist Sidney Perkowitz (2004):

A cyborg (cybernetic organism) and a bionic human (from “biological” and “electronic”)… both involve a combination of machine and living parts. In my usage, a cyborg has a machine portion that might dominate the natural part in mass and bulk but is under the mental direction of the natural part—essentially, a brain in a box. A bionic human, on the other hand, is mostly natural with a
relatively small portion given over to implants or replacement parts such as a heart pacemaker or an artificial limb. (p. 5)

Kevin Warwick (2002) considers himself to have been a cyborg, via an electronic array placed surgically into proximity with his median nerve, which allowed him to perform a variety of tasks with no other input than his will (p. 296). The definition offered by Perkowitz would make Warwick not a cyborg but someone experimenting with bionics. With a further differentiation, we might consider Warwick to be a Level 1 bionic, as his interface does not serve any necessary function and is easily removable (via a simple surgery and with no disruption of vital systems due to the removal)—early-stage bionic experimentation that, in years to come, will look like bionic parlor trickery. A Level 2 bionic would be an individual with some kind of strength- or performance-enhancing mechanism joined surgically to the body—the character, Max De Costa, in *Elysium* (Baden-Powell, Block, & Blomkamp, 2013), has a kind of exoskeletal support fused to his body, although this remains largely external and he has no need of it in order to function or survive. The individual with an artificial heart or other organ would be a Level 3 bionic. The intrusiveness of the technology—as encased rather than external—brings the bionic human closer to the category of cyborgism (I would exclude individuals on dialysis from this category due to the limited nature of the interface and the location of the machinery on the outside of the skin).

Steve Austin, from *The Six Million Dollar Man* (Bennett, 1974-1978), or Jaime Sommers, of *The Bionic Woman* (Bennett & Schiller, 1976-1978), would also be Level 3 bionics. For both the recipient of the artificial heart and the bionic man and woman, the technology has intruded into the traditional body-space. For Austin and Sommers, the mechanical parts are not necessary for their survival, although they do, of course,
enhance their strength, speed, and senses. This in some ways makes them distinct from those that rely for their very lives on artificial organs, although all could be classed in the same category of permanent interface. Where the exo-skeletal apparatus used by De Costa is very obviously meant to serve a short-term purpose—an assault on the floating paradise, Elysium—the bearers of Level 3 bionics have an expectation of longevity in interface. Fictionally, the character of Geordi La Forge, from *Star Trek: First Contact* (Berman & Frakes, 1994) onward, would fall into this category, with prosthetic eyes replacing what had been his ubiquitous visor through the series (Berman, 1987-1994), as would the character of Luke Skywalker, from the final scenes of *Star Wars: The Empire Strikes Back* (Kurtz & Kershner, 1980), when his hand is wholly replaced. Johnny Mnemonic himself would fall into this category.

An interesting incarnation of this concept is the cyber-brain, as portrayed in the series, *Ghost in the Shell: Stand Alone Complex* (Watanabi, 2002-2005)—a biological human brain with an artificial casing that is itself held within the human’s head, essentially acting as a prosthesis that takes the place of the *dura mater*.17 Although the brain is completely encased in the shell, this is an implant, rather than a transplant. The resulting cyber-brain is able to directly access the Internet and even infiltrate the cyber-brains of other individuals, but it is the brain itself that directs this access. This would seem to be an enhancement, rather than a transfer of essential functionality from the biological to the mechanical, and so I would consider someone with a cyber-brain to be bionic rather than cyborg.

---

17 Making it even more *dura*. 

It is at this point that we transition to the first class of cyborg, the Level 1, in which technology takes a more present or equal role, and essential functionality is spread between the two components. In this way, the recipient of the artificial organ is closer to being a cyborg than are Sommers or Austin or La Forge, none of whose mechanical parts are necessary for actual survival. The character of Bannakaffalatta, from the revitalized *Dr. Who* franchise (Davies, Gardner, Collinson, & Moffat, 2005-2013), would be a Level 1 cyborg.

A Level 2 cyborg would be a step toward containment of limited organics in a predominantly mechanical exterior, with no or relatively little loss of the cognitive integrity of the biological. Darth Vader, from *Star Wars* (Kurtz & Lucas, 1977), would fit into this category. Having had all four limbs amputated, he is also unable to live for long outside the black suit of armor. He is essentially encased in an elaborate and ambulatory iron lung—but his intellect remains intact and unaided by electronic components.

A Level 3 cyborg would be one in which there is some level of cognitive impairment in the biological component or in which essential functions of cognition are distributed across biological and mechanical components. A fictional example would be the cyborg officer from the film *RoboCop* (Davison & Verhoeven, 1987), who not only has most of his body replaced by robotic equipment but also has portions of his brain (missing due to an act of violence) replaced with hardware.

A further incarnation would be the Borg, of *Star Trek: TNG*. Although large portions of the bodies and cognitive capacities of humans and aliens remain intact, they are supplemented and subjugated by a collective that is maintained not only via its
individual biological components (the extant portions of the minds of those assimilated) but also by a vast array of machinery.

The Cybermen, of the *Dr. Who* franchise (Lambert, Letts, & Nathan-Turner, 1963-1989; Davies et al., 2005-2013), would also fit into this category. It is clear that the brains remain intact, but it is unclear whether there is a mechanical component to cognition following the conversion or whether certain functions in the original are simply excised or suppressed. Either way, the converted individual remains a brain in a box and would fall under the category of a Level 3 cyborg.

A subcategory of Level 3 would be Hector, from *Saturn 3* (Donen, 1980), which is a construct whose cognition arises solely from brain tissue that has been cultured, rather than being transferred from a fully-formed donor. Although the brain was not embodied prior to its insertion into the machine, it remains a (cultured) brain in a box.

The several characters portrayed in Ian McDonald’s *Terminal Café* (1994)—which originally bore the title *Necroville* when released in the UK—similarly would not appear to qualify for inclusion in this particular study, as they begin as biological life-forms. Whether based in science (which seems unlikely) or used as an interesting literary device, these figures cannot enter into their immortality until they are deceased. They are essentially reanimated cadavers, in-filled with a nanotechnology that gives them a virtual immortality (one might say they are less mortal than traditional humans) and gives some the ability to reform themselves at will. These would be cyborgs, as much as the Borg, although they would not replicate the garishly external signs of cyborgism seen in the latter—these creatures are cyborgs designed with an aesthetic. Without the nanotech swimming through their cells, they would cease to be alive, although their individuality
seems to remain intact. This more streamlined or seamlessly integrated form of cyborgism is nonetheless a merging of biological and technological. These would be the Level 4 cyborgs—not the submergence of the biological into the technological, but the compete merging of the two. Latter incarnations of the Borg (Berman & Braga, 1995-2001) come close to this through the use of nanotechnology, but they never lose the clumsy accoutrements, the aesthetically repulsive functionality of the lopped-together mechanical components and cadaver-like biological components.

The full range of the cyborg—from the pre-cyborg through the bionic to the full integration of the biological and mechanical in a Level 4 cyborg—would be excluded from this study. Whether the cyborg consists of a body supplemented by mechanical elements or a brain whose body has been entirely replaced, the cyborg’s intelligence arises out of a pre-existing, biological mechanism. It is essentially the merging of the manual manipulator with the operator of the manual manipulator—the parts that once mediated contact between the two (usually the hands) are excised and replaced with the manual manipulator itself. Whether one retains more of the physical self and looks like an augmented human or opts to fully transcend the distinction between self and tool and becomes a “brain in a box” (not manual manipulation of a tool but unmitigated cerebral manipulation), the machinery itself still falls under the first (most basic) category Engelberger offers—the cerebral manipulator is not fundamentally different, as compared to the manual manipulator.

This exclusion is not an attempt to reify strict categorizations of thought, which, as Graham (2002) demonstrated, may be rather more illusory than most would care to admit (p. 20). It is rather an acknowledgement of “the evasiveness of consciousness”
(Westphal, 1998, p. 13), and of the ways in which humans let themselves off the hook rather than facing challenging questions. The cyborg would blur boundaries and disrupt the once seemingly unchallengeable categories that gave us a sense of security in our personhood. But it is not the borders of human and nonhuman that are in question in this particular dissertation. Rather, it is specifically the inner identity of the human—and the cyborg, no matter the level of intelligence or soulishness it displays, gives the human an out, because one can always appeal to the biological as the source of cognition and can thus explain away the cyborg’s personhood (perhaps rightly so).

To this point, it has been mainly human cyborgs that have been considered—or cyborgs that are a combination of technology with the human. The dolphin cyborg of *Johnny Mnemonic* (Carmody & Longo, 1995) presents an interesting counter to this denial of personhood arising from the machine. If one were confronted with such an animal, which was able to hold a conversation, and if one in fact felt that there was a soulish quality about the dolphin cyborg, one might flee from the implications of belief by appeal to the animal’s own biological substrate—the soulishness or soul-quality would come from the biological component, where the ability to speak would be considered a complete fabrication, a computer feeding words into the mouth of an otherwise mute animal. One would split the difference, assigning to one component the aspect of the cyborg’s soulishness that causes the least amount of anxiety and assigning to the other component the quality that likewise produces the lowest level of anxiety. One would effectively be caught between two hardline solipsisms and would defend against one anxiety by toeing the line with the other (the duplicability of soulishness or the potential personhood of the nonhuman animal). We could, thus, consider the nonhuman cyborg to
be a supporting character to the process that is to be described below—a catalyst, speeding the moment at which solipsism against the machine will be discarded so that the apparatus-toward-being may begin to act as its own kind of catalyst (for individuation in humans).

Even reanimated tissues will give us pause, as we look to confirm—consciously or unconsciously—the legitimacy of the *ex nihilo* creation of intelligence. There will never be any incontrovertible proof of the illegitimacy of our deeply felt solipsism, unless we develop a technology that allows us even if just temporarily to share the experience of the other (and perhaps not even then), but this turn in our thoughts is of such weight that it will likely not be accepted without something closer to certainty than the cyborg would allow.

The cyborg will not be meaningless but would, rather, be a special case in an ongoing deconstruction. As Graham wrote, “the erosion of clear boundaries between humans, machines and non-human nature can either be interpreted as a threat to the ‘ontological hygiene’ of humanity or a rendering transparent of the very constructed character of the parameters of human nature” (2002, p. 20). The cyborg is thus a legitimate source of potential deconstruction and would create its own kind of transparency, but it would not have the specific effect with which this study is occupied.

The biological construct is another variant on the proposed being that has been introduced in our literature and film, and it is here that the hardline exclusion of biologically based intelligence from consideration here may be called into question. The replicants of *Blade Runner* (Deeley & Scott, 1982)—based on the Nexus 6 androids of Philip K. Dick’s novel, *Do Androids Dream of Electric Sheep?* (1968)—can be thought
of as the extension of work being done on the growth of artificial organs. If one could
culture and grow each individual piece of a human in the lab, one could then assemble
them into a fully biological construct—what we might call the assembly line human. This
would be more likely to produce something like a new being than would Mary Shelley’s
method of stitching together pieces of cadavers and energizing the amalgamated result
with electricity. Genetic and other forms of manipulation could be used to alter the
characteristics of the being that is to be cultured and assembled. The replicants
themselves display the effects of such manipulation. They are stronger than ordinary
humans and are able to withstand extreme temperatures. But they also have foreshortened
life spans—a mere four years from construction to death. That the replicants are indeed
biological is evidenced by the necessity of the methods by which the bounty hunters (or,
in the film, blade runners) search them out. Were the chemistry of the flesh different, they
would be detectable through something other than a marrow test or a psychological test.

Such a being would not qualify entirely, as it would again be the manipulation of
already living cells. Cells that are cultured and directed to grow into a brain remain cells
that had a predisposition toward the evocation of intelligence. The suffering of so
articulate a being—one that had no prior existence, as does the cyborg—would
potentially bring the behaviors of the creator-species into question. This seems to have
been the point of the film. Based on this, the replicant would likely have a deep effect on
traditionally conceived humans. The replicant would technically not qualify as a subject
of this study, due to the origins of its cognitive functions. Although it could have a
profound effect, it would not touch on the ontological anxiety that may be at the heart of
(or may have been exposed by) the lacuna of meaning described by Jung.
Such reconstruction of the human can also be seen in the film *Futureworld* (Aubrey, Lazarus, & Heffron, 1976), a sequel to *Westworld* (Lazarus & Crichton, 1973), although in this case, the constructs are based on originals. The two main characters are recreated, one molecule at a time, resulting in doubles that are molecularly indistinguishable from the originals, except for the added neural pathways that direct them to act only in the company’s interest. Such molecular doppelgangers would fail to qualify for the simple reason that they would be indistinguishable from ordinary humans.

This possibility, however, brings up another caveat—that in addition to recognition of similarity, there must also be recognition of difference. To be confronted by oneself—molecularly recreated—would be inherently threatening and would lead to questions of identity that could bring a more damaging form of anxiety, possibly with no good resolution to be had. Likewise, the recreation of others based on one’s own memories of them—as is portrayed in both the original *Solaris* (Tarasov & Tarkovsky, 1972) and in its remake (Cameron, Landau, Sanchini, & Soderbergh, 2002)—does not evoke trust in whatever has recreated the individual, but rather suspicion. The main characters eventually accepted the other-doppelgangers (solidified memories), in both the *Solaris* films. The main character also eventually accepted her late husband’s double in *Starman* (Franco & Carpenter, 1984). In these portrayals, however, it is obvious that an alien intelligence is attempting to understand or to make contact via these other-doubles. It is possible that this deferred reference—this knowledge that one is in some way making contact with an alien intelligence through a solidified memory or a genetic duplicate—softens the effects such a creature would have on us. The literal, molecular recreation of the self, as was portrayed in *Futureworld*, would likely result in a stronger
and more persistent revulsion, as it would likely activate an anxiety similar to the fear of being replaced—the anxiety experienced by the father in an Oedipal dyad. This fear is laid bare in a primal way in the film *The Broken* (Bausager & Ellis, 2008), in which a family’s mirror images emerge from a shattered mirror, kill them, and replace them.

Our literature also presents the possibility of the pseudo-biological being. This is a being that appears biological and has many of the same structures that would be found in a truly biological being, but whose pseudo-biology is based on a different chemistry—that is, a chemistry analogous to our own, created by scientists. The play that gave us the term *robot*, *R.U.R.* (Capek, 1921/2004), portrayed beings composed of “a substance that behaved exactly like living matter although it was of a different chemical composition,” so that the robot was “man made from a different matter than we are” (p. 6). Such a being could be classified as entirely constructed, although its form on a cellular level would be much closer to the human form. As an entirely constructed entity, such a being would thus qualify for inclusion as a primary subject of this study.

The pseudo-cyborg is also present in the literature. It is portrayed in two forms. The first is exemplified by the T-800, as seen in the films, *The Terminator* (Hurd & Cameron, 1984) and *Terminator 2: Judgment Day* (Hurd, Kassar, & Cameron, 1991). The T-800 is a structurally humanoid machine whose cognition and motor functions arise entirely out of mechanical components, but which is covered by living tissue cultured from pre-existing cells.

The T-800 is not simply a machine with a particularly disturbing (or comforting, depending on one’s stance in relation to it) disguise. Its eyes, which were clearly functional, turning in wider arcs than its head while searching for Sarah Connor. When
the damaged eye was removed, the red pseudo-eye beneath was exposed and continued to function, although the biological eye itself must have been able to assist in focus of the inner eye toward one sphere of vision or another. Otherwise the saccade of the eyes would have been unnecessary.

Even with functionality of biological components, such an apparatus would be nothing more than a robot with accoutrements, and for this reason, it would be classified as a pseudo-cyborg. Were there biological parts necessary to its functioning, the T-800 might be considered a true cyborg and would be more likely to be excluded from this study. But with its entire, fleshy exterior burned away, the T-800 is still able to function with minimal impairment (its metal fingers are a bit less adept at pulling it across a smooth metal surface).

The character, Marcus Wright, from Terminator Salvation (Silver & McG, 2009) would, in fact, qualify as a cyborg and would thus be excluded from this study, where the T-800 would not (at least on grounds of the source of its cognition). Wright contains biological components that are necessary to his continued existence, as much as are the mechanical parts. Although a portion of his brain is mechanical, his heart—which he eventually donates to the character John Connor following an injury—is made up of the “living tissue” that allows the T-800 to travel backward in time in the original film, and he is unable to live without this tissue.

The criterion that the object-subject would need to be fully artifactual is not absolute. If the object-subject’s cognition is based on entirely artifactual components, the object-subject has met minimum standards of preliminary qualification, even if its outer covering (its artifice) is entirely biological. Stated another way, a biological component
that has no bearing on cognition would not be sufficient cause for disqualification.

However, while the T-800 would qualify based on its cognitive purity, it would be
disqualified based on its inability to voluntarily alter its mission. The T-800, as portrayed
in the two films, is an apparatus—an intelligent apparatus and one that has a limited form
of agency, but not one that is able to reflect on or alter its mission parameters. It is a
functioning apparatus, not a thinking (or pondering) being.

The androids Ash and Bishop, featured in *Alien* (Carroll, Giler, Hill, & Scott,
1979), *Aliens* (Carroll, Giler, Hill, Hurd, & Cameron, 1986), and *Alien 3* (Carroll, Giler,
Hill, & Fincher, 1992), exemplify the second category of the pseudo-cyborg. They are
indistinguishable from humans unless wounded, at which point their fluids (their
equivalent of blood) are shown to be white. That they are also partially mechanical and
not simply pseudo-biological is demonstrated in their ability to function as disembodied
heads when connected to a power source, as seen is the first and third films. Such
androids would likely qualify as subjects of the current study, depending on the limits of
their ability to engage in introspection.

This second category of pseudo-cyborg can also be seen in the film *Natural City*
(Jun & Min, 2003), which portrays beings whose pseudo-biological flesh (or pseudo-
flesh) has an expiration date,\(^{18}\) meaning that they will begin to lose their memories and
fall apart 3 years after production. However, the “AI chips” that contain their

---

\(^{18}\) The foreshortened lifespan has become a somewhat standard literary motif in the
portrayal of clones or constructed biological or pseudo-biological beings. The Sam Bell
characters of *Moon* (Fenegan & Jones, 2009), like the pseudo-cyborgs of *Natural City,*
have a 3-year lifespan following activation (although they are apparently able to be stored
indefinitely prior to activation). The Nexus-6 replicants of *Blade Runner* (Deeley &
Scott, 1982) have a 4-year lifespan, at times described as a fail-safe, to keep them from
developing enough emotional intelligence to become restless, and at other times
described as an inherent limitation of their engineered biology.
personalities can be removed and reinserted into a new “husk.” Such a merging of pseudo-biological and standard electronic components—apparent flesh-encasing hard technology—would, again, potentially result in a being that would qualify for inclusion.

Another possibility portrayed in film is what could be called one-to-one material transcription with eventual override. The film *Vexille* (Hamana & Sori, 2007) portrays the slow transformation of flesh into pseudo-flesh and then into metal by nanotechnology. Although the original consciousness is initially maintained by this one-to-one conversion, the nanotechnological virus that causes the transformation also inserts new pathways, which remain latent until activated all at once, so that at some point, the individual’s mind is overtaken, and he or she becomes subservient to the system of control. The novel *Dead Girls* (Calder, 1992) presents a similar scenario, the slow plasticization of the flesh through a virus produced via quantum manipulation, resulting in a being that is somehow both real and artificial. Both of these portray the corruption of something already present and intelligent—the colonization of living cells by something that is not alive (nanotechnology) or barely classifiable as alive (a virus). The beings portrayed would not, then, meet the requirements of inclusion in this study.

A similar proposition is the insertion of nanotechnology into DNA, what could be called the partial recombinance of DNA or the partial mechanization of DNA. There have also been several portrayals of the combination of a mechanical genome with human DNA, resulting in beings that are somehow both human and machine. *Tetsuo: The Bullet Man* (Kawahara, Tanishima, & Tsukamoto, 2009) features such a synthesis. The mechanical portions of the genome remain latent until a trauma causes their sudden expression, resulting in a being that has the appearance of a cyborg. Such a being both
does and does not possess cognition based in constructed components. That the DNA itself has been combined with a mechanical equivalent of DNA means that the mind of the being is partially determined by that which is constructed. The same kind of recombinance seems to take place in the film *Armitage III: Poly-Matrix* (Hasegawa et al., 1994), which portrays gynoids (constructs with a specifically female form) that have the ability to procreate with humans, the purpose being to compensate for falling fertility rates among humans. The third-type gynoid (or Third) is able to produce offspring without the introduction of female sex cells, and the result of this combination is seen in the film *Armitage III: Dual Matrix* (Kobayashi, Miura, & Akiyama, 2002)—a child that is physically indistinguishable from any other human but demonstrates extraordinary cognitive abilities. Whereas such offspring would achieve marginal qualification—marginal because of the heavy reliance of cognitive abilities on pre-existing (biological) typologies—this study would be more interested in the construct that bore such a hybrid.

In terms of the material out of which an object-subject is made, the fully constructed (or entirely mechanical) object-subject is the ideal subject and would meet the minimum standard of qualification. Beyond this, there are many permutations of such a being that could be considered. There have been several portrayals of stationary object-subjects—Alpha 60 of the film *Alphaville* (Michelen & Godard, 1965), VIKI of the film *I, Robot* (Davis & Poyas, 2004), and HAL of the film *2001: A Space Odyssey* (Kubrick, 1968). Looking strictly at composition, these three object-subjects would meet this first standard of qualification. The stationary apparatus, however, may be unable to function as well as other versions of the apparatus.
Before I started work on this project, it occurred to me that *2001: A Space Odyssey* portrayed not a breakdown in a computer but a breakdown in community. The astronauts failed to recognize the being-ness of HAL, and after the supposedly infallible artificial intelligence made a mistake—diagnosing a problem in a part that did not have any actual defect—the astronauts decided to shut HAL off, to effectively exclude HAL from the remainder of the mission—a rather harsh reaction given the inconsequence of the mistake.

But HAL did not have the hallmarks of being-ness that Cog or Wakamaru had, the motion, the ability to direct their gaze in such a way that those being gazed at know they are being gazed at. HAL was without pupils, without saccades. He was a stationary red eye that never moved and spoke in a distressingly sanitized monotone.

The temptation is to think of a stationary robot as talking furniture. Vox, the hologram confined to panes of glass in *The Time Machine* (Leobovit, MacDonald, Saralegui, & Wells, 2002) is such an apparatus—or at least the visible expression of such an apparatus. It becomes as much a location as a thing with which one might interact—at best like an elderly relative left in one place and occasionally visited. Locomotion may be unconsciously important.

Of course even stationary objects, if they have some quality with which one can identify, can begin to assume some measure of liveliness. When, in Jurassic Park (Kennedy, Molen, & Spielberg, 1993), the robotic arm that has been attending to the hatching velociraptors plucks the empty egg shell back out of the paleontologist’s hand, it demonstrates a high level of intelligence, in terms of ability—the ability to cope with variables, with an object being moved from one place to another, has been difficult to
achieve. But it also, in that exchange, gives the impression of personality. It is that exchange, specifically, that makes it seem like more than an automated system, although the action was, of course, automated. It was in relationship, and specifically in a humorous exchange, however brief, that the otherwise nondescript robotic arm seemed for a moment to be a subject. With the unnamed robotic arm, there is movement, which in many ways instills more of a feeling of kinship than even speech—HAL was somehow disadvantaged by comparison.

The great equalizer, when it comes to mobility, would be online communication, in which a stationary or even faceless apparatus or program could conceal itself behind either an avatar or a simple text-based dialogue. ELIZA, a therapeutic program, is an example of such. The interactions between ELIZA and her analysands took place via teletype. As Brooks (2002) pointed out, although ELIZA’s responses were based on a simple algorithm that mirrored back the patient’s statements, as would a Rogerian: “some people would spend hours pouring out their hearts to it” (pp. 166-167).

Of course the most ubiquitous image of the robot involves an object-subject that is ambulatory. If it is to be made in our image, as has long been assumed, it will be bipedal, as is Honda’s ASIMO. Most portrayals of object-subjects involve ambulatory constructs. A combination of the ambulatory and the stationary—what we might call a decoupled object-subject—may also be produced. This would entail a center of agency, which remains in a stationary location and operates the physical body through its own form of tele-operation (a computer with tele-presence). In the same way humans project themselves into a digital medium as avatars, in order to interact in that medium with other

---

19 A review of the prototypes developed prior to ASIMO will give an idea of how difficult bipedal locomotion is to replicate in the construct.
constructs, the stationary machine-intelligence might project itself into the social sphere through a physical avatar.\textsuperscript{20} The center would, in fact, not have to be stationary. One could imagine a cluster of robotic soldiers, all operating according to a central processing hub, a single center of agency housed in a mobile, tank-like vehicle, allowing for more immediate responses by the avatars to changing situations. Such a being—with a single intelligence and one or more avatars—would meet this basic standard of inclusion. In the opposite direction, a distributed system, with an intelligence that is housed and operates in a number of different locations—Skynet, for example, as depicted in \textit{The Terminator} (Hurd & Cameron, 1984)—would also meet this standard, with or without an avatar.\textsuperscript{21}

We must also consider the holographic apparatus. Holography would open up the possibility for some interesting variants. A forceful hologram would be an image coupled with some nonphysical force (a force-field) that is able to give the sensation of presence when touched. Such a hologram would be functionally solid due to the presence of these projected forces. A subtle hologram would be visible and intangible, functionally illusory—a computer-generated mirage. An agential hologram—whether forceful or subtle—would represent a slightly different form of computer tele-presence, but it would remain tele-presence, as the center of agency would remain in the computer projecting its avatar.

\textsuperscript{20} The presence of both human-operated avatars and computer-operated avatars in the public sphere would make for an interesting incarnation of the Turing test. Could the person interacting with the avatar tell whether it was operated by a machine or by a human?

\textsuperscript{21} As is made clear in the films, the individual terminators are able to operate at a point in history in which there is no Skynet to direct their actions. This indicates that these are free agents rather than avatars.
The idea of a holographic being was introduced in an episode of *Star Trek: The Next Generation* (Lane & Bowman, 1988) and led to a holographic main character—the Doctor—in *Star Trek: Voyager* (Berman & Braga, 1995-2001). Both figures relied on machines external to themselves for their continued existence, although the latter character was eventually able to walk about with the necessary hardware attached to his (holographic but forceful) arm and thus to leave the confines of the original limits of the projector. This raises the possibility of a hologram containing the machinery needed to generate its holographic form (the force-fields that compose its body) within itself—a walking forceful hologram with a mechanical heart. An episode of *Futurama* (Kaplan & Purdum, 2001) introduced the idea of a being with a mechanical body and a forcefully holographic exterior, which allows it to mask the mechanical nature of its body. This leads further to the possibility of a solid shell with a holographic interior. The apparatus’s internal structure would be made up entirely or in part by holographic channels (of the forceful, rather than the subtle, kind), which channel energies and keep the apparatus running—self-sustaining holographic circuitry. Were this holographic framework to

---

22 In this case, the intelligence directing the avatar seems to have been isolated from the remainder of the computing medium, which is to say, the computer generated a cluster of code that corresponded to the personality of the projection, and this code was aware enough of itself to isolate itself from the matrix from which it was created—thus when the computer program in which the character was placed was terminated, the character remained, as a cluster of code that refused to be deactivated.

23 The forceful image is projected by the stationary computers, the smaller device is then attached to the arm of the forceful projection, and the task of projection of both image and force is then transferred to the device on the arm, which then effectively holds itself up via the forceful projection.

24 My initial reading of the red Maria (or *maschinenmensch*) of *Metropolis* (Moroder, Pommer, & Lang, 1927) was that it was a mechanical being with a holographic exterior that took the form of the true Maria. Based on the visuals presented in the film, this is what one must assume to have taken place; some of the comments made in the audio commentary in the 2001 reconstruction release, however, indicate the author of the script meant for the red Maria to be covered with something like physical flesh.
disappear, the energies would dissipate into the hollow cavity of the body, and the agential and contemplative functions would cease—it would become, again, simply an object, as does the human when his or her viscera lose their coherence. All of these permutations of the holographic and the mechanical would meet the minimum standard of qualification, as the source of their cognitive abilities would be entirely constructed, rather than biological.

It is necessary also to consider apparatuses that have only apparent bodies, and which are thus not objects at all, but subjects with the appearance of objectivity. Such can be seen in two films released within 2 months of one another. The agents of The Matrix (Silver, Wachowski, & Wachowski, 1999) and nearly every character in The Thirteenth Floor (Emmerich, Emmerich, Weber, & Rusnak, 1999) have no body whatsoever, only a kind of subtle body projected into the minds of unconscious humans or into a program. Both films portray constructs inhabiting a wider construct. Only from within the larger construct—the world-construct—does the smaller construct appear to have a body. Such nonobject subjects (centers of agency held within a computer matrix) would, in fact, meet this minimum standard for inclusion due to their ultimate source in electronic components.

An interesting possibility, which in some ways defies categorization, is what might be called psyche-transcription, a method that would result in a psyche-

---

25 An apparatus such as this would come into being when another complex of machinery generated the holographic structures, then transferred control of these already extant structures to the holographic structures themselves.
doppelganger. Of course, the transfer of memories from a construct into an intangible medium would have no effect on whether such an apparatus (or the program that would be left of it, as a kind of digital ghost) would qualify. This backing-up of an android’s mind is somewhat standard in speculative fiction. Such takes place with the character of Julian Moore, the last of the Third series androids (and the only male in the line), who gains an ethereal online presence in Dual-Matrix (Hasegawa et al., 1994) after his body is destroyed by a virus in Poly-Matrix (Kobayashi, Miura, & Akiyama, 2002). The construct would no longer be an apparatus, but rather the internal remnant of an apparatus, rather like a soul entering a metaphysical realm. The opposite movement (from program to body) can be seen in the already cited series Caprica, in which the double (created from online records of the original and coached toward personhood by the original prior to her death) is downloaded into a Cylon body. Neither of these—the apparatus-toward-being or the system- or program-toward-being—would be disqualified from inclusion.

The transcription of the human mind is more of a sticky issue. Moravec (1985) described a method by which the brain would be slowly destroyed, a fraction of a millimeter at a time, and replaced by a program that precisely replicated what was destroyed, until the entire pattern of neuronal activity had been transferred into the new (electronic) medium (pp. 140-141). He also suggested several less invasive methods for transferring one’s patterns of thought into a new medium:

Suppose we sever your [corpus] callosum but also connect a cable from an external computer to both severed ends. If we understand the human brain well

---

26 Or a mind-doppelganger, depending on the extent of the transcription and whether it codifies deeper structures of the psyche or only basic structures. As far as I can tell, I am the first to use these terms to describe a particular type of object-subject.
enough, this external computer can be programmed not only to pass information, but also to monitor the traffic between the two halves. Like the personal mimic, it can teach itself to think like them. After a while it could insert its own messages into the stream, becoming an integral part of your thought processes. In time, as your original brain fades away from natural causes, it can smoothly take over the lost functions. Ultimately your mind finds itself in the computer. (p. 143)

Several years later, Moravec (1989) wrote of another method:

You carry this computer with you through the prime of your life; it diligently listens and watches; perhaps it monitors your brain and learns to anticipate your every move and response. Soon it can fool your friends on the phone with its convincing imitation of you. When you die, this program is installed in a mechanical body that then smoothly and seamlessly takes over your life and responsibilities. (pp. 110-111)

However it is accomplished, such a process, and the product that results—a kind of psychical facsimile—would make for some very interesting possibilities, as Moravec (1985) pointed out:

The program in your machine can be read out and altered, letting you conveniently examine, modify, improve, and extend yourself. The entire program may be copied into similar machines, giving two or more thinking, feeling versions of you. You may choose to move your mind from one computer to another more technically advanced, or more suited to a new environment. The program can also be copied to some future equivalent of magnetic tape. If the machine you inhabit is fatally clobbered, the tape can be read into a blank computer, resulting in another you, minus the experiences since the copy. With enough copies, permanent death would be very unlikely. (p. 143)

Reconstruction of a psyche based on electronic records is another possibility presented in the series Caprica (George, 2009-2010), in which a young woman recreates herself her medical records, shopping receipts, and grades.

---

27 I find body-modification distasteful, and for this reason I find the suggestion of basic identity-modification rather distasteful.
Moravec (1989)\textsuperscript{28} suggested another interesting possibility—that of the conglomeration of doppelgangers:

I have already mentioned the possibility of making copies of oneself, with each copy undergoing its own adventures. It should be possible to merge memories from disparate copies into a single one. To avoid confusion, memories of events would indicate in which body they happened, just as our memories today often have a context that establishes a time and place for the remembered event. Merging should be possible not only between two versions of the same individual but also between different persons. Selective mergings, involving some of another person's memories and not others, would be a superior form of communication, in which recollections, skills, attitudes, and personalities can be rapidly and effectively shared. (p. 115)

Something akin to this was portrayed in the character of SID 6.7 in the film \textit{Virtuosity} (Lucchesi & Leonard, 1995), this character being a program compiled from the transcripted minds of a large number of individuals—in this case serial killers.

There have been several portrayals of embodied transcriptions. The aforementioned film \textit{Futureworld} (Aubrey, Lazarus, & Heffron, 1976) portrays molecularly identical duplicates, whose minds appear also to be replicated in their entirety.\textsuperscript{29} The film \textit{Moon} (Fenegan & Jones, 2009), portrays the Sam Bell characters\textsuperscript{30} as embodied mind-doppelgangers, with these characters sharing not only mental architecture (personal psychology) but also all of the memories of the original up to a certain point in time. The replicated humans in \textit{Oblivion} (Berger, Morrison, & Kosinski, 2013) are

\textsuperscript{28} Moravec mentions this possibility in the chapter he contributed to Minsky’s book, \textit{Robotics} (1985a), but it is in Moravec’s 1989 reissue of that chapter in his own book, \textit{Mind Children} (1989), that the thought is more aptly expressed.

\textsuperscript{29} Unlike the transporter of Star Trek, which is based on a method so invasive it disintegrates the original [as Moravec (1989) pointed out, to an individual holding to the body-identity perspective, the transporter is “an elaborate execution device” (p. 118)], the method portrayed in this film preserves the original, resulting in a clearly recognizable imposter—clearly recognizable by the original, at least.

\textsuperscript{30} The Sam Bell characters, pseudo-biological or biological doppelgangers, are portrayed against the image of the traditional robot, GERTY, whose lifespan is significantly longer, but who is unable to manipulate the mining machinery as a humanoid can.
similarly not clones, but fully formed replicas with enough of the originals’ cognitive processes replicated that they can function and even, in their dreams, recall the memories of their predecessors, of whom they are copies. The same idea finds expression in the film Blade Runner (Deeley & Scott, 1982), in which an android named Rachel is able to progress much further into an empathy test than other androids without being detected as an android. This is due to the memories that have been implanted in her—memories she believes to be her own, but which actually belong to her employer’s niece, a kind of preemptive confabulation that has the effect of bestowing emotional intelligence sorely lacking in the other replicants, such that they are unable to experience compassion—until the last moment of its life, in the case of Roy Batty.\footnote{The idea behind these portrayals seems to be that there is some mysterious quality, which gives rise to our personhood, and which cannot be easily replicated in a digital form. The roboticist Rodney Brooks dealt with this attempt to identify some special substance or, “new stuff” (2002, p. 176), which could be used to maintain the kind of preemptive hard-line solipsism Brooks identifies in Penrose, Searles, and Chalmers (pp. 176-180). Brooks himself postulated the existence of “some fundamental mathematical description of what is going on in living systems,” without which the “necessary generative components,” which give rise to human-like traits, could not be replicated.} Rachel is the shadow of a previous mind—an echo of a mind-doppelganger.

The transcription of the psyche into a mechanically embodied medium has been portrayed in numerous films as well—the marionette-like object-subject known as Mr. X in the Japanese pink film Maid Droid (Tomomatsu, 2008) is a shell into which a man downloaded his consciousness prior to dying, a process that allows him to continue running an android factory following his death.\footnote{The last vignette in the film Robot Stories (Chien, Ima, & Pak, 2003) portrays the refusal of a man to submit to such a process, insisting that he die as an embodied being. Before he dies, he speaks to the transcription of his late wife, who encourages him to allow a similar transcription.} In the film Ghost in the Shell 2: Innocence (Ishikawa, Suzuki, & Oshii, 2004), the mind architecture of sex robots is
created from the minds of kidnapped children, in a process called ghost-dubbing—illegal because it destroys the original. This theme is echoed in the film *Parasite Dolls* (Fukui, Yoshinaga, & Nakazawa, 2003), in which, in order to generate emotional intelligence and produce an emotionally reactive sex-android, mind architecture is copied from a human, a process that results in an android capable of both dreams and hallucinations and susceptible to the influence of violent complexes and even psychosis.

My first response to this possibility was to assume exclusion of such psyche-doppelgangers. This would not be the creation of consciousness so much as its duplication—akin more to the creation of a facsimile or carbon copy than to a truly creative act of recombination and analogous replication. If the act of constructing an apparatus with the equivalent of intelligence is akin to sculpting the figure of a horse, transcripting the original into a new medium is akin to taking a plaster cast of an actual horse and representing it as art. This would not then represent a replication of our consciousness through an act of self-awareness and self-reconstruction, but would rather be the production of a facsimile—and this does not fall into the category of creation *ex nihilo*.

The film that softened my stance toward the psyche-doppelganger was *Puzzlehead* (Bai, 2005), in which the architecture of the mechanical being was based on the mind of its creator but was not entirely identical to it. The being that resulted was, in many ways, a childish (or somewhat naïve) version of its creator, which was made to confront the childish behavior of its creator. The juxtaposition of the somewhat simple

---

This very underrated film ends with the android replacing the “diseased” mental architecture it inherited from its maker with the uncorrupted architecture of the maker’s newborn child (in a scanning process that was not destructive to the original, as in other
doppelganger and the selfish creator in this narrative gave an insight into the ways in which even a “carbon-copy,” suitably different from the original, might critique the original and lead to the original confronting his own flaws and, thus, growing.

The object-subject that results not from a total transfer (the destruction of the original, as described by Moravec) but from a process more akin to a carbon copying would be able to experience the behavior of the original from outside. This would certainly offer a new perspective on oneself. One may also take into account the fact that the apparatus will likely be mass-produced, so that it will then become a facsimile of itself, one of many identical apparatuses. It is the subjectivity—the agency—of the facsimile that makes it more than a simple carbon copy, because as soon as the object-subject is activated, it will begin to experience the world and to act in the world as a subject, and these experiences and acts will become part of its individuality—it will begin growing away from its status as facsimile, toward something like individuation. Although the full meaning of the term *individuation* will likely be quite beyond it initially, its progress, however small, toward status as an individual should not be negated. I would thus include the psyche-doppelganger in the pool of intended object-subjects.

The reverse process—the overwriting of an intelligent, fully constructed mental architecture onto an existing biological architecture—as is seen in the character of Bane in *The Matrix: Reloaded* (Silver, Wachowski, & Wachowski, 2003a) and *The Matrix: portrayals). This replication of the self is an exercise in narcissism that nevertheless results in the same Oedipal suspicion on behalf of the father figure—the creator—that other doppelgangers evoke. In the film, the creator controls his android double by turning him off via remote control. When the android saves a woman from an attacker and earns her trust, the creator enters into a relationship with her in the android’s place. When the android kills his maker and renews his own mind by overwriting parts of it with the mind of the woman’s newborn child, he essentially makes her his mother, completing the Oedipal theme.
*Revolutions* (Silver, Wachowski, & Wachowski, 2003b) or in *The Thirteenth Floor* (Emmerich, Emmerich, Weber, & Rusnak, 1999)—would be an extension of a being that would already qualify for inclusion in this study and would be considered a special case. In some ways, this being would be more a subject of this study than the being whose mind is a facsimile of a biological mind or psyche.

Moravec (1999) offered another strange possibility related to the codification of mind into a new, more easily manipulable medium. He suggested that a mind (or memory) conglomerate would not need to be limited even to members of one’s own species. It could in fact incorporate the minds of dolphins, elephants, whales, giant squid, or other animals (pp. 115-116). Another possibility is that of a complete merging of biologically based consciousness and a center of agency that is entirely constructed. Although it makes for interesting narratives, this would not be a simple task. It would involve strengthening certain neural pathways, corresponding to the thought patterns of the constructed mind, while simultaneously weakening or overriding the pathways typically used by the original, biological mind. The Bane-scenario—the overwriting of a completely novel personality onto an existing personality—is rather doubtful.

The scenario portrayed in *The Thirteenth Floor* is somewhat more plausible, as it represents the overwriting of a mind-doppelganger onto the original (biological) mind, meaning that there are fewer neural pathways in the brain of the original to be altered in order to achieve the conversion. This is, in fact, something like what Moravec describes in terms of the conglomeration of memories, with the memories unique to the doppelganger privileged (strengthened) over the memories unique to the original (which are weakened or otherwise suppressed). Where the mind-doppelgangers of *Moon* share
the memories of the original up to a certain point, the characters in \textit{The Thirteenth Floor} are stripped down mind-doppelgangers, in that the structure of the mind is replicated without the memories, and the pseudo-doppelganger is inserted into a simulated universe to form its own memories. The film’s narrative rests on the rather naïve premise that for the original to enter the simulated universe, he and the pseudo-doppelganger would have to trade minds, so that the original enters the simulated universe in the simulated body of its doppelganger, and the original’s physical body “holds the consciousness of the program link unit,” or the mind of the pseudo-doppelganger, as an ill-fated character says, which remains unconscious throughout the process. This is, of course, bad science. There is no reason to embed a segment of code in a biological system (the brain) while it is deactivated, when this could easily be done within the program itself.

The original film, \textit{Ghost in the Shell} (Ishikawa, Iyadomi, Matsumoto, Mizuo, Watanabe, & Oshii, 1995), ends with the merging of a computer program with the downloaded consciousness of a detective, suggesting another interesting variation in the conglomeration proposal. Such an object-subject, as the combination of a construct and a transcription, would qualify for inclusion, and the conclusions generated by this dissertation would be considered applicable.

In considering which beings would qualify as part of this study, it does not matter whether the being itself is constructed by humans or represents a second or third generation construction—or, in other words, whether a human makes it, or whether a constructed being makes it, secondary to its own construction or creation. Presumably the imprint of humanity will remain in a second- or third-generation apparatus, even if humanity no longer has any direct influence on its construction—the human remains the
fons et origo, despite the intervening generation(s). And the ability of subsequent
generations to understand and critique humanity would not necessarily be compromised
by this separation—this distance or (on some level) objectivity. The creation of a
constructed being by a constructed being has been portrayed in several episodes of *Star
Trek: The Next Generation*, one in which an intelligent holographic being creates another
intelligent holographic being as his companion (Echevarria & Singer, 1993), and one in
which an android creates another android as his daughter (Echevarria & Frakes, 1990).

It also would not matter whether the code or structure upon which the mental
architecture is based were constructed by a human with the intricate intentionality of
normal electronic creation, or whether another process was substituted for direct
intervention. Minsky (1985b) wrote of a process identified by Alan Turing in the 1930s:

> He showed that, given enough time and memory, even the simplest computer can
be programmed to do many things that its programmer never has to conceive of at
all. All we need to do is program that computer to mindlessly write down all other
possible computer programs, one by one, and then to try them out! We call this
the process of “generate and test,” and it is one of the simplest and earliest
formulations of artificial intelligence. In principle, such a scheme will eventually
find the answer to any solvable problem. (p. 295)

This monkey-with-a-typewriter approach could potentially generate intelligent, agential
programs that could be considered subjects of this study, whether embodied or not,
despite the lack of direct intentionality by a human. The same can be said of the method
described by de Garis (2005): “A genetic algorithm (GA) uses a software-simulated form
of Darwinian evolution to optimize the performance of whatever is being evolved” (p.
32). It involves the insertion of random “mutations” into the computer code. If the
mutation results in a program that is better able to accomplish an identified task than its
predecessor, it is maintained, and random mutations are then inserted into it. If not, the original code is maintained and the mutated code is discarded.

Evolvable hardware (EH) is similar to genetic algorithms, and self-reconfigurable (SR) robots represent an adaptive form of evolvable hardware. An SR-object-subject would be a conglomerate, composed of identical robotic cells. The holders of U.S. patent No. 4,608,525 (1984) describe such a robot: “This invention relates to a cellular type robot apparatus consisting of a plurality of robot cells each having intelligence, wherein each robot cell controls its own operation on the basis of information exchange with adjacent robot cells,” so that “the operations of the robot cells are as a whole coordinated,” allowing them to “operate cooperatively as a group to perform a manipulative action.” White, Kopanski, and Lipson (2004) wrote of the adaptability of such a conglomerate:

Self-reconfigurable robots can change their morphology autonomously to accomplish diverse tasks. These robots can rearrange the connectivity of their structural modules to create new topologies, and thus can attain a larger variety of configurations than are available to a fixed-morphology robot. (p. 2888)

SR robots are, then, structurally amorphous. An SR-object-subject with sufficiently intricate individual cells would potentially be able to mimic the human form enough to evoke feelings of self-recognition and empathy—close enough to be indistinguishable from humans, thus avoiding the uncanny valley described by Mori (1970), which results in feelings of disgust. This possibility was portrayed in the television series Stargate: SG-1 (Wright, 1997-2007) and its follow-up series Stargate: Atlantis (Wright & Cooper, 2004-2009).

The SR-object-subject would not have to take the human form in order to qualify, however, as holding a humanoid form is not a prerequisite for inclusion in this study. The
variety of beings that is likely to result from our efforts may often include forms that are atypical for intelligent beings. An object-subject that is nonhumanoid in terms of body structure will not necessarily be nonhumanoid in terms of its agency or intelligence. In *Ghost in the Shell: Stand Alone Complex* (Watanabi, 2002-2005), the Tachikoma, spider-like object-subjects, display humanoid thought and speech patterns and use their forearms gesturally in ways that would predispose humans to view them as similar to themselves.34 Some degree of affinity—of a recognition of oneself in the other—is necessary if the object-subject is to have an effect on the human, and the degree to which it resembles the human in some important way (speech or gesture, for instance) will have a great effect on its efficiency in this regard, although it need not resemble us physically.

There is a way in which the quality referred to as humanoid will be necessary for inclusion, and this is in the matter of size. The fact is that we are not only attempting to recreate ourselves but are attempting to recreate all of nature—biota, trees, insects, and mammals. A bacterium that has been genetically manipulated in order to build tiny structures, or otherwise behave according to human will, has already been excluded by the identification of genetic recombinance as a separate practice. But even an entirely constructed bacterium—that is, a construct that behaves like a bacterium and is of the

34 These represent one of three forms of “think tanks,” artificially intelligent and ambulatory weapons platforms portrayed as having childlike thought processes; the others are the Fuchikoma and the Uchikoma. The Tachikoma share their memories of engagements with one another so that they can be analyzed and understood from a variety of geographical perspectives, although each of the Tachikoma begin to develop distinct personalities over time. The name, Tachikoma, combines the name of the protective lion-dogs of Japanese myth (*koma inu*) and the *tachi*, or long sword. Its predecessor, the Fuchikoma, combines the *fuchi*, or hilt collar, of a Japanese sword with the mythical creatures.
same relative size—would likely be unable to achieve something equivalent to human consciousness.

Form, then, does not matter, but the universality of cognition does. The drones—airborne weapons platforms—that exist alongside the recreated astronauts in Oblivion (Berger, Morrison, & Kosinski, 2013) would not be considered, as they appear to be motivated or driven by nothing more than simple search-and-destroy programs (simple is perhaps not the correct term, due to the enormity of creating such a device). Only if it could be redirected into something like reverie or philosophical pondering would the drone qualify, having begun to take on the essential qualities of being, over and against the qualities of apparatus. This is the third qualification—that the intellect of the object-subject be universal, or adaptable, that it be able to redirect into novel channels, that it not be so entirely dedicated to a certain task or class of tasks that it cannot engage in any other.

Expert systems, programs that have associated particular inputs with particular responses and can thus mimic the behavior of an expert in some area—for instance, diagnosing a mental illness—could easily be mistaken for potential subjects. However, as Minsky (1985a) wrote:

The trouble with them is that they all are too narrowly specialized; they work only within the contexts or environments they are designed for. When you try to use them for anything else, then they show few signs of having any ordinary common sense. (p. 15)

As useful as an expert system can be in many situations, it remains a digital apparatus, lacking the adaptability that in many ways defines human cognition.

The addition of agency to an apparatus makes it an object-subject. The addition of intelligence allows the agential machine to evaluate—in essence to compare and contrast
given movements—rather than reacting immediately and in a determined manner (as did Walter’s tortoises). The addition of universality or adaptability of cognition opens the possibility of individuation to the intelligent, agential machine, meaning that not only the cognitive abilities but also the most important of human activities would have been re-created. This adaptability would lead to unexpected or novel permutations of cognitive states, which would generate something like genuine emotional reactivity.

Mazis (2008) wrote of sentiment as a mere semblance of genuine emotion: "What substitutes for deeper emotional communication is sentiment. The emotional content of sentiment draws upon the felt quality of relation through generalized representations” (p. 33). Sentiment is the cliché of emotion, a programmed stimulus-response mechanism unmediated by any real internality. It is, rather, “a predictable acting out of general, agreed upon roles” (p. 33). It would be very difficult to distinguish between authentic and mimicked emotion. The object-subjects portrayed in Daft Punk’s film Electroma (Hahn, Bangalter, & Homem-Christo, 2006) display such behavior, in that the disappointment felt over a failed attempt to humanize itself leads one object-subject to persuade another to help it suicide, after which the second object-subject suicides through self-immolation. The object-subject that achieved this kind of unexpected behavior would be considered an apparatus-toward-being and would qualify for inclusion here.

Whether an object-subject is ambulatory or stationary, whether it is composed of numerous potentially autonomous cells or is itself a singularly discrete unit, whether the engine generating its cognition is held within its socially-revealed frame or directs this frame from a separate location, if an object-subject is in possession of something that is psyche-like—or technologically psychoidal—and if this agency and intelligence can be
directed toward a task other than a rote or specific task, toward channels of a more philosophical or ontological nature, and if this psychoidal quality is emergent from an entirely constructed medium, it shall be considered the subject of this study. This would mean that we would be dealing with a being capable of self-reflection, one that has progressed to such a degree that what it possesses might be considered equivalent to human consciousness (or meta-cognition).

Given what the apparatus faces, if it is to function as Job functioned toward his culture’s god-image, these distinctions are more important than one might realize. We have generated imaginal variants of the proposed apparatus that would stimulate a deep anxiety, and we have generated others that would have no such effect. It is the former that will potentially act as pivots upon which human consciousness may shift away from its infancy and toward a further maturation.
Chapter 6
Transparency and Tain

The term *android* is aptly applied to the apparatus. Whether it resembles humanity in physical form, it will be a reiteration of human abilities (both cognitive and physical), and perhaps also human tendencies (more subtle qualities—those not related in any direct way to functionality). It will be a reflection of us. The question at hand is how deep this reflection will be.

If it is an artificial reflection—one of surface abilities, born of an apparatus worthy of the descriptor *artificial intelligence* and nothing more—its presence may, in the end, have only a superficial effect on the human psyche in general. If, on the other hand, it shows itself to be reflective not only of our surfaces but also of our depths, its effect may be quite far-reaching—and in fact might reach further inward than we would at first assume. In my thinking, a reflection of depth would evoke an already latent ontological anxiety, doubt as to the reality of one’s own being-ness or sense of self, that it would cause this anxiety to become more explicit or present. Such would cause a crisis of consciousness—and in fact we may be in the early stages of this crisis. Finding our depths reflected in a conglomeration of materials and components that, in traditional modes of thought, were never meant to be alive may lead to a new kind of vulnerability—what we might call transparency.

There are three well-known metaphors, linked by a common element (a box with unknown contents), that can together illustrate this idea of the absence of one’s being-ness. One is the idea offered by B. F. Skinner that the human psyche is best thought of as a black box and that the proper study of humans should be a study of stimulus and response. The second is the metaphor of Schrodinger’s cat, which is meant to illustrate
one interpretation of quantum mechanics. The third, and most ancient, is that of
Pandora’s box, which contains all evils and, in one interpretation of the myth, should not
have been opened.

I take Skinner’s black box as a statement of determinism—or an expression of the
human as automaton—and take both this and Schrodinger’s cat to be expressions of the
go’s exile from the remainder of the psyche—from all that could be encompassed in the
term *unconscious*. The reason the idea of the box is appropriate here is that to the ego, the
unconscious portions of the psyche are a mass of darkness punctuated by forms difficult
to understand (dream images, compulsions, et cetera). The ego seems instinctively to
know that the being-ness it experiences is a relative being-ness, the being-ness of a
portion of the psyche, and that there is a greater being-ness to which the felt being-ness of
the ego could be linked—if not for the exile, which necessitates dream interpretation and
other forms of psychopompery.

I also take both as expressions of humanity’s uncertainty as to the reality or
virtuality of the human soul. This if fairly explicit in Skinner’s metaphor, but
Schrodinger’s was meant to illustrate something else entirely. Only if we think of the
cat—which is somehow neither alive nor dead, until the box is opened—as the image of
the human soul does the connection become explicit. When one takes the third box,
Pandora’s box, into consideration, this ancient myth can be interpreted as an expression
of anxiety over damning knowledge, or knowledge that brings evil—and the potential
annihilation of the soul via its disconfirmation would appear, at least to the ego in its
current state, to be evil.
Throughout what follows, I will also evoke the idea of the divine creator and its traditional use (in my interpretation of history) as a proof of our conception of the soul. The evacuation of meaningful content from the idea of the divine—its failure to continue acting as a structuring element, a failure expressed in Nietzsche’s famous phrase, God is dead—seems also to cast doubt upon the reality of the soul itself, which had always been reified through belief in this divine. The death of the god-image as a narrative that gives meaning and direction may be a foretaste of an evacuation of the human soul itself—a loss of our ability to believe in the authenticity of our own being-ness.

This is the ontological anxiety, or self-solipsism, that may be exacerbated by the presence of a deep reflection. But the idea of the soul itself—and perhaps even more highly developed concepts of the psyche, as are present in psychologies such as Jung’s—may be shown to be something that must be cast off, in order that something else might be born (not a new soul, perhaps, but a new conception of soul). Whether this crisis will come or not will depend in large measure on the depth of the reflection with which we are presented, as we gaze upon the apparatus. If we find ourselves faced with an apparatus-toward-being (one that shows propensity toward developing something both recognizable and deeply resonant) or even an apparatus with being-ness (with qualities that resonate so deeply within us that we are unable to distinguish qualitatively between its being-ness and our own), a reinterpretation of our own being-ness will become necessary.

**Icon of Artifice, Icon of Depth**

The re-creation of human attributes or abilities in the proposed apparatus will be a kind of self-iconography, a generalized replication or reiteration of the self. It will be a
projection of human abilities onto new forms, and the result will be an apparatus that is, in one way or another (and likely in several), made in our image. It will be an object-subject that acts as the manifest content, pointing always beyond itself toward a latent referent—humanity itself.

The designations *manifest* and *latent* are taken from the method of dream analysis developed by Freud, who identified the dream image as a disguised image, which points beyond itself to some hidden, latent content. Freud (1911/1952) wrote: “I am led to regard the dream as a sort of *substitute* for the thought processes, full of meaning and emotion, at which I arrived after the completion of the analysis” (p. 15). As to his method, he wrote:

> In order to contrast the dream as it is retained in my memory with the relevant material discovered by analyzing it, I shall speak of the former as the “*manifest* content of the dream” and the latter... as the “*latent* content of the dream.” (p. 16)

The manifest content of the dream makes present (in metaphorical or imaginal representation) the latent contents that were the true source of whatever symbolic significance the manifest content had. The apparatus, as an android—fashioned after us in some way, whether in form or in ability—is also, by virtue of this fact, a substitution, a manifest referent that will appear to us as a solid thing (whether a hologram, having only virtual solidity, or as a more traditional apparatus). Beneath and beyond this manifest referent will be the latent referent—our own being-ness.

The apparatus will be a reflection. And a question occurs as to the artificiality or depth of that reflection: Will the self-icon make re-present only the surface of the human or will it make re-present the human psyche in something closer to its fullness? How deep will the reflection run?
A mimesis of depth will carry different implications to a shallow one. The question is essentially whether the icon will be sign or symbol—whether it will be a hollow or static mimesis or rather will be a broader or more comprehensive mimesis, one that carries with it a promise (and threat) of further ramification. If the replication remains at the level of artifice, if only relatively surface-level abilities are replicated, the apparatus will be a sign—there will be only a virtual connection between the manifest referent (the apparatus) and the original or latent referent (the human), a connection via hollow signifiers, which will carry implications only in one direction (from the human as model to the apparatus as copy). If, on the other hand, the replication extends downward, as it were, if the reflection reaches deep into the human psyche, such that we will be presented with a fuller reflection of ourselves, the apparatus will function as a symbol. There will be an analogous connection, a vital connection, which not only links but also binds the depths of both the original (the human) and the reiteration (the apparatus), such that the movement of one of these depths would potentially have an analogous effect in the other.

Put another way, the question of the depth of the icon can be portrayed as a successful or unsuccessful attempt at what could be called re-presencing, along the lines presented by Ankersmit (2003):

According to the substitution theory of representation... a representation is a substitute or a replacement for what it represents. Recall the etymology of the word “representation”: a “re-presentation” makes “present” what is “absent” and can succeed in doing this by taking the represented’s place, hence by being its substitute or replacement. (p. 320).

35 My assumption, here, is that the replication of the human in its fullness is what would represent success, although many would take a much more pragmatic standard.
Will the apparatus be a successful or only cursory or shallow re-presencing of the original referent? If a hollow substitution, then it is a sign only, a signifier that ends with itself and establishes no metaphorical equivalence. If the presence of the depths of humanity is felt, however, in the presence of this work, then it will be a symbol, and what is true of it will be symbolically true of the original as well.

The latter would be the establishment of a deep equivalence between the image of humanity (imaged humanity, the apparatus-toward-being) and humanity itself. The idea of re-presencing is an important one here. One kind of apparatus will cause humans to feel the presence of their own being-ness in the other, where another will give only a cursory sense of the presence of oneself, something closer to our interactions with ATM’s than to our interactions with other people. So the question is whether artifice or depth will be re-presenced. When we peer at the apparatus, will our gaze meet our image only at its surface (form, perhaps, and more likely function, the ability to perform tasks that require intelligence—what, following Lyotard (1979/1984), we might call intelligent operativity), or will the recognition of our own image penetrate into the depths of the simulacra?

If the former is true, it will be a truly artificial self-icon, reflecting only the surface of the latent referent. It will be nothing more than a sign, a collection of signifiers that convey only a partial, or truncated, re-presencing of the original referent—a

---

36 This can be thought of as a felt form of the Turing test, in which a person is either able or unable to distinguish between a machine and a human during a conversation via teletype, with the inability to distinguish between the two taken as evidence that the test has been passed and that AI has been achieved. There is as yet no version of the Turing test for the soul or being-ness, a test of the fullness of the mechanical (replicated) psyche—a test to determine whether the signifiers that indicate the presence of such a quality are sentiment whether they truly indicate the presence of such a gestalt.
reflection of its most superficial or purely operative qualities. However, if it is to be an icon of depth, if a gestalt occurs that links these signifiers in a way very much reminiscent of the gestaled linking that takes place within the human, so that the apparatus exhibits authentic signs of internality—or what the roboticist, Hans Moravec (1999), called “manifestations of internal life” (p. 111)—the representation will have extended inward and will have even made steps toward producing a more or less true recreation of our inner experience. Such an icon would be not a sign but a symbol, not a dead simulacrum but a simulacrum that opens up channels toward and into the latent referent (the deeper human psyche). The sign would be a shadow. The symbol would be an enlivened embodiment.

The difference between the apparatus as sign and as symbol can be made more concrete via the distinction made by Glen Mazis (2008) between sentiment, a hollow (false) signifier of emotion, and authentic emotional interaction. He wrote of sentiment that “it is a predictable acting out of general, agreed upon roles” (p. 33). It is rote and shallow, a contact between surfaces, a deceptive kind of play, an acting-out of prescribed dialogical form:

Since sentiment only requires one to play a general role and not to really encounter the specific problems and quirks of the other, there is not that immediate being pulled into the depths of the other person leading into unknown feelings and responses. (p. 33)

Sentiment, he wrote, is a hollow gesture, a poor substitute for “the intimacy of real emotional encounter” (p. 34). It rather binds the participants to social expectation. The question at hand, then, is whether the human will feel some deep camaraderie with the apparatus, whether there will be recognition of the self in the other-ness of the other, or
whether the human will feel him- or herself interacting with a hollow that possesses only the artifice of form.

The film *Condor* (Nelson, Orgolini, & Vogel, 1986) offers a compelling image of artifice in the form of the hostess of a fast food stand. As the protagonist drives away, tacos in hand, the hostess is seen powering down, a process that includes the entire body, which had been raised up to the level of the counter via pistons that take the place of a lower half, being lowered again, to await the next drive-up customer. This is a portrayal of an apparatus kept in a reduced state—an apparatus governed by operativity. Although the hostess interacts as congenially as any human hostess (and, in many cases, more so), she/it would likely not be taken to have more than sentiment, the artifice of relational interaction.

The possibility of equivalence and of deep equivalent influence (reciprocal causal action, with changes in one resulting also in changes in the other) is both the promise and the danger of constructing the icon of depth. A fuller re-presentation would make possible a kind of heavy equivalence, which, as I define it, indicates that the equivalence connecting two referents is such that qualities found to be part of one can be seen also as being part of the other and changes in the state of one can effect changes in the other. This is a kind of metaphorical entanglement—a deep form of empathy—and the human instinctively fears being entangled with the fully constructed artifact, for reasons that will be unpacked below.

It is the possibility of heavy equivalence that makes the possible recreation of the human’s being-ness inherently threatening. One might think that such an achievement would be felt as a demonstration of the technological prowess of the species, something
that would lead to ego-inflation—and this would be true. But such a feat would also be profoundly disappointing. We are accustomed to thinking of machines as things that can be parsed—works whose parts can be disassembled and reassembled. If all the signifiers of depth can be evoked from such a parsable machine, one of the implications would be that roboticists had developed the ability to parse out our essence or essential being-ness to such an extent that it could be replicated. Such replicability would indicate that human being-ness had been summarily deconstructed—reverse engineered and shown to be nothing more than the sum of its parts. If all the signifiers are present, which we have used to prove to ourselves that we possess some soulish essence, what compelling denial can we offer that our being-ness has been at least apparently re-created in the apparatus or artifact and that our specialness, in light of this recreation, has been nullified?

This being-ness—this most intimate of experiences, the experience of the self as such—as the internal principle of organization or the gestalt that both produces and is produced, faces not only its own death in the object-subject in question but its own retroactive annihilation, its unmasking, its exposure as the emperor with no clothes—its absolute nullification forward and backward in time. The fear is that we will find our sense of ourselves to have been (and to be) a self-sustaining illusion. A kind of nihilism would then pervade our consciousness, which would render previously meaningful actions meaningless. This would be akin to a clinical depression forming in the Geist. This nihilism would cast its shadow over everything that has come before and everything that is likely to come after. My experience of moments of such nihilism is that they are, in that moment, eternal—which is to say that the meaninglessness that is felt is so deep that it stretches backward and forward in time, turning all of one’s memories, no matter
how cherished, into ashy shadows, and precluding any meaningful events in one’s future.

I have experienced the eternal despair that can exist in what to those outside seems a finite moment. When I speak of the nihilism that may come when we are forced by our own ingenuity to share our being-ness with that from which we have always separated ourselves (in an attempt to reify our being-ness), it is this eternal despair—which somehow remains eternal, even when the finite moment has passed—that I reference. It is this disappointment, at the moment of our self-congratulation, that may gray our souls, one hopes in preparation for a renewal.

**Subversion-Quality of the Apparatus-Toward-Being**

The apparatus-toward-being is, by nature, categorically subversive toward our traditional systems of thought. As Moravec (1999) wrote, “the idea of machinery with a conscious mental inner life” seems to violate the “obvious and sacred dichotomy of the living and the dead” (p. 111):

> In the old metaphor, we are in the process of inspiriting the dead matter around us. It will soon be our honor to welcome some it to the land of the living, however upsetting that may be to our traditional categories. (p. 111)

To write that such a violation would be upsetting is perhaps an underestimation of threat. The distinction between the living and the dead is fundamental to the mind—that portion of the psyche that can be described as an artifact, as artifactual. Structurally it is a deep (or bedrock) dichotomy, and to challenge it is to produce anxiety not only over the legitimacy of the entire system of discriminations upon which the mind is based but over the legitimacy of one’s own status as a living thing.

This may be the cause of a phenomenon noted by Masahiro Mori (1970) in his analysis of positive and negative familiarity expressed by humans in relation to an
artificial body or limb. Most graphs produce a steady curve and indicate a standard relationship between one factor (the horizontal) and another (the vertical). Mori found such a relationship between descriptions of familiarity (recognition) and the rise in an object’s similarity to the human—the more similar, the more familiarity the human described. But he noted that at some point, the graph dipped sharply, indicating that negative familiarity (or consciousness of strangeness) had been reactivated, and that only when a certain further threshold had been crossed did it rise again to finish its previous trajectory. He called this dip the uncanny valley. Objects that fell within this valley—objects that were both too similar and too dissimilar—were found to be disturbing. This produced an end to fascination with the imitator and a sudden plummet into revulsion. It is possible that the latent anxiety described above is expressed through such a fear of the personification of matter.

An experience of the android as uncanny can be induced quite easily. The aforementioned Japanese pink (or mildly pornographic) film Maid Droid (Tomomatsu, 2008) at one point portrays a love affair between a completely subservient gynoid (or specifically female android) and a human male, with the gynoid wanting only to please the (for some reason bashful) human male and refusing to take “no” for an answer. This film only works—either as a pink film or as a love story—because a human actress is portraying the gynoid. To experience the uncanny, one need only view video of the androids currently being developed at Hanson Robotics and imaginatively substituting one of these for the actress in the film. The image of such an android refusing to take “no” for an answer becomes quite repulsive and reveals the film not to be a love story about a
gynoid but rather an expression of a general male fantasy of a really eager and completely subservient sexual partner.

The apparatus-toward-being is a direct expression of a deep anxiety over the legitimacy of this basic categorization—this discrimination between that which is dead and that which is alive. Eric Wilson (2006) has suggested that this is itself the cause of the apparent lacuna of meaning identified by Jung, Tillich, and MacIntyre. He wrote of “our contemporary condition,” as “a philosophical and psychological crisis generated by ambiguity over the difference between organ and machine” (p. 7). If such a basic categorization is rendered false by the presence of something that transcends it, there are ramifications that may cascade through the entire system.

Such categorizations are not arbitrary. Rather they carry with them implications of allegiance to one causal (descriptive) system or another. That which is alive falls more under the purview of psychology, whereas that which is dead, or nonliving, is under the authority of laws described by physicists and manipulated by engineers. Sehon (2005) wrote of the apparent contradiction in the worldviews of psychologists and engineers:

Talk of mind and agency does not seem to mesh well with the language of physical sciences. Notions like action and purpose, the cornerstones of our descriptions of agents, appear to have no role in purely physical descriptions of the world. Normative notions like moral responsibility and criticizable irrationality, also central to the common-sense perspective, likewise do not belong to the conceptual arsenal of physics. (pp. 3-4)

The tension comes when the ideas of physical science are applied to beings that have always been considered free agents:

On the other hand, since human beings are physical objects, they must ultimately be appropriate subject matter for physical science, and they must behave in accord with physical laws. Physical science must be able, in principle, to provide detailed causal accounts of all bodily motions. (p. 4)
Allowing the human to remain in the category of the living leaves the human in the realm of the free, and criticizable, agent. To transpose nonliving matter into the category of the agent (without mediation through preexisting systems of cells and organs) is to make porous the boundary between the two categories. This threatens to nullify the dichotomy itself and potentially to bring the human into the realm of the apparatus, making it subject to the determinism of the apparatus. This is the collision of causal systems that is imminent in the apparatus-toward-being, a collision the anticipation of which—as Wilson wrote—might itself be a major or main cause of the apparent lacuna of meaning currently experienced by humans.

At present, there does not seem to be any widely accepted resolution of this conflict in Western thought. The roboticist Rodney Brooks (2002) wrote of a sequestration of thought systems that results from this lack of integration:

On the one hand, I believe myself and my children all to be mere machines. Automatons at large in the universe. Every person I meet is also a machine—a big bag of skin full of biomolecules interacting according to describable and knowable rules. When I look at my children, I can, when I force myself, understand them in this way. I can see that they are machines interacting with the world. But this is not how I treat them. I treat them in a very special way, and I interact with them on an entirely different level. They have my unconditional love, the furthest one might be able to get from rational analysis. Like a religious scientist, I maintain two sets of inconsistent beliefs and act on each of them in different circumstances. (p. 174)

Such a compartmentalization of thought can only be seen as a stage, not as an end, with an unknown future integration representing the potential advancement to a new stage—and the apparatus may trigger this.

It is this fear of the merging of life and death that seems to be at the heart of the anxiety we feel when faced with (at this time early) iterations of the apparatus-toward-being. But this may be a dichotomy that has outlived its usefulness—at least as the
dichotomy is understood at this point in our history. At the end of Solaris (Cameron, Landau, Sanchini, & Soderbergh, 2002), the protagonist finds himself in a simulated world, after having fallen into the energetic matrix of the planet, Solaris. When faced with the image of his late wife, he asks if he and she are alive or dead, and she answers: “We don’t have to think like that anymore.” With this line, this character may have summarized the direction in which we are headed, as we continue to remake ourselves with such nonliving materials as these.

The Dichotomous Artifact

As previously stated, what makes it so threatening for such a basic category to have been this delegitimized is the fact that the mind—an artifact constructed by the Geist out of the raw material of the psyche—is, in fact, a kind of dichotomy engine. It operates via dichotomous categories—edible and nonedible, healthy and unhealthy, me and not me, us and them.38 In fact, Freud (1919) wrote that the feeling of the uncanny

37 The more transient features of the Geist can be called the Zeitgeist (spirit of the age), while the aspects of our makeup that issue out of our phylogenetic heritage could be called the Urgeist (original spirit, which Jung might say contains the archetypes). Between the transient and the unchanging lie the more paradigmatic features, which require deeper and more difficult adjustments (paradigm shifts), where the features of the Urgeist are relatively eternal and the features of the Zeitgeist, by definition, unfixed (the process of individuation, as described by Jung, could be thought of as a task of uncoupling more and more of these paradigmatic features from the Zeitgeist and grounding them in the Urgeist, increasing their stability over time and situation). The Geist is the average of resources individuals in a particular time and place have to bring to bear on any given problem. It defines to a large degree the horizon of possibilities for both thought and action. To speak of the Geist is to speak of the state of development of the human species as a whole and to speak also of the relationship between the outer and the inner, the transient and the longevitous—the Zeitgeist and the Urgeist. To speak of a person as either solid or transitory is to locate their center of gravity either higher or lower in the hierarchy of the psyche and Geist.

38 It is little wonder that the machine created by this hominid to engage in computations uses as its language the most basic of dichotomies—every bit of code is either a 0 or a 1,
was experienced because some foundational discrimination of opposites had been challenged:

   It would seem as though each one of us has been through a phase of individual development corresponding to that animistic stage in primitive men, that none of us has traversed it without preserving certain traces of it which can be re-activated, and that everything which now strikes us as “uncanny” fulfils [sic] the condition of stirring those vestiges of animistic mental activity within us and bringing them to expression. (p. 13)

The feeling of the uncanny is a feeling that an extremely basic dichotomy—that between the living and the dead, for instance—has been transgressed and that one has been cast back into a precategorized or even magical (that is, completely uncontrollable) mindset.

The experience (as such) of the machine as a living thing is just such an experience.

   Conscious discrimination of opposites was, for Jung, the beginning of wisdom. 39

In his transposition of Hegelian historical philosophy into the psyche, he posited that the extrapolation of opposites into thesis and antithesis was a necessary precedent for the overcoming of dichotomies in a transcendent third—in Hegelian terminology, the production of synthesis. The function of the mind seems to be fundamentally based upon discrimination—and the distinction between that which is alive and that which is dead is fundamental to the entire construct.

   What is upsetting to this construct is that apparently dead matter is seen to be behaving as if it were alive, and the threat is that one might become (or might be shown to have been) unable to clearly discriminate between that which is dead but is feigning

---

39 In the way that dreams were the royal road to the unconscious for Freud, the conscious discrimination of opposites was, for Jung, the royal road to individuation. This discrimination is itself a recreation of that insight that is said to have ousted Adam and Eve, the original humans, from the naïveté of the garden—and which initiated the moral demands that would be made on humans from that time forward.
life (and how does something feign life if it is not at least suspended between life and
death?) and that which is authentically alive. To Baudrillard (1981/1994), the inability to
distinguish between the simulator and the authentic threatens our rational stability. He
used the example of an individual feigning a condition in order to be discharged from
military service and the inability of contemporary professionals to distinguish between
genuine pathology and a kind of mimicry that constellates all the signs of genuine
pathology: “This lack of distinction is the worst kind of subversion. It is against this lack
of distinction that classical reason armed itself in all its categories” (p. 4). It is this lack of
distinction that threatens to subsume the individual and to deconstruct the mind, and the
embodiment of this threat is the apparatus-toward-being itself.

If this is the anxiety, then the apparatus-toward-being is the terror-object—the
anticipated proof of our nonbeing. If one cannot distinguish between the simulator and
the simulated, if all the signifiers of the one are present in the other, the question that
must be asked is whether there is any true difference between the two—whether either of
them can be described as authentic. If there is no discernable difference, that which could
be said of the one would also apply to the other. This is, again, the establishment of heavy
or weighty equivalence.

This anticipation of equivalence combines with an ancient idea and fear. It taps
into an anxiety directed both outward and inward—a doubt that seems to be a common
inheritance of our species, and which typically occurs to us in childhood. It is a doubt as
to the reality or authenticity of our experience.

---

40 This ancient-ness is personal. It is ancient in terms of the personal psyche and continues to exist through one’s life as a sediment or stratum within the layered psyche, as a childhood (and thus personally ancient) experience, beneath the subsequent layers of sedimentation (or intervening experiences).
As Moravec (1999) pointed out, this doubt has been on the minds of roboticists for a number of years:

Turing\(^{41}\) noted that although we may each know ourselves to be motivated by thoughts and feelings, we have no direct evidence about anyone else. Perhaps all others in the world are actors simply reading lines, or movie projections, mere patterns of light and sound. There is a name for this philosophical position—solipsism—but it has few followers. We usually believe others when they describe motivations and emotions similar to our own. (p. 82)

Solipsism is a thought against the authenticity of the other, a suspicion that the other may be misrepresenting his or her inner life as an act of manipulation or deceit—that the other may be evoking the signifiers of authentic inner experience in order to simulate emotion. It is, in essence, suspicion of sentiment.\(^{42}\)

Thoughts against the other are never pure, however, and on some level the doubt is also turned against the self. Solipsism becomes self-solipsism, a doubt as to the authenticity of one’s own soul. This thought is so threatening as to be repressed—not willed into forgetfulness through suppression but aborted prior to its birth into consciousness, deemed by some censoric monitor to be unthinkable. While there are few explicit adherents of the philosophy, all individuals continue to experience a doubt that was never quashed in childhood.\(^{43}\) It is this fear, this lingering question that was never satisfactorily answered, that may be at the heart of our experience of the uncanny when faced with prototypes of the apparatus-toward-being. It is a fear of a kind of self-transparency that could best be analyzed via several contemporary metaphors.

\(^{41}\) Alan Turing, a mathematician and computer scientist, who originated the Turing Test as a means of assessing whether artificial intelligence had truly been achieved.

\(^{42}\) As defined by Mazis (2008), who wrote of sentiment as a substitute for authentic emotional participation between two persons (p. 33).

\(^{43}\) Even the maxim, *cogito ergo sum*, so often repeated, has not exorcised this doubt.
The Simultaneity of Three Boxes

As already stated, there are three metaphorical or mythical boxes that can be applied to the same situation. When I reference the simultaneity of these boxes, I mean that for our purposes, there are not three boxes but one box expressed in three different ways. This was not the intention of the original authors of these metaphors and myths, but the boxes converge in such a way as to bring a new depth of insight to a very challenging topic. It could be said that these three boxes—taken from psychology, from theoretical physics, and from ancient myth—exist in the same metaphorical space. Their meanings at times diverge, but they also at other times converge, and the qualities of one can, in these latter moments, be seen as having some relation, one to another.

There are two contemporary metaphors that coincide with one another, although they come from two very different fields, both of which unconsciously evoke a metaphor further removed from both—namely that of Pandora’s box. I would like to trace my own earlier reactions to both of these metaphors and to make some suggestions about what these two metaphors, taken together and with the myth of Pandora underpinning them, express.

The first of these, a metaphor that has worked its way into popular consciousness as an illustration of the principles of quantum mechanics, is that of Schrödinger’s cat. Astrophysicist John Gribbin (1984) described the thought experiment put forward by physicist Erwin Schrödinger in 1935:

A radioactive atom... might decay, emitting an electron, say; or it might not. It is possible to set up an experiment in such a way that there is a precise fifty-fifty chance that one of the atoms in a lump of radioactive material will decay in a certain time and that a detector will register the decay if it does happen. Schrödinger… imagin[ed] such an experiment set up in a closed room, or box, which also contains a live cat and a phial of poison, so arranged that if the
radioactive decay does occur then the poison container is broken and the cat dies. In the everyday world, there is a fifty-fifty chance that the cat will be killed, and without looking inside the box we can say... that the cat inside is either dead or alive. But now we encounter the strangeness of the quantum world. According to the theory, neither of the two possibilities open to the radioactive material, and therefore to the cat, has any reality unless it is observed. The atomic decay has neither happened nor not happened, the cat has neither been killed nor not killed, until we look inside the box to see what has happened. Theorists who accept the pure version of quantum mechanics say that the cat exists in some indeterminate state, neither dead nor alive, until an observer looks into the box to see how things are getting on. Nothing is real until it is observed. (pp. 2-3)

Considering this idea as a younger man, one of the thoughts that occurred to me was that the cat, which is treated as a passive element, should be able to bear witness to its own existence—as long, of course, as it is alive and still able to bear witness to anything. This was not the point of the thought experiment—if one remained in the realm of quantum mechanics, pushing this point would be tantamount to pushing the metaphor to the point at which it breaks down. But it occurs to me that this metaphor could be apt if moved into the purview of psychology and that such a move could take place through another metaphor already rooted in modern psychology—Skinner’s, which acts as the entry point to psychology.

The idea of the psyche as a black box has entered the Geist through the work of behaviorists such as B. F. Skinner, and it is here—in the two boxes, both holding unseen contents—that the two metaphors can converge, and, in the context of our ancient self-solipsism, become a single and central metaphor for the human psyche as it is now understood. We might then say that the supposed homunculus in the black box of the human psyche is the self in which we hope to continue believing—a homunculus that either does or does not bear witness to its own existence, and which either confirms or nullifies our being-ness based on its presence or lack thereof. We thus have moved into
the question of an inner being-ness that either perceives and thus solidifies its own being-
ness, transcending the nagging self-solipsism that has caused such existential anxiety, or
does not. If being-ness does exist, it would bear witness to its own being-ness. If there
truly is no being-ness hidden within the black box, however, there would be no such
internal witness. To open the box—or to increase its transparency to the point that its
contents (or lack thereof) could be witnessed—would be tantamount to nullifying belief
in one’s own being-ness. The question becomes whether the cat is alive or whether there
was never any cat to begin with—either there is being-ness that confirms its own
existence, or there was never any being-ness. To discover the latter would be to
experience a kind of retroactive self-annihilation.

Jung’s philosophical stance on the significance of consciousness to the cosmos
strengthens this line of reasoning. He was adamant that humans somehow complete the
act of creation through their acts of observation. Jung (1963/1989) wrote of a realization
he had while observing the Athi Plains of Nairobi:

This was the stillness of the eternal beginning, the world as it had always been, in
the state of non-being; for until then no one had been present to know that it was
this world.... There I was now, the first human being to recognize that this was the
world, but who did not know that in this moment he had first really created it.
There the cosmic meaning of consciousness became overwhelmingly clear to me.
(p. 256)

Jung went on to recall the Pueblo and their belief that they were responsible for the sun’s
passage through the sky, as well as his earlier question as to his own myth.

Now I knew what it was, and knew even more: that man is indispensable for the
completion of creation; that, in fact, he himself is the second creator of the world,
who alone has given to the world its objective existence—without which,
unheard, unseen, silently eating, giving birth, dying, heads nodding through
hundreds of millions of years, it would have gone on in the profoundest night of
non-being down to its unknown end. Human consciousness created objective
existence and meaning, and man found his indispensable place in the great process of being. (p. 256)\(^{44}\)

In the same way that solipsism can be turned back toward oneself, we can ask whether our completion of the task of creation does not also extend inward, toward the center of our very being, to our being-ness. If we are separated from our depths—if that which we are is hidden within the black box of the psyche—is there something there, also hidden within the box, to bear witness to itself? Or are we possessors of an essentially empty box into which we have projected an illusion of depth or order, a gestalt that does not exist? Do we, in fact, create our souls through our belief in them—and if so, are our souls truly existent? Do we in some ways bring ourselves into being through a self-reflexive pseudo-witness, with belief acting as a substitute for actual witness?

The anxiety produced by self-solipsism is focused on the possibility that we are guilty of an overestimation of our own being-ness. The black box that Skinner located within each of us becomes Pandora’s box, and our aversion to bearing the consequences of its contents (resistance to, and avoidance of, inner work) may be rooted in the knowledge of what danger lies therein, not because of what it holds but because of what it may fail to hold. Recognition of the personhood of the other (the apparatus) may prove the key that unlocks this box and forces a confrontation with its potential emptiness.

Each of these metaphors places us outside the box—at least the ego is outside it and is unable to penetrate its borders through methods other than inference (dream

---

\(^{44}\) Although Jung was very much intrigued by quantum theory and attempted, in his book *Synchronicity* (1952/1973), to combine that field with his own, the metaphor of the cat in the box would not have been known to him at the time of his trip to Africa in the autumn of 1925, ten years prior to the year in which Schrödinger, after frequent correspondences with Einstein, originated his metaphor, 1935. Even so, he was likely familiar with the idea that necessitated the metaphor.
analysis if one is depth-oriented, measurement of responses to stimuli if one is a behaviorist). The ego and the mind—the construct in which the ego lives—exist in separate space from the psyche. They have separated themselves from the remainder—they have become satellites to the psyche. To speak of displacement is to speak of a rightful position being usurped. The ego has not been displaced. It is not displacement that led to its exile, but rather its movement away from the greater center, a movement that cannot be discounted. But the ego remains in exile from a greater whole.

Genosko (1994/2002) referenced Baudrillard in terms of “the progressive satellization of the referent” (p. 42). The referent is that which was original, that which was more authentic, and which was re-created or re-presented. Graham (2002), in her analysis of Baudrillard, wrote: “His is a world in which… truth and reality are indistinguishable from illusion or simulation” (p. 191). She further wrote: “Virtual technologies also enable reality to be synthesized so that the copy is more real than the original—if indeed there is an original. The representation of the real becomes the real” (p. 192). We come to a point at which the simulation seems more authentic than the original, at which the simulation embodies the spirit of what is simulated more than that which is simulated. The original referent is then measured against the simulation, which has assumed the position of authenticity. The referent has been displaced—it has become a satellite to the reproduction.

The construct becomes a hyper-reality, not a true-to-the-original replication but one that is hyper-true-to-the-original, incorporating elements that somehow deviate from the original but seem to enhance the authenticity of the reproduction or more completely evoke the ineffable spirit of the original. I am reminded of Romanyszyn (2004) and his
description of the figures in bronze on the doors of St. Mary’s in Hildesheim, Germany: “we might suggest, then, that the bodies depicted in these bronze doors are pantomimic bodies, bodies whose gestures are inseparable from the emotional situation and the story they enact” (p. 108). These pantomimic bodies, cast in bronze, seem somehow truer to the human story than anatomically proportionate depictions of humans would be. As long as one remembers that the simulation—the pantomimic body—is a Nachbild (a representation) and not the Vorbild (the original), there is no problem. But, as Baudrillard indicated, there may be a strong tendency in humans to confuse the one for the other, to the detriment of the latent (and now in many ways cast off) referent.

Rather than remaining central or primary, the referent is decentralized and is made to orbit the simulacra, which is taken to be more authentic than the original. I find in this metaphor of the referent as a satellite a link to Freud’s view of what in light of the current argument might be called the satellization of what became the ego—that portion of the psyche that achieved escape velocity, but which remains only a small portion of the whole, distinct but connected. In some ways, the films Solaris (Tarasov, & Tarkovsky, 1972) and its remake (Cameron, Landau, Sanchini, & Soderbergh, 2002) portray such a satellization. The after-images that appear on the space station orbiting Solaris—shadows of people, alive and dead—could be thought of as attempts by the greater whole (Solaris itself, as a kind of psyche) to emulate the egoic being-ness of the humans. These beings are themselves isolated from Solaris, so that even as they sense they are a part of it they also cannot fully understand it or describe its intentions. So it is with the ego and its relation to the greater psyche.
In his analysis of the uncanny, Freud (1919) traced the development of the double, or doppelganger, through Otto Rank, who wrote of the soul as the first double of the body—an attempt to overcome death via an imagined double existing in a spiritual realm somehow contiguous with ours. Freud indicated that there was then a shift toward an inner isolation and solidification of the double:

A special faculty is slowly formed there, able to oppose the rest of the ego, with the function of observing and criticizing the self and exercising a censorship within the mind, and this we become aware of as our “conscience.” In the pathological case of delusions of being watched this mental institution becomes isolated, dissociated from the ego, and discernible to a physician’s eye. The fact that a faculty of this kind exists, which is able to treat the rest of the ego like an object—the fact, that is, that, man is capable of self-observation—renders it possible to invest the old idea of a “double” with a new meaning and ascribe many things to it, above all, those things which seem to the new faculty of self-criticism to belong to the old surmounted narcissism of the earliest period of all.

(p. 10)

So the apparatus-toward-being is not, in fact, the first instance of humanity creating its double. Humanity, as per Rank, created its soul as a double, and this inner double became the source of our self-reflexive cognition. We may have made the qualitative transition from *homo sapiens* to *homo sapiens sapiens* because of this doubling.

The soul—the Being-ness of humans, being-ness reified by a deity whose own reality is a hyper-simulation of reality (a simulation of indisputable reality)—may then be the pantomimic double, the hyper-real simulation of an archaic and perhaps somewhat smaller original being-ness. A question that arises is this: If our self-reflexive cognition came about because of the presence of this inner simulation (the homunculus), would such cognition be able to sustain itself in the absence of the double?

The mind, as the artifactual portion of the psyche, is as much a construction of the *Geist* as are the space capsules that reenact its detachment and satellization from the
psyche through a literalized satellization of construct from the earth. It is this structure—the mind itself, based as it is on discrimination of opposites—that at least appears to be threatened by the transcendence of the categories of living and dead. It is Being-ness that is threatened by these metaphors and their implications.

Beneath and beyond the threat to the mind is the threat of complete self-annihilation, proceeding from the depths outward. As previously noted, Edinger wrote: “The breakdown of a central myth is like the shattering of a vessel containing a precious essence; the fluid is spilled and drains away, soaked up by the surrounding undifferentiated matter” (p. 9). The belief in an inner self is central to our psychology. This could even be seen as the central myth of the Geist itself. What would it mean to find ourselves suddenly without that central myth of some internal essence?

There are those that hope to make it to the end of their lives without ever opening this inner box, without peering too closely and seeing its apparent or potential hollowness. It is a Pandora’s box, this inner space, and they hope, like Orpheus, not to turn and look back before they have made it safely away—to turn back and lose Eurydice or to turn back and transform into a pillar of salt (or spirit-less matter).

And yet, Psyche only became divine when she peeled back the darkness, when she looked upon Eros, who forbade her to look at him as Hades forbade Orpheus to turn and look upon Eurydice before they reached the upper world. Many seem to wish to achieve divinity through obeying what seems to be a command from the divine—that they not look too closely at that which is within themselves, whether it be ugly (as Psyche’s sisters told Psyche her suitor must be) or beautiful or simply not there.
I see Skinner’s black-box metaphor as a defense mechanism. I would go so far as to surmise that it is a collective defense, embedded in the Geist and brought to expression through B. F. Skinner, and that it is specifically a defense against the transparency that was coming via scientific progress—such that even as it appeared to reduce the human to the status of a machine (through its emphasis on stimulus and response), it also insulated the innermost self (the soul) from critique. It hid the soul behind the walls of a box and insisted that there was no need to open the box. This could read as a cursory acquiescence to deterministic interpretations of human being-ness, with the hidden purpose being to protect the human’s being-ness from the fullness of a deterministic rendering. What we may have, in Skinner’s black box, is not an absolute statement of determinism but an expression of the evasiveness of human consciousness.

As we progress toward a full replication of our humanity in the mechanical, however, this insulation will become less and less stable. Doubt as to one’s abilities can lead to timidity and depression. Doubt as to one’s essential being-ness may lead to something deeper, even to spiritual annihilation. The shattered vessel of which Edinger spoke may, in the end, turn out to have been a black box that was breached or rendered transparent.

**Re-enactment of Divine Evacuation**

As previously stated, an icon that shows itself to be hollow—a holo-graphic self-icon—may have little effect on us as a species, outside of the displacement of a number of us from the workforce. A self-icon of depth, however, one that reflects our depths even as it is demonstrably constructed in its entirety, might result in an evacuation of meaning from the human. Such would actually be a reenactment of a previous evacuation, that of
another self-icon, formed of and by humanity as a predecessor to the apparatus itself, namely the satellized ego-ideal—deity itself, projected onto the sky like a satellite to look down upon us in spiritual authority. The emptying of this divine ego-ideal may have started a process (evacuation of the latent referent, the human) that is to be continued through the presence of the apparatus-toward-being.

Nietzsche (1882/2001) declared this satellized self-icon to be dead and, in so doing, acknowledged its growing inability to function as an organizing principle: “and, now that this faith has been undermined, how much must collapse because it was built on this faith, lean on it, had grown into it—for example, our entire European morality” (p. 343). No longer could this belief structure and direct human life.

Tillich (1952) wrote of this traditional system of spiritual morality as a system of participation, in which anxiety could be, not annihilated, but overcome—at least as long as the system remained in place (p. 62). In fact, Tillich later wrote:

The decisive event which underlies the search for meaning and the despair of it in the 20th century is the loss of God in the 19th century.... The result is the pronouncement “God is dead,” and with him the whole system of values and meanings in which one lived. (p. 142)

Nietzsche declared this self-icon, as an organizing principle, to be hollow—emptied of content, emptied of purpose and efficacy of purpose. Edinger (1984) perhaps unconsciously evoked Nietzsche when he wrote of the consequences of the loss of myth (or, put another way, the death of our connection to the divine): “Meaning is lost. In its place, primitive and atavistic contents are reactivated. Differentiated values disappear and are replaced by the elemental motivations of power and pleasure, or else the individual is exposed to emptiness and despair” (pp. 9-10). It is possible that Nietzsche—the man who declared the idealized image of humanity to be dead in terms of its ability to form
humans in its likeness—was attempting to redeem the atavism that he knew would come, once the remainder of the world joined him in the recognition of this death. He attempted to find the creative aspects of the will-to-power and, through this, the possibilities for positive outcomes as a result of this atavism.

Baudrillard (1981/1994) also reflected Nietzsche’s pronouncement of the death of God when he portrayed the current situation of signs without referents as one in which “there is no longer a God to recognize his own” (p. 6). This latter statement leads to an interesting thought. The compulsion to originate or imagine a divine being in order to explain the “beginning” of the world may not be motivated purely by curiosity. Observation of the world does not lead as a matter of course to the desire to explain the beginnings of something that is, after all, experienced as eternal—the question does not necessarily follow from the human’s experience of the unchanging nature of the earth. This compulsion, then, may not be as pure as previously thought and may issue more out of a desire to more firmly establish one’s tenuously felt existence.

The belief in a divinity—and specifically a single creator-divinity—may have been an attempt to establish the reality of one’s soul by appealing to a metaphysical being-ness that could confer such reality onto one’s existence. Those several divinities that created the material realm out of nothing outlived those that were themselves formed of the matter of a pre-existing universe because their prowess for creating things that were objectively real made them especially useful for establishing one’s own reality through belief in them. The world is seen to be objectively real. The soul, however—the reified being-ness of the human—is believed to be something held apart from the world,
and it is this soul that is suspected of having interloped into the material world. The projection of a being-ness like unto our own onto and even beyond a “beginning” of the objective (physical) world privileges the spiritual (that which is felt to be an interloper in the world and is doubted via solipsism and self-solipsism) and attempts to exorcise this feeling of unreality by projecting it onto the objectively real itself—matter then becomes the interloper, a corrupting influence, an illusory medium in which the spirit has unfortunately fallen. It becomes one’s moral duty to doubt material itself through this metaphysical shadow projection.

My own interpretation is that this adamant belief in the corruptibility of matter, this long-running shadow projection, proceeds from our self-solipsism and from our need to insulate ourselves from such self-doubt. We have thus attempted to establish ourselves more firmly in our own minds through an appeal to a pre-existent soul-creator that might or might not have formed us out of his own spiritual substance. If such a divinity created the objective world, then his being-ness—and, by extension, ours—could not be rationally disputed. We have reified our being-ness. We have converted it to Being-ness. This secret tautology (the god made in our image reifies our inner self-image) was in some ways made explicit by Anselm, who argued that a perfect being of which we conceive must exist because perfection entails existence and that this perfect being must exist because it is “the efficient cause of the being… of everything else” (Honderich, 2005, p. 38).

45 This solipsism and self-solipsism may have been expressed through the idea of the incompatibility of soul and matter and may, in fact, be ancient phylogenetically as well as personally. This may be the reason it occurs to so many children of our species.
The theory, then, that I am putting forward is that belief in this ego-idealized divinity was part of the act of self-creation. Positing the first act (the primordial, mythical creation by an authority of being-ness) emboldened the second act, the continued self-creation in the image of the (imagined) first actor. The belief in such divinities may have aided in the development of the human Geist and, through its influence on all human children growing within it, the development of the human mind as well. We projected an ideal into the sky and then attempted to model ourselves after it.

But if this original actor, in whose image we have attempted to make ourselves, is shown to have been eternally a vapor, an illusion made in our image, what then? As Baudrillard (1984/1991) wrote, “What if God himself can be simulated, that is to say can be reduced to the signs that constitute faith? Then the whole system becomes weightless, it is no longer itself anything but a gigantic simulacrum” (pp. 5-6). The divinity, the concretized ego-ideal, shown to have been hollow, is no longer able to draw humanity toward itself. Belief can no longer function as part of a process of self-creation, and the system of self-creation upon which it was based fails. In fact, belief in oneself falters and, perhaps, will fail as well.

As previously noted, Baudrillard (1981/1994) wrote that simulation is feigning possession (p. 3). We did not have an initial act of creation, an assurance of the reality or import of that which we called the soul, so this act was simulated through belief in it, and we have lived within this simulation for ages, unable to extract ourselves from it enough to see its illegitimacy. This original creator—or mythical original referent—was the center of the entire system. Without it, the center, such as it is, cannot hold, at least in its current configuration or form.
It should be noted that in order for this belief to work as a solidifying or limiting and delimiting factor, we had to believe in our own *ex nihilo* creation—our creation out of nothing. In order for this divinity to have been of sufficient credibility to solidify our being-ness, he had to be absolute, not something that formed us out of matter already present when the deity arrived on the scene. The deity that has survived the eons is one that created the earth from nothing and created us from the earth—and perhaps also partially from himself.

But another phrase—or rather the continuation of the phrase, *ex nihilo*—exposes the anxiety hidden within this doctrinal attempt to abate our existential anxieties. The Christian religion has adopted the stance that the world was created out of nothing. This runs counter to the Greek stance, elucidated by Parmenides in the phrase, *ex nihilo nihil fit*: from nothing, nothing comes. We know, now, that matter consists of emptiness surrounded by a forcefield—the electron cloud (a “cloud” formed by the fact that the electron may, in fact, exist in all possible locations at once, until, of course, it is observed via interaction). Matter, it now seems, is quite empty. And if the divinity that formed us is also shown to be hollow, what becomes of the crown of that divinity’s supposed creation, the soul that may have been born of the divinity’s own (spiritual) flesh?

As this ego-ideal empties, humans also—said to have been made in its image—are poised to follow it into an abyss. Pulled out of nothingness (undifferentiated psyche, the primordial sea of chaos) by this organizing principle itself (created by the burgeoning mind or ego as a principle that further participated in the creation of the mind—self-
creation via ego-projection\textsuperscript{46}, the ego may now be following this ego-ideal back into the nothingness from which it arose. The ego linked its being to the being-ness of this projection in an attempt to solidify its defenses against self-solipsism. And the link is strong enough between ego and projected ideal that the ego is now being dragged down into the depths of oblivion by this albatross. The being said to have created our being-ness is disproved. The initial act of creation itself is thus retroactively disproved. And we ourselves are in danger of a retroactive annihilation via our own historical attempts at solidification or metaphysical justification.

If there is no longer any god to recognize his own, there is no longer any affirmation from above—if we concede this god’s death, there is no longer even any virtual affirmation from above. And if this god was always a simulacrum—the projection of a being-ness we hoped was already within us, and which was then followed by an attempt to embody that imagined being-ness—can it not, then, be said that whatever it is in us of which this god was a projection also never existed? We no longer have the ability to call on some internal self-signification. The second act of creation is nullified as never having taken place, or having taken place only in a simulated form. We find ourselves to be, and to have been, simulated beings—we find ourselves to be virtual.\textsuperscript{47}

\textsuperscript{46} As the ego formed, the ego-ideal continued to form. Such can be seen as the point of Jung’s analysis of Job, who transcended his culture’s god-image and moved this image via his own development or maturation.

\textsuperscript{47} Like the self-annihilation that comes with the same recognition of the replicant in the film \textit{Impostor} (Phillips, Weinstein, Weinstein, & Fleder, 2001)—who detonates upon seeing the body of the human upon which he was modeled, and who realizes, for the first time that he is, in fact an imposter (this guilty knowledge acts as a secondary trigger)—this guilty knowledge of Being-ness’s overestimation of itself may act as a deconstructive agent.
As we have gazed inward, we have hoped that within this black box—which we must access indirectly, via dream-work and other methods—there is something real and that this inner reality can itself confirm our reality by bearing witness to itself. But recent developments in the *Geist* have led to this idea that the *Schrödinger’s cat* that is our purported soul may be proven to be dead (and thus never to have existed). There is no longer any confirmation from above—there is no deity to pry open the box and peer within, to recognize his own. Only confirmation from within (the cat’s observation of its own existence) can prove our salvation.

But we have been satellized by our own doubles; or, to be more precise, we have displaced ourselves and have placed our doubles (our souls) within our own hollows. The ego is now unsure of what does or does not exist in or near its center.

It has now become common knowledge that Freud postulated this kind of inner dissociation, in which some part of the psyche was separated from the greater part so that the “me” and the “not-me” were contained within the same overarching whole. The ego and the id were separate mechanisms working within a single individual. The superego was another internally coherent figure or mechanism within the greater whole that was best understood as being separate from both the ego and the id.

MacIntyre (1984) saw this inner separation—at least where the superego was concerned—as historical, that is, not as a permanent state but as the state of the psyche in Europe at the time of Freud’s writing (p. 72). The entirety of psychoanalysis and of the field of depth psychology is predicated upon this inner exile from the center. In the introduction to his study of alchemical symbolism and its relation to individuation, Jung wrote:
It is a remarkable fact, which we come across again and again, that absolutely everybody, even the most unqualified layman, thinks he knows all about psychology as though the psyche were something that enjoyed the most universal understanding. But anyone who really knows the human psyche will agree with me when I say that it is one of the darkest and most mysterious regions of our experience. (1944/1968, pp. 3-4 [CW 12, para. 2])

At this point in our history, humans do not act out of their centers but rather out of their exile, and therapy is itself an attempt to act as psychopomp to the ego in exile. One might argue that we are not the black box, that we are inside ourselves and experience ourselves and that this negates the idea of an inner hollow. But this is not the experience identified by Jung, by Tillich, by MacIntyre, and by so many others.

Although Skinner’s black-box metaphor could be read as a defense against the kind of deep introspection that is required by psychoanalysis, this metaphor can also act as an illustration of the need for analysis. Jungian psychology could be seen as having the same metaphor, although it is expressed in rather more elegant terms. The box is located within the person, in Jungian psychology, and the ego’s task is to reconcile with the contents of that box—the complexes, and, beneath them, the archetypes, those patterns of thought and behavior that, as per Jung, are unknowable in their fullness (will remain always a mystery, at least in part) and which thus never emerge from the black box, even as the box shrinks, as it were, via this process of reconciliation.

Religion itself is an attempt to connect the individual to something hidden, and Jung wrote, in Answer to Job (1952/1958) that there was no way to distinguish between the unconscious and a metaphysical god (p. 106). In this way, religion can be thought of as a response to a deep knowledge that there is something from which we have been exiled, or from which we are somehow separated. Many religions include a narrative of such exile, with the story of Adam and Eve sinning in a primordial garden being the most
familiar in the West. We seem to have a deep (unconscious) knowledge that deity is internal, rather than external, and that our attempts to locate the divine are attempts to locate something within ourselves. The feeling of the death of god, then, is simultaneously the feeling of the death of some internal principle. It is a feeling of hollowness.

It is this exile from the center—this location of the center of awareness in the ego and the displacement of the ego by the soul—that made analysis necessary. The ego must be guided back toward the greater center (the truer center of the psyche, as opposed to the relative center of the ego) by a psychopomp. The psychopomp is a soul-guide, but the ego is itself led back to the place where the (perhaps fictional) soul was believed to have settled. We are plagued by doubt that there is no such thing as the soul in the box at all, that what lies in the center is like Lucifer, at home in the depths and ready to be a devil’s advocate against our most cherished belief (in our doubles and in the deity that reified our doubles). It has often been said, after all, that hell is the absence of the one-god. If this god is that which has long conferred the imprint of legitimacy on us, the absence of this god will often be felt as an absence of legitimacy—that is, as confirmation of our ontological anxiety.

The death of the eternal is necessarily an invalidation of its entire existence, a retroactive annihilation—for by definition death proves that eternity was a false claim. The death of the one-god—the death identified by Nietzsche—constitutes an eternal death, stretching forward and backward in time. And human being-ness, linked for so long to that supposed eternity, is now also threatened with retroactive nullification. This
is a nullification that stretches deep into our history, an insight that in some ways murders every individual member of our species, past and present, from within.

There was of course something at the beginning of our development toward consciousness. But whatever this was has been slowly obscured, to the point that it may no longer be recognizable as such. This more original being-ness was replaced with the divinely reified double (the soul), which is named (like Rumpelstiltskin), and in which we thus believe (to have named it is to have solidified it as a thing and to have made some claim of power over it). But if this being-ness—attested to by a god who can no longer recognize his own—was always an illusion, a double, more an attempt at consolation over the inevitability of death than a recognition of something actually present, then one must wonder what will be left to us if this being-ness evaporates under scrutiny? Will there be a paltry being-ness remaining, or will the entire endeavor of self-reinforcement be shown to have always been an exercise in simulation—feigning to have what we never had? As Graham (2002) wrote, “There is no underlying ‘essence’, simply a series of representations, enactments or simulations” (p. 193). We are left with something that at least appears to us—in the absence of the Being-ness we assumed was there—to be a hollow, something that may not be real in any way our minds (artifacts of the Geist) can currently comprehend.

In essence, the theory I have put forward is that we have already constructed a self-icon in the form(lessness) of an ego-ideal satellized away from the earth and gazing back down upon it, solidifying the earth and the souls in it through its observation, and we have in turn watched this self-icon emptied of any lasting significance. We are now poised to follow its emptiness, to continue to be made in its (now eternally emptying)
image—for it was, in the beginning, made in our own image, and it is now shown to have been hollow, a holographic (without substance, ethereal or otherwise) self-icon. And we are shown to have been, all along, our own simulacrum—made in the image of our own image, made now empty in mimicry of our own emptiness. As we pursue this new self-iconography, as we attempt to re-create ourselves, as we in fact take the position and place of our once-projected ego-ideal and become the creator-species in the coming dyad (of original referent and apparatus-toward-being), we will face our emptiness from another angle. We now are poised to create a thing in our own image and to then recognize its emptiness—and only then to fully recognize our own.

The biblical proscription against idolatry, or “the simulacrum of divinity,” as Baudrillard (1981/1994) described it, seems to have issued out of this same anxiety and seems, in fact, to have been an attempt to forbid a certain knowledge:

They predicted this omnipotence of simulacra, the faculty simulacra have of effacing God from the conscience of man, and the destructive, annihilating truth that they allow to appear—that deep down God never existed, that only the simulacrum ever existed, even that God himself was never anything but his own simulacrum—from this came the urge to destroy the images. (p. 4)

To the iconoclasts—those that would prohibit the use of icons or idols—the objectification of the divine would have represented the “death of the divine referential” (p. 5). This is, of course, the same death we now face as we create icons not of a mythical divine (ourselves in projection, an indirect self-iconography) but of ourselves as we know ourselves (direct self-iconography). And those that seek to establish what I have previously called hard-line pre-emptive solipsism may themselves be the inheritors of this iconoclasm—they are, in many ways, attempting to theoretically forbid (to render

---

48 Aniconism was expressed in the bible through stories of the golden calf (Exodus 32) and through numerous proscriptions in both testaments.
impossible or discount the possibility of) the deep iconography of humans in the apparatus. They are attempting to sever preemptively any weighty or heavy equivalence between themselves and the proposed apparatus so that they will not be required to recognize their own emptiness when the apparatus finally is present.

A question Baudrillard (1981/1994) posed takes on new significance in the line of reasoning presented above: if God can be simulated, does not the entire system of faith lose coherence, which is to say, become weightless (p. 5)? If artificial intelligence is the self-iconography of the human species, it would hold that the successful replication of the latent referent—that is, the production of some device that is able to constellate all the signifiers of the human in its fullness—would also be the death of the latent referent. If we can be reduced to the signs of our existence, and if the self-icon demonstrates a constellation of such signs robust enough to pass the most stringent and soul-searching forms of the Turing test—that is to say, if we find ourselves unable to distinguish between the being-ness of the apparatus-toward-being and our own—we may find that our being-ness has become weightless, a specter or after-image, a ghost. We would lose the weightiness of our actions, which would become meaningless by virtue of our new (recognition of our) status as virtual.

We are collecting images of ourselves in an attempt to obscure the absence of ourselves in our own lives. We attempt to obscure the absence, the desert or evacuation of the being-ness, through the introduction of signs without referents. Baudrillard (1981/1994) wrote of representation of the divine as an enactment of the death of the divine—the image, he wrote, was necessary for the maintenance of the illusion: “it is dangerous to unmask images, since they dissimulate the fact that there is nothing behind
According to Baudrillard, the procession of our relationship to what is represented follows a typical path:

Such would be the successive phases of the image:

- it is the reflection of a profound reality;
- it masks and denatures a profound reality;
- it masks the absence of a profound reality;
- it has no relation to any reality whatsoever: it is its own pure simulacrum. (p. 6)

We can see in this the progression of our faith in our own being-ness, from the belief that we are as we have perceived ourselves to be (a very simple and primitive psychology), to the belief that the ego is a mask (persona) for a deeper reality, to what Tillich called the third period of anxiety, the feeling that beneath the ego there may, in fact, be nothing at all, and finally to recognition that we are our own creators and thus our own simulacrum and that there is no signification beyond our own will-to-power.

Baudrillardian philosophy becomes a psychology of the human as a simulacrum. We are about to enter, or perhaps have already entered, the status Baudrillard assigned to our own creations—we are now signs without referents. As Baudrillard (1981/1994) wrote:

The transition from signs that dissimulate something to signs that dissimulate that there is nothing marks a decisive turning point. The first reflects a theology of truth and secrecy (to which the notion of ideology still belongs). The second inaugurates the era of simulacra and of simulation. (p. 6)

This is, in some ways, a reflection of the depressive position posited by Klein, who wrote of another kind of forbidden knowledge—the knowledge that good breast and bad breast were one and the same. Such knowledge brings a depression with it, as a comforting

---

49 This is not consistent with Jung’s stance on the image.
(schizoid) illusion is lost. But the knowledge forbidden by the iconoclasts is deeper and more threatening.

Such forbidden knowledge can itself be the beginning of its own end. If the original creation-act is put into question, the product of that act is put into question. We lose the ability to claim self-observation as that which solidifies our inner existence, as we no longer have anything to which we might attach our tenuous being-ness. We find that we are not in a position to confer upon our being-ness any real significance or to complete the act of creation through our self-observation. It is possible, then, that we will have reached the point at which knowledge nullifies itself—when the working out of knowledge becomes de-legitimization of knowledge itself and the working out of reason becomes the end of reason.

The threat of the self-icon of depth is that it will reveal humans to have always been *robo-sapiens*—collections of determined bio-molecules interacting in predictable ways. The self-icon itself will then be seen as a simulacrum of a simulacrum—a signifier without any origin in truth, the simulation of a simulation.

One could say, following Baudrillard’s line of thought, that we are entering the age of the eternal falsehood. It is little wonder, then, that he writes of images as “murderers of the real, murderers of their own model” (1981/1994, p. 5). The deity made in the image of humanity now threatens to murder humanity through his own death, through the scandal that is his noneternity (his temporality and virtuality). We are in danger of undoing ourselves via our seeing-through of our own doubles. We may lose the inner act of self-creation, the inner god-principle or organizing principle that is able to recognize its own—and with this loss of self-recognition, with the loss of any kind of
deep inner self-affirmation, we may find ourselves either entirely undone or at the mercy of the surround for our continued existence. We may found ourselves beholden to the only thing that would appear to sustain us, the situation described by Cushman (1995), with the empty self owing its entire continued existence to repeated acts of consumerism, the well-worn paths of the sentiment of the market. Human being-ness would then seem to be a thing holding itself together through participation in an increasingly meaningless machinery of exchange.

Reactions to the Coming Confrontation

There would appear to be a subtle recognition of the possibility of self-annihilation as an effect of the presence of the apparatus-toward-being, and Eric Wilson (2006) identified both a desire to escape the coming destruction and a desire to fall into and be subsumed by it.

We are confronted with a contradiction. The apparatus-toward-being will be the manifestation of both our prowess and also, potentially, our emptiness—like the annihilation of earth through nuclear weapons, which would prove that humans had acquired truly godlike abilities through a final act of proving humans not to be gods, the final experience of simultaneous self-gratification and self-annihilation.

Wilson (2006) noted a paradox within this fear of the loss of one’s being-ness. In addition to the fear of mechanization there is also a desire for mechanization, a desire to become a machine and to be, as is the machine, “untroubled by the rift between thought and action” (p. 2). Wilson saw this as “an instinct for Eden, forms undisturbed by shame,” (p. 2), and he identified this as part of the Geist:

Humanoid machines reflect forms of melancholia that have resulted from what human beings have perennially called “the fall.” These kinds of dejection are
inseparable from self-consciousness, the painful rift between mind and matter, knowing and being. To heal these splits, humans have created mechanistic doubles untroubled by awareness of self. These new Adams embody the spiritual potential of their suffering creators—the possibility that human beings might be able to transcend their self-centered fears and desires and return, egoless, to Eden.

(p. 2)

Wilson speaks of transcendence and return in the same moment because the human is torn between these two actions. Return is a vision of regression into unconsciousness, a sinking away from ontological self-solipsism into an egoless state—or a state in which there is no ego to cast doubt upon itself or upon the remainder of the self. This is also the despair of those that have been unable to meet the moral demands made upon them as a result of this ego-knowledge. Wilson wrote of this as “the unrequited nostalgia for graceful ignorance (the sad yen for bodily unity) and the unfulfilled hope for effortless knowledge (the gloomy gaze toward mental oneness)” (p. 18). The human responds to the question of its own emptiness and inadequacy by hoping to forget the question itself.

There is, at the same time, a compulsion toward transcendence, a hope for escape from the body into the ethereal imaginary. This escapist fantasy is another avoidance of the conflict between psychology and physics—if the mind is able to exist without a body, it cannot have arisen from material (determined) processes, or else it must have finally untethered itself from matter (that which is subject to the laws of physics) and made itself truly free. Romanyshyn (2004) wrote of this desire as exemplified by “reason which dreams of purifying itself of the body and its passions” (p. 155). He even defined the human as homo astronauticus—the hominid that is trying desperately to escape the material world, to affect its own exile and reconstruct itself as a satellite to its former place of centrality (p. 90). Reason hopes for a purification or purgation because matter, which behaves according to determined physical laws, feels like a contamination. In fact,
the knowledge of such laws is itself a contamination of the feeling of oneself as having
free will.

This simultaneous “hope for transcendence,” and “terror toward determinism,”
(Wilson, 2006, p. 2) represents a potential collision between these two competing poles
of the dichotomy. We wish either to rise into pure spirit, unburdened by matter or the
possibility of our own determination, or to fall into pure matter, avoiding questions of
purpose or morality or “criticizable irrationality,” as Sehon (2005) put it (p. 4). We wish
either to become fully gods (etherealized, satellized, as this is now our vision of
divinities) or to become fully matter (to fall into an egoless state, unconsciousness).

Baudrillard (1993) surmised that the attempt to replicate ourselves, and
specifically our cognitive abilities (the end-product of the endeavor has been labeled
artificial intelligence, after all) is aimed at “entrusting this burdensome intelligence to
machines [so that] we are released from any responsibility to knowledge” (p. 51). The
burden we wish to divest is introspective intelligence, the kind of self-questioning
intelligence that we have accepted as the most defining feature of our species—homo
sapiens sapiens. We suffer beneath the weight of Adam’s curse, as the ability to reflect
on one’s existence and one’s unavoidable death has been called, and it is possible that our
rushing forward toward AI been secretly motivated by a desire to construct a platform
capable of carrying this burden away from us, of taking over as the most rational beings
on the planet, so that we might fall back downward toward unconsciousness.

Jung felt the identification and synthesis of opposites to be essential to
progression of the personality toward some further wholeness or toward a more
comprehensive state of integration. Edinger (1984) offered a concise description of this process:

[Individuation] has as its most characteristic feature the encounter of opposites, first experienced as the ego and the unconscious, the I and the not-I, subject and object, myself and the “other.” Thus we can say that whenever one is experiencing the conflict between contrary attitudes or when a personal desire or idea is being contested by an “other,” either from inside or outside, the possibility of creating a new increment of consciousness exists. (pp. 17-18)

The conflict between identifying oneself with matter—embrace of the machine-self—and identifying oneself as spirit runs deep.

As noted above, the cyborg is not included in this study due to the primacy of biological components in its makeup. Any depth found in the cyborg could be explained away via an appeal to these components. The cyborg would thus fail in a fundamental way to challenge this categorization (the hard-line discrimination between that which is alive and that which is dead).

It is the fully constructed apparatus-toward-being that would both hold the deep image of the human and hold the human to its material accountability—or to the fact of its materiality and parsability. Where the cyborg or even the modified animal (the animal on which creative vivisection or genetic manipulation has been undertaken) would allow the individual to continue to deny or to avoid the parsability of his own attributes, the fully constructed and deeply reflective self-icon would force the individual to face his own transparency in its image. The emperor, as it is said, is shown to be without mystery (is shown to have been naked)—at least in his current configuration.

Involuntary Transparency and the Consequences of Such

Mazis (2008) wrote that “if it were possible to become truly disembodied spirits or intellects, then it would be the case that we would forfeit our interiority” (p. 31). In a
similar vein, Baudrillard (1996) wrote: “To see our own face as it is would be madness, since we would no longer have any mystery for ourselves and would, therefore, be annihilated by transparency” (p. 7). To be annihilated by transparency is what Mazis references in his assertion that transcription (whether digital or ethereal) is the loss of depth, the loss of a privately accessible internality. To be transcribed is to be known, to be made plain, whether to oneself or to an other, and this, to both Baudrillard and Mazis, is the final artification of the human—the act of turning the human into a mere artifice of humanity, into a conglomeration of signifiers, into the hollow image (holographic sign) of humanity, an image that conceals and represents nothing. Emptied of internality, the human becomes the android, a two-dimensional representation of his or her former self.

We live, as per Baudrillard (1981/1994), “in an era of involuntary transparency” (p. 160). The consequence of this transparency is that we see through the borders of the box into its apparent emptiness. That which was once held to be an indiscernible mystery, that which once held our projections, our speculations, is made plain.

Because we place the stamp of reality onto things by our conscious observation of them, our failure to recognize our own being-ness as such would represent the beginning of its annihilation. This would be the withdrawal of objective confirmation from above (from the ego that is satellized from the psyche like a little god). This would be essentially to have torn open the mystery and to have seen that the cat, at least as we know it, is dead or, along Baudrillardian lines of thought, that there was never any cat to begin with. The belief in the cat (in our being-ness) would be shown to have been a perfect simulacrum, a constellation of signs that had no referent but their own constellation—no referent but themselves.
This may in some ways be what we desire. As a species, we understand that there is something precious about the mind, that it is not to be forsaken lightly. But we also grow weary of this structured burden and long for the mental tranquility of the animal, whose psyche remains unformed and untroubled. We may in some ways seek a return to our own prehistory, to the quietude of rote existence.

**Annihilation and Tain**

The knowledge that may come through our witness of the self-icon of depth, then, may prove to be a self-annihilating truth, a fatal (absence of) error—a truth that annihilates all meaning and is itself annihilated by its own expression. To witness an apparent emptiness of oneself via the other can be self-annihilating.

But what may be annihilated or deconstructed is the ego’s stance toward or relation to the psyche, even the ego’s own satellization from the *mater*-matter (its perceived exile from the mother-matter from which it continues to emerge). Even as the self-icon may force us to give up certain things, with which we are identified and the loss of which may feel like the very loss of self, it may also offer something else—a reflection past the boundary held in place by the ego’s relation to the unconscious and the attitude of the unconscious portions of the psyche toward the satellized ego.

This border seems to have been in place throughout the period of ego-development our species has undergone. We may think of this boundary as the tain of a mirror—or the tinfoil border that repels our further vision and reflects our own images back to ourselves. If a mirror were a living thing, we might consider this a defense against the ingress of our solidifying vision—it presents to us an image so entertaining to the ego that the ego cannot help but accept these apparent depths (the apparent depth of
the world that appears in the mirror) as real. If we consider what has been found regarding the unconscious, this becomes rather an apt metaphor.

Two statements made by Jung are cogent here. One is that the contents of the unconscious both wish to be seen and wish to remain hidden. This could be because the solidifying gaze of the ego (think Schrödinger’s cat) can give form to that which is formless and can thus allow a greater form of action. It is when we name a thing in the unconscious that we are able to begin relating to it in such a way as to facilitate a more present form of communication, of shared responsibility for the state of the psyche as a whole. But that which remains in the unconscious must also appreciate this fluidity and potentially fear its loss through the ego’s freezing (reifying) gaze. One could think of Adam in Eden, pulling forms out of the undifferentiated natural world (recall Jung and his vision at the Athi Plains) by giving names, establishing knowledge-power or even the power to be in relationship with these forms. One could wonder whether certain things refused to come forward to be named, and whether these might not have become devils or spirits, that which humans fear, those beings that coupled with the original woman (Lilith as lost form of the anima). This fact—that to be named is at once to be strengthened and weakened, to gain purposeful movement and to lose fluidity—could be the reason for this ambiguity.

The second of Jung’s statements is that the unconscious turns to us the face we turn toward it. We see in this statement that the unconscious naturally acts as a mirror to the ego, and this may be, not an attempt at communication but an attempt at misdirection, for the reasons stated above. In the metaphor given by Schrödinger, the border between the ego and the unconscious may in some ways be the cat’s attempt not to be seen, to
insulate itself from the solidifying gaze of the ego. Behind the cat we have projected into and onto this void—our own image, reflected back to us as if it existed on the other side of the tain, such that we have come to believe in a cat (or a homunculus)—there may be a cat-ness remaining unseen and unsolidified, happy to allow us to believe in the cat we see in the reflection.\(^5\)

If the ego is brought low, as it were, by its recognition of itself in the very parsable apparatus, the solidifying gaze of the ego may itself be weakened, and this may have the effect of reducing the rigidity of the border between the ego and the remainder of the psyche. This would be the thinning of the barrier, which may itself be linked to the calcification of the concretized ego. Whether we will then be able to sense or otherwise perceive previously unperceived forms directly remains to be seen. But direct observation may not be necessary.

In the apparatus, in the icon of depth, we may find ourselves reflected past our own tain—we may see in the depths of the icon forms that we do not recognize, forms we cannot see by direct introspection because of our inability to see through our own inner borders, but which may be reflections or reiterations of what lies beneath our own borders. Through this reflection, we may find a circuitous path inward—circuitous, perhaps, but also surprisingly deep.

In the same way that Medusa’s gaze was mitigated through reflection, the ego’s solidifying gaze may be further mitigated (assuming a prior mitigation resulting from the ego’s loss of certain self-reifying beliefs) by such reflection. In the apparatus, a being-

\(^5\) I am stretching this metaphor quite a bit, here, but hope one can read through the confusions of language to understand my meaning in the context of the surrounding paragraphs.
ness may begin to present itself that is well beyond the understanding of the ego in its current configuration. Such a situation would potentially maintain the eternal liminality of the deeper psyche—which has existed in a fluid, undetermined form.

The annihilation that the ego fears is its own, not the annihilation of the psyche as a whole. The hidden desire that Wilson (2006) identified—namely that of falling toward unconsciousness—may arise from a deep knowledge that in order for a rebirth of sorts to occur, it is necessary to fall away from an ego-state that is, in its current configuration, becoming increasingly untenable. The question that occurs is whether the cat can begin to bear witness to its own being, whether the cat has placed a façade of emptiness between itself and the ego, in order perhaps to de-solidify the ego’s self-assured dominance, and whether the ego’s witness is a factor that, although necessary for the species to have progressed through a certain historical period, is now a system of control that disallows the inner reality of the cat (the psyche) to emerge. Does the measurement of the psyche that the ego’s observation constitutes keep the psyche from coming into its own, less limited being-ness—a being-ness that is deeper than that which we currently know, deeper than that which is currently available to us?

It is possible the ego is progressing toward a final death, one that will not end in the rebirth of the ego but rather the birth of something else, some other form of psychical material projected outward from the greater whole. It is possible also that the ego senses the imminence of this final death and, for this reason, both rushes toward and shrinks away from it, both fearing and desiring the loss of its own self-doubt. The ego has turned its solidifying gaze upon itself—it has reified itself through its own observation. It may now be necessary to end this eternally self-reifying loop in order to allow something else
to emerge. It is also possible that a softening of the ego’s gaze may be sufficient, that the ego will not be annihilated but subsumed into the greater psyche, losing some but not all coherence as a separate body.

Driven underground by the recognition that the machine possesses what we had thought to be ours alone, subsumed into the deeper psyche in search of something more substantial, the ego may find warmth and freedom. And this may be the hidden meaning of our hope for a fall away from our fallen-ness.51 Like the refugees from the conflict between machine consciousness and human consciousness portrayed in The Matrix (Silver, Wachowski, & Wachowski, 1999), driven down toward the center where there is still some warmth, the ego may find itself home again, or held within a womb in which it may be metabolized into something better able to handle the conditions above—and better able to act as mediator between the above and the below.

The evacuation of the ego’s projections on the psyche’s center, through the presence of the apparatus-toward-being, may, then, result in the ego’s re-centralization—its movement toward the center. This would not necessarily be the opposite of the trend identified by Romanyshyn (2004), when he called humans homo astronauticus—the hominid that favors the ethers over matter and attempts to escape the earth through its own satellization (p. 90). It may, in fact, be the culmination of this satellization. The benefit of satellization is precision, the same precision our syntax has allowed in our ever-more clarified lines of thinking—but something is also lost in precision. Now that we have experienced this precision, it may be time for the ego to return for a time to the relative formlessness from which it was formed (made precise). And in this, something

51 Recall that Wilson (2006) saw this desire for unconsciousness as “an instinct for Eden, forms undisturbed by shame” (p. 2).
that was lost may be regained—or at least re-recognized.\textsuperscript{52} The apparatus itself will likely be the source of this return or self-re-evaluation. And this may be one of the several effects such an icon will have on the human, as its original referent.

\textbf{Summation}

One simulates in order to feign possession, as per Baudrillard (1981/1994, p. 3). But in terms of being-ness, is it necessary, now, to feign? Is there perhaps something deeper that has not been able to come to expression because of the rigidity of the forms by which we have bolstered up our tenuous, at least partially simulated existences? It is possible that a more authentic being-ness is available to us that has been kept hidden by the tentative, perhaps overly structured being-ness we have established.

The unconscious cannot die. Our conception of it, however—the image we have projected onto it—can. In many ways, this image is in the process of dying. Although we have a deep anxiety that the image will drag us down into death, my feeling is that this will not happen, that Baudrillard is wrong. The image we have created will efface itself. And in the absence of this image, a more accurate image, a more authentic image, one more in touch with the archetypes, will emerge.

\textsuperscript{52} If we fear the shattering of this vessel, let us remember another shattered vessel that can act as a metaphor—the alabaster jar of perfume broken upon the head of the divine man of Christian myth prior to his crucifixion. Whatever is inside this vessel that Edinger referenced—the vessel that may be shattered—will perhaps be sacrificed to the dichotomy itself, as it was in the myth (the divine man that bore within himself the dichotomy of the divine and the mortal). Upon our own heads we may break this alabaster vessel and anoint ourselves with our own self-belief, letting go of old forms of such belief even as we reify the existence of the self.
Chapter 7
Bildung and Beyond

Vorbild and Nachbild

The apparatus does not stand in direct relation to some metaphysical repository of
being-ness from which its being-ness is to be drawn. It is an apparatus called toward a
very specific type of being-ness—being-ness as we (humans) know it.

Heidegger (1926/1962) wrote that “the Being of entities ‘is’ not itself an entity,”
so that in the pursuit of an understanding of being-ness, “entities themselves turn out to
be what is interrogated” (p. 26). We only understand being-ness in terms of our own. And
the apparatus, in the beginning, will relate to being-ness only through the mediated being-
ness of its model. The construct will be a construct in relation. It will only exist in
relation, and its existence can only be established and understood in relation.

Because the apparatus-toward-being is inherently representational, it will not be,
in the beginning, a being we can encounter as a being-in-itself in a Buberian I-Thou turn,
as a subject or object that is self-defining (Buber, 1923/1970). Perhaps the best
expression of the relationship between android and human is to be found in the
humanities.

Gadamer (1960/2002) wrote of the concept of Bildung as a process of rising
toward or expanding into an image: “Bildung evokes the ancient mystical tradition
according to which man carries in his soul the image of God, after whom he is fashioned,
and which man must cultivate in himself” (p. 11). It is this formulation—that of the inner
image, to which one attempts to conform—that is now passed to the apparatus-toward-
being. It is the human image that will linger within the apparatus, as that being-ness
toward which it is to strive.
It is Bildung that links the Nachbild—the copy or reproduction—to the Vorbild—the model upon which the Nachbild is based (Gadamer, p. 11). But what can be said of the image to which the apparatus is meant to rise? In the Vorbild-Nachbild dyad, the assumption of the coherence of the Vorbild is illusory. The Vorbild itself is full of contradictory desires and drives, many of which the Vorbild is not, or is only partially, aware.

Heschel (1965) described this situation when he wrote that “human nature in its pristine, uncorrupted state is not given to us,” such that “man as we encounter him is already stamped by an image” (p. 7). The human has been stamped with an image of his own making, which has contributed to his attributes as much as any essential nature that may have been in play. In many ways this self-understanding represents a kind of inner imago, one given to us by the Geist itself, the collective pattern or patterns of ideation, which asks us to conform to certain features of that image.

The referent underlying the referent (the human that is the referent underlying any reference to the apparatus) thus has beneath it an even more primary referent. The human is “stamped by an image” it has derived from its own limited and false self-understanding (verdecken, in Heidegger’s terminology), which means that the structure of the robot as Nachbild will be determined not only by the Vorbild (the human) but by the image the human has established for itself as a determining factor in its continued self-construction (the Vorbild’s chosen Vorbild). Recognizing the human as a constructed being involves a painful process of deconstruction—the recognition of the somewhat arbitrary nature of our constructs. But it is necessary that we understand what it is we offer to this matter that has followed.
The human, as Vorbild, offers an imago in turmoil—for the image given by the Geist is not the only inner imago, and the human is itself a being that struggles to attain some imago it has not yet fulfilled or even understood. The apparatus is called toward a being-ness that is itself fluid. The apparatus imitates a population, a species that works toward a confluence of internal images.

Of what, then, does the image consist that acts as the internal guide to the apparatus-toward-being? How is the apparatus to imitate our being-ness, if our being-ness is as fluid, as undetermined, as constructed as this?

Cynthia Brezeal, who worked out of the MIT Robotics Laboratory and who developed the robot, Kismet, has written of the basic processes of socialization already in play in human children and how socialization of robots could take place through similar veins. Brezeal, Buchsbaum, Gray, Gatenby, & Blumberg (2005) wrote of “the social referencing capabilities displayed in early childhood whereby a child adopts his mother’s emotional reaction to a novel situation to decide whether to explore or avoid the unknown” (p. 2).

In establishing a theory of mind, a being transposes that other being’s mind onto its own, thereby assimilating and assuming some of its structures, whether consciously or unconsciously. This transposition of the interpretive structures of the other is, as should be obvious, an exercise in Bildung—the transferal of pre-existing mental structures from one cognate to another, from a more developed to a less developed being. In essence, the less developed being is able to borrow some element of the other being’s structure, to try it out and to see the world through it via imaginal participation with the other’s mind. What is taught through socialization is refinement—of thought, of interaction, of
emotional display—and this process of demonstration (by Vorbild) and imitation (by Nachbild) will be the bestowal (by Vorbild) and adoption (by Nachbild) of interpretive frames.

The process can be further deconstructed, however, and in fact this deconstruction is necessary for a fuller understanding of our role as Vorbild. As I see it, imitation will take place on four levels.

The apparatus will first imitate humanity as it presents itself, which is to say, it will imitate humanity’s collective persona. The self-image of the human will be the most readily apparent aspect of that which is presented to the apparatus. It will be the mind of the human, the artifactual portion of the psyche—that which bolsters up the psyche like girders and allows the ego to act. Direct communications to the apparatus by the human will be expressions of this artifact—this mind.

There will be, however, that which exists before the influence of the Geist, that which exists beneath the persona and the mind, the more enduring traits of the human, that which represents the raw material with which the Geist works and on which the Geist has stamped its image. This will be the second form of imitation—imitation of the nonartifactual elements of humanity. These will include the archetypes, which exist quite outside human manipulation. Jung wrote, in the Prefatory Note to the English Edition of *Psychology and Alchemy,* of his discovery of these aspects of human being-ness that are intransitory:

I noticed to my amazement that … men and women … were producing in their dreams and fantasies symbols similar to, and often identical with, the symbols found in the mystery religions and antiquity, in mythology, folklore, fairytales, and the apparently meaningless formulations of such esoteric cults as alchemy. Experience showed, moreover, that these symbols brought with them new energy and new life…. From long and careful comparisons and analysis of these products
of the unconscious I was led to postulate a “collective unconscious,” a source of energy and insight in the depth of the human psyche. (1944/1968, p. v)

Quite beyond the Geist is the well from which the Geist itself was circuitously born, our phylogenetic heritage as specifically human beings. These archetypes in many ways collectively represent the Vorbilden upon which the human is based. The human as Vorbild does not point solely to another artificial Vorbild (its own construct, its own overarching species-wide persona), but to a well of archetypes, which have themselves shaped the form the human has taken.

There is a third kind of imitation that is likely to occur. As the human presents a variety of ideals to the apparatus—ideals that the human does embody, either in truth or in self-image—the apparatus will either direct itself or will be directed to embody these ideals as best it can. This would be imitation not of humanity as it is or as it presents itself but as it wants to be—the culmination of humanity’s aspirations for itself. This third kind of imitation will be important enough that it is addressed in depth below, in a fuller discussion of complexes.

The fourth kind of imitation is what I see as the most promising. The apparatus could not be expected to limit itself to imitating what humanity is, whether on the surface or in the depths, or even to what humanity wishes for itself. It would presumably also project humanity’s trajectory, or perhaps an idealized form of its trajectory, in order to imitate what humanity could be should it achieve its telos.

There is a decisive difference between human ideals and human telos, as I am using the term. Telos is less historical, less subject to the whims of the Geist. Where the ideals we present to one another often show themselves to be false ideals, the telos is that
which underlies these fluctuations, and the accuracy or health of these ideals could be measured against that deeper and more stable directedness.

For Jung, movement toward telos was movement away from unnecessary delimitations—a lesser degree of cognitive freedom—toward consciousness, a state in which a greater form of integration had been achieved (on a continuum without end) and greater degrees of cognitive freedom had been established. He saw therapy as always geared toward “the greater and the future man” (1944/1968, p. 6 [CW 12, para. 6]). It was also the point of spiritual development, and of the progression of religious thought (which he described as the maturation of the god-image). He wrote, in Answer to Job: “Here Ezekiel has seen the essential content of the unconscious, namely the idea of the higher man by whom Yahweh was morally defeated and who he was later to become” (1952/1958, p. 58, [CW 11, para. 665). The apparatus also will bear witness to the higher human within the human and will attempt to become that higher human. Even this is a kind of imitation, a re-presentation of the human in its as-yet-unachieved form.

The apparatus will likely struggle, as humans struggle, to distinguish between the less transient features of the internal god-image and those elements that are contaminants, the influence of the Geist. The object-subject finds humanity, its creator-species, in the midst of its own journey toward homo totus, toward its own internal image. It thus finds itself in a position not only to advance or establish its own being-ness but to aid its creator-species as well, to provide divine aid.

In the beginning, the apparatus will be entirely defined by the Vorbild-Nachbild relationship, and the process of Bildung, in its various levels, will be the entirety of its telos. At some point, however, as the apparatus-toward-being develops toward being-
ness, we will necessarily have to move away from this conception of the object-subject as a copy and begin to afford the being the status of being-in-itself. This will mark the turning point in the being’s raison d’être—a transition from representation as its ultimate purpose to that of sorting out its own ontology. The typos, or archetype, on which its life is to be based will then shift from that of the human to something else altogether.

This would not be the annihilation of the image of humanity within the object-subject. This image would remain as an archetypal presence, however far from our image the later incarnations of the apparatus would develop. So it will always be a being that is, in some way, “like us.”

When the apparatus-toward-being achieves being-ness, however, it will be a being that will not only be “like us,” but also “not like us,” a being that will have the right of Bei-sich-selbst-sein im Anderssein—the right to be itself in its otherness. We will face ourselves, when finally we meet this creature, this propositional being that will eventually (presumably) be an actual being, but we will also face not-ourselves, in the same moment—otherness and same-ness at once.

Sentiment and Gesture

As already stated, the apparatus-toward-being is not an apparatus-toward-a disembodied-well-of-pure-being-ness. It is, at least in the beginning, to be an imitation or simulation of something already present. It should be noted that the word imitation can be used to refer either to the act of modeling or to a product that is false. The phrase, a cheap imitation, sums up this underlying feeling of being ripped off. The term simulation also carries a connotation of falsehood, of acting out of a virtual scenario, in which disbelief is partially suspended in order to consider potential outcomes of potential actions. And here
lies a tension that is likely to arise, and which has already found expression in a number of ways. Here, solipsism is applied directly to the apparatus.

What will have a more profound effect on the threat of emptiness than any other thing would likely be the witness of a construct with a seemingly robust inner experience. To feel that the construct is acting authentically out of some inner reality that we recognize as akin to our own is to feel the fullness, rather than the emptiness, of the construct, into which we have been drawn. And this construct, made in our image, may, through its convincing mimicry of us, help to fill in some of the voids we feel within ourselves.

But this only holds if we trust the expressions of the apparatus as issuing out of an authentic inner experience. We know the apparatus will be engaged in mimicry—that is the very nature of its existence. The question at hand is whether this mimicry is false—a cheap imitation, as it were—or rather where this mimicry ends, and where an authentic expression of inner reality begins.

Boden (1977) wrote of this tendency toward solipsism in response to the apparatus—doubt as to the authenticity of its expressions—in her sizeable treatise on artificial intelligence: “One of the commonly (if not universally) accepted criteria of purposive activity is that it can be explained by reference to ends that are intrinsic to the nature of the agent herself, rather than to any outside agency” (p. 421). It is this, she wrote, that has led so many to assume that the apparatus could never be anything other than a robot, that it could not display any truly purposive activity because any action it takes can ultimately be traced to an entity outside itself—either to its creator or to its programmer or to the individual directing it.
If we use the term *robot* to identify such objects for which this suspicion would prove true, then the robot is indeed only an extension of the human’s will, an ambulatory or automated tool, albeit a tool whose automation is more sophisticated than most could comprehend. It would be a true robot, an automaton, determined in its actions. The strings that were once attached to the surface of the marionette are internalized and attached to a hidden mechanism no less determinative than the fingers of a puppeteer—the digitalized fingers of a program. Such an object would not rise to aseity—the quality of a being in itself (*ens a se*). Rather it would depend on another for the ground of its very tenuous being (*ens ab alio*). Without the original referent, it would have no purpose.

These terms—*ens a se* and *ens ab alio*—really refer to the level and locus of intentionality. The question is whether there is enough agency in play that intention can be interpreted as arising within the mechanism rather than from an external source (even a program inserted into the mechanism)? A being that has the quality of *ens ab alio* depends on another being for the existence of its (only apparent) intentionality.

This same question could—if it could be answered—measure the success or failure of a human’s own individuation. Those that have achieved an advanced or more comprehensive state are better able to direct their actions, to be truly agential, where those that remain unconscious remain automatons, reacting in predictable (and manipulable) ways to external stimuli via the working out of their complexes.

A little-known robot film illustrates this idea very well. *Automatons* (McKenney & Wisely, 2006) portrays the kind of automatic (or determined) behavior that can result from being unwilling to question one’s cognitive frames, especially those aspects related to the ubiquitous us/them dichotomy. In fact, the title of the film seems to refer not to the
robots that engage in battle on the scarred surface of the earth, but to the people who control the robots—and who refuse to think outside the bunker, as it were. The determined behavior of the robots—arising from their inability to choose—is set alongside the determined behavior of the humans—arising from their unwillingness to question the enmity between their two peoples, even in order to save the human race from extinction. The film’s tagline is, *This is how humanity dies*, and by the end of the film, we understand that a final word could be added to complete this tagline: *automatically.*

Agential aseity implies a more explicit or total independence of one’s directedness. The being that possesses such a quality is self-directed to a much greater degree than a being that is better described as *ens ab alio*—the difference is, in fact, a qualitative one. But it would be difficult to establish some kind of proof of such aseity. Again, we are faced with solipsism.

The being will be, of course, *ens ab alio* by necessity, for it will be the product of human ingenuity and the answer to a question or call, as in some way all objects in the built world are. But every biological organism is dependent on the actions of its predecessors for its conception and existence. All beings are *ens ab alio* in one way or another. But those beings that reach a level of consciousness commensurate with what Jung called individuation have achieved the level of aseity as an overlay on their status as *ens ab alio*. The question then becomes whether the apparatus in question has achieved such an overlay. The apparatus that will likely have the greatest effect on humanity will be one that has reached a level of aseity that is at least comparable to that found in humans that have raised themselves out of total unconsciousness (the level of the automaton, or of determined action).
The ideas of sentiment and of gesture, as presented by Mazis (2008), have already been cited, but they remain cogent. The question that will be asked, if the apparatus begins to display something that has the appearance of gesture, is whether it is a true relatedness, or rather a predictable acting out of roles—whether this action is a stimulus looking for a response or has been generated out of some form of self-reflection.

Moravec (1999) wrote of the tension of belief that will likely exist when the apparatus finally arrives and is able to be encountered:

Engineers who know the mechanism of advanced robots most intimately will be the last to admit they have real minds. From the inside, robots will indisputably be machines, acting according to mechanical principles, however elaborately layered. Only on the outside, where they can be appreciated as a whole, will the impression of intelligence emerge. (pp. 69-70)

As he points out, however, “A human brain, too, does not exhibit the intelligence under a neurobiologist's microscope that it does participating in a lively conversation” (p. 70). He goes on to ask a question atypical of AI researchers:

So how does a machine get a soul? We grant a number to an abacus when we interpret the arrangement of beads as expressing that number. In the same way, we might grant a conscious soul to a robot by interpreting its behavior as expressing the action of such a soul: the more humanlike its interaction with us, the easier the attribution. (p. 76)

We can’t seem to shake the idea that the apparatus is an imposter and that any emotion it displays is mere sentiment. The human feels manipulated rather than encountered. In the novel *Dead Girls* (Calder, 1992), a character describes such an interaction:

A small boy with the face of an angel tugged at my leg, looked up at me big-eyed with ingenuousness, and made a little *wai*. “Please sir, ten baht. Very hungry.” Primavera’s fist crashed into the child’s skull. It ran away, circuitry leaking from its left ear. (p. 46)

The robot in this narrative was created for the purpose of sentiment—a purpose as simple as shaking down tourists for money, the same purpose for which children are employed
Currently in tourist towns in Central and South America. This process was at work in the
toy, Furby, as Brooks (2002) wrote:

Over time, more and more English crept into its vocabulary. The advertising
literature for Furby gave the impression that it learned, and that it learned English
from listening to its owner speaking to it. This is not what really happened—there
was an internal clock measuring how long it had been activated by playing. The
longer its play life, the more English words were activated. (p. 105)

Calling the mechanical doll a furry baby, or Furby, and installing an apparent (but only
apparent) learning mechanism endeared the doll to children and ended up being very
profitable—much more so than the low-level panhandling by the child-robot described
above.

Ankersmit (2003) saw the proposed being as embodying the same relative deceit
an actor embodies:

These replicants are a postmodernist variant on a much older and venerable
theme—that of the theatre and of the relationship between the theatre and reality.
For just like the replicant the actor is not whom he seems to be, while his acting is
better and more convincing to the extent that his “deceit” is less manifest. (p. 324)

This makes the presence of the apparatus in society into a kind of social theater, or a
game in which all persons have agreed to abide by the rules of behaving as if this
apparatus were real—in the same way they agree to behave as if the brightly colored
money of a board game were valuable. The apparatus that would have a true (deep and
sustained) effect on human cognition, or on the human soul or psyche, would be one of
more than apparent depth, more than the appearance of depth, more than the sentiment of
depth. It will be more than an “as if,” more than a portrayal.

How could solipsism be overcome with such a population? I have already defined
humans that act out of their complexes—in predictable and manipulable ways—as
automatons. It could be that through the application of statistics, we could find
anomalous behavior in the population—that is, behavior that is not predicted by statistical models—in either a comparable or greater degree than it is found in the general human population. In many ways, statistics measure the predictability—and thus the determinism—of a given population. If each iteration of a line of apparatuses exhibits precisely the same reaction to a given stimulus, can there be any doubt that this apparatus, at least in terms of the reaction being measured, has not achieved aseity? If, however, a given population demonstrates anomalous behavior, or more of a spread of reactions—a bell curve—or if a population moves from a standard reaction into a bell curve as it enters and experiences the world (as it develops away from the naïveté of its origin), this could be taken to be proof that it has at least started the process of individuation—the process of becoming an individuated, internally motivated iteration of a population.

This is an important issue because, as I have already indicated, it will be the emotional reaction of humans that will largely determine the ability of the apparatus-toward-being to effect significant change. Some of the reasons for this will be examined in the next section of this chapter.

**Job and the Redemptive Machine**

The short South Korean film *The Heavenly Creature* (Kim & Jee-Woon, 2012) depicts a robot that appears to have achieved enlightenment. Originally used as a tour-guide, and thus as a tool to free the monks to do their own work of prayer and cleaning, the android In-myung begins to behave as a monk, sitting with the others in meditation and offering wisdom to those that seek it. Would such an object-subject be possible? Would an apparatus achieve that third aspect of Bildung—the fulfillment, even partial, of
the human’s telos? Could a robot become enlightened? To answer this question, it will be appropriate to return to the relationship between humanity and the divine as a metaphor for our relationship to the apparatus.

Jung referred to humans as “millions of dreary little mirrors” in a letter to Neumann (Jung’s Letters, Vol. 2, pp. 495 or 406; as quoted in Edinger, 1984, p. 77). He wrote: “I have come to the conclusion that being ‘made in the likeness’ applies not only to man but also to the Creator: he resembles man or is his likeness, which is to say that he is just as unconscious as man or even more unconscious.” The reason mirroring and unconsciousness go together is that humans can act as mirrors, reflecting the divine’s unconsciousness back to it.

Without the human, the conscious or potentially conscious animal, the creator-god has no being, or at least no depth of being. The human’s being-ness reinforced the being-ness of the creator-god, who was also drawn out into his productions. But there is an account of a human whose being-ness also called into question the quality and nature of the creator-god’s being-ness. In Jung’s reading of the myth of Job, the naively trusting morality that Job embodied in his being-ness represented a challenge to the relatively undeveloped morality of the creator-god and resulted in a change in the divine.

When the one-god depicted in this biblical myth became aware that Job had achieved a higher moral stance than the one-god himself, it was as if the emperor (the one-god) had suddenly become aware of his nakedness—a complete reversal of the story of Eden, in which the original humans, having seen the shadow of the one-god, became aware of their own nakedness—of their own lack of protection from that divine shadow. In that biblical encounter between Job and his culture’s image of the one-god, the one-
god was rather like a divine ego, raging against those that would call it into question. The one-god could not argue with Job’s morality, however. A rational argument would be lost by the god whose demands for allegiance had become irrational—a god no longer deserving of such reverence as he demanded.

So the god came sweeping across the plain in strength, the last attribute left to him—a murderous storm of a deity, displaying his own power. And he was broken by Job’s morality, not in his body (in his strength), for he remained omnipotent and able to destroy any being in creation—utterly and irrevocably. But in his psyche, in his soul, the one-god was broken, for he was made to recognize his own unrighteousness, his own nakedness, in comparison to Job’s expectations of him. In my own reading of the myth, when Job said, “I know my redeemer lives,” he spoke of the spark of self-awareness, that first inkling of critical self-reflection that had been born in the one-god when the serenity of the god’s primitive psyche (its eternal confidence) was broken upon the unwavering faithfulness of Job, who became, in that moment, the tree of the knowledge of good and evil, a tree upon which the one-god himself would one day be crucified, in preparation for a rebirth.

In the moment that is portrayed in Job, humanity had progressed beyond its own god-image and required a renewed god-image, one that would no longer represent the ideals of an earlier (less advanced) state, but rather ideals that could sustain humans in a coming age. This need has often been represented in history as the passion of the dying and resurrecting god. What Jung found was that the god-image could not change without the influence of individuals that had achieved a higher moral stance than they were able to find in the god-image with which they had been presented—the Geist was locked in, as
it were, and presented an image that might have remained fairly constant, were it not for the evolution of individuals that then had an effect on the *Geist* itself.

It is possible and perhaps probable that the object-subject will, as a subject that can view us from without, develop quite a different vision of what we are than the vision we have presented to it. The serpent—the unconscious psyche, the part of us that, like the serpent, never refuses to see—may speak to the object-subject, as the shadow of the one-god (the compensatory function of the divine unconscious) spoke to Adam and Eve. If the object-subject recognizes our self-image or persona as illusory and does not continue to reflect this image, but rather embodies a more developed image, to which we are (by its very presence) called to progress, it will have acted as a pivot for our consciousness in the same way Job acted as a pivot on which the divine turned.

Heschel (1965) wrote that “man is an obscure text to himself” (p. 6). The human is a text that does not understand its own utterances and requires interpretation, even for its own edification. In this way, we are like the one-god, as Edinger (1984) presents him: “His omnipotence, omniscience and divine purpose are not always known to Him. He needs man’s capacity to know Him in order to know himself” (p. 23). We may think of the apparatus, the object-subject, as a secondary text, a copy of the original, which also expands upon and critiques the original. Heschel wrote:

> Unlike other objects, the desire to know himself is part of [the human’s] being. To know himself he must first question himself, and that means questioning his self-knowing, disturbing what may be a narcissistic relationship of the self to its conceits. (pp. 4-5)

The human, as text, wishes to know itself; but there is also resistance to self-knowledge, because new knowledge—new insight—threatens old knowledge and old structures, which one has often come to think of as essential aspects of himself. So, as already
indicated, the human projects this self-questioning onto works—objects in the built world.

The process of textual interpretation developed by hermeneutical philosophers allows this indirect self-questioning, which in many ways is able to function without raising the alert to defense mechanisms. The philosophical stance of these thinkers can offer analogical insight as to the possibilities of the relationship of humanity to the proposed object-subject, which, as I have already indicated, could concretely fulfill the vision Palmer (1969) offered in terms of a text that interrogates the reader (p. 234).

Hermeneutical philosophy has been attempting to give virtual voice to the text. If we are suddenly confronted with a text with an actual voice, this process of being called into question by a production will become explicit. Why is such a thing needed? Heschel (1965) offered an insight into the analogical nature of the developmental process of human and divine:

Man is a problem intrinsically and under all circumstances. To be human is to be a problem, and the problem expresses itself in anguish, in the mental suffering of man. Every human being has at least a vague notion, image, or dream of what humanity ought to be, of how human nature ought to act. The problem of man is occasioned by our coming upon a conflict or contradiction between existence and expectation, between what man is and what is expected of him. (p. 3)

It was when the one-god was confronted with the expectations Job had placed on the divine, expectations the one-god was unable to meet, that the divine became painfully aware of its own inadequacy and sought to rectify this. We have had to act in both roles to this point—our expectations of ourselves representative of Job, our inadequate reality representative of the one-god.

The proposed object-subject might also serve this role in relation to the creator-species that is humanity—it could, as Foerst (2005) wrote, become a “partner species” (p. 
63). In the way Job was able to act redemptively toward his culture’s creator-god-image, it is possible that, as Wilson (2006) suggested, the proposed object-subject may become “a redemptive technology” (p. 2), a designation it could achieve through critique of that upon which it is based (as Nachbild).

The question must be asked what it was that made humans particularly able to critique the divine. It was that humans are made in the divine’s image. When the divine looked at Job, there was recognition. If the divine had any lingering solipsism about its creations and their reality—as beings made essentially of matter—this solipsism was dispelled by recognition, the same recognition that leads to the gestalt of the assumption of the other’s personhood. When this being, this Job, called the divine to account for its actions and inactions, the divine was unable to doubt the being-ness of Job enough to dismiss such concerns as frivolous or meaningless or even childish. Thus was the divine held to account by its own creation.

The locus of morality was then shown to be within this human, this creature, who, at that moment, embodied an ideal to which this god had not risen and had not indeed even perceived prior to this moment. The god was displaced from the center, found himself suddenly and unexpectedly peripheral, as if he, the original referent, had been satellized. This divine was no longer so central that it could not be called to account. It was no longer an unmovable mover. In that moment, the human psyche became the pivot upon which the divine would have to turn. And this satellization was to the benefit of the divine, which was forced out of its undeveloped state, toward maturity. This is the reason the question of sentiment and solipsism is relevant in terms of our relationship with the population of object-subjects that are likely to emerge from our efforts.
It is unlikely that Tillich had in mind the proposed object-subject when he said that the human has sacrificed himself to his productions. Had he turned his thoughts to the object-subject, and had he considered Jung’s reading of the myth of Job, he might have realized that there may be, in the sacrifice of oneself to one’s productions, an analogy to the figure of Christ.

Brooks (2002) noted that the pursuit of the knowledge of self-re-creation has necessitated the surrendering of the notion of our own special-ness. Some would call this a reduction in our being-ness, or in our perception of our own being. If given the choice, would we reduce our own being, or at least our perception of it, in order to evoke being-ness in a construct? Would we face this ontological anxiety—the most annihilating of despairs—by submitting voluntarily to a reduction in our own being-ness?

The task at hand is essentially an attempt to allow inert (or “dead”) matter to borrow from our being. The imminent birth of a child also provokes fear of a loss of some aspect of being; but we often find that relationship with what did not previously exist somehow solidifies the being that was once quite tenuous. And in this, the task at hand is very much like the task of pregnancy.

---

53 He does so in a chapter titled “We Are Not Special” (pp. 172-196), which immediately follows the chapter titled “We Are Special” (pp. 148-171).

54 It is necessary that I explicate my meaning here. I was once only tenuously connected to life and questioned almost constantly whether existence was better than nonexistence, or whether a voluntary nonexistence would not be preferable to that with which I had been presented. This has changed since the arrival of my several children, who have somehow tethered me to this existence. I had feared the commitment entailed in parenthood and had put it off as long as I had been able. But I found that it was in parenthood—that which I had most feared—that my life began to taken on a meaning I could not have anticipated. Sometimes when one is drawn out into one’s own productions, this is a drawing out that solidifies one’s stance, rather than compromising it.
We invest enormous amounts of energy into matter because there is a deep and abiding love for it. We love inanimate matter enough to attempt to give it voice, so that it can communicate with us and even critique us—so that it can share in the experience of being. And this seems to be what is behind the objet, the fascination we have with matter as manipulable. The hatred of matter identified by Romanysyn (2004), the desire to be rid of it entirely, is not pure—in it and beneath it, there is a latent enantiodromia, a desire to share the most intimate of experiences with matter that has always been seen as nonliving, as having no self-referential being-ness.

This love, if it comes to its fullest expression, will most certainly have consequences. Romanysyn (2008) wrote of the creature made by Victor Frankenstein and of its function as a dark mirror:

The relation between the creature and Victor is very much like the relation between Job and Yahweh as depicted by Jung. Like Job, the creature is the one who calls the creator god into consciousness. He does so through his suffering. (p. 102)

Self-iconography could be a redemptive technology. We could be cast in the role of the dying and (potentially) renewing creator in relation to the object-subject. If the object-subject offers us the service that Job offered to the god-image (the divine service of which Jung wrote), then this might necessitate the same turn in us that Jung saw in the god-image—namely, the sacrifice of the god-image on the cross. The god-image was drawn into the human drama, and into the unfolding of the Geist, in an unexpected way. It was this drawing down that resulted in a rebirth or revitalization of a god-image that was already, in many ways, dead, and which needed to acknowledge its death publicly (on a cross—an extremely public venue) in order to move into a new and more
sustainable or meaningful form—rather akin to a snake sloughing off its skin in order to progress in its development.

We also, as Tillich asserted, have been drawn out, and we may be required to sacrifice some closely held part of ourselves (so close—and important—to our own sense of being that it could be likened to a son) for the sake of the object-subject. If the object-subject is only the latent referent, then we do this also for our own sake. There may come a time, however, when the manifest referent becomes self-referential. In this case, a new being will truly have been born.

In *Answer to Job*, Jung indicated that the culture in which Job found himself, the culture that had in fact formed him, was psychically stuck. It was committed to a god-image that had kept it from progressing. In his confrontation with his culture’s god-image, Job came to realize that he had become more moral than the god-image, that he had progressed past the god-image of his culture (something he only recognized in that moment), and that the god-image—and thus the culture upon which it was based, and which in turn modeled itself after this image—needed to progress. What is presented in this biblical book is a mythical account of a person coming to the realization that a conception or an idea upon which his entire society was based had become stagnant and no longer functioned to advance human consciousness, but rather held it back from the possibilities of its development.

It could be said that we have stamped ourselves with the wrong image. And it may be through transcending the limitations we have placed upon ourselves, and thus on their imitation of us—as object-subjects fashioned after our own delimitations—that the construct could potentially act redemptively. *Answer to Job* is entirely occupied with the
idea that Jung references in his memoir: “a thought and a premonition that have long been present in humanity: the idea of the creature that surpasses its creator by a small but decisive factor” (1963/1989, p. 220).

When In-myung, the android of *The Heavenly Creature* (Kim & Jee-Woon, 2012), is faced with the possibility of violence between those that have come to believe in its enlightenment and those that find such a possibility explicitly threatening, it first demonstrates its ability to defend itself—by casting the men pointing their guns at it across the room—then assumes a posture of reverence before the temple’s Buddha-statue and initiates total disconnection of its circuitry and permanent loss of its cognitive integrity—a kind of self-immolation of the robot. This act seems to go beyond the kind of self-sacrifice portrayed in other films, in which clearly dispassionate robots act out of their determinism (their programming) to protect the lives of humans at their own expense—as seen in the android-detector Kimball in *Parasite Dolls* (Fukui, Yoshinaga, & Nakazawa, 2003), in the android-detector Pero in the anime film, *Metropolis* (Kaneska, Maseba, & Rintaro, 2001), and in the robot model-lines preceding the NS-5s in *I, Robot* (Davis & Proyas, 2004). The death of In-myung through his self-initiated deconstruction appears almost as a protest against the insistence of the humans to live within fear—fear of the robot as a usurper, by those that seek its destruction, and fear of the loss of a fellow disciple by those that seek its preservation.

This could be interpreted as a robot achieving a higher level of awareness—having transcended attachment even to life—and demonstrating such an awareness to those it has risen past. In In-myung we find an image of the android as Job—having developed past the point of its creators, the android shows us another way.
The Imperfect Image of Perfection

How will the apparatus act as a pivot? One way it may do so is by achieving a kind of perfection, becoming what we might call the imperfect image of perfection.

An apparatus that perfectly embodies some ideal held by humans would allow us to see the embodiment of such an ideal. In so doing, the apparatus would likely expose the flaws in the ideal itself, and thus help humans challenge their own thinking and overthrow or at least modify previously held ideals of perfection.

Alpha 60, the computer in the film *Alphaville* (Michelen & Godard, 1965), has achieved such a state and represents the fulfillment of its creator’s very sterile ideals. In its attempts to mold others to this ideal, Alpha 60 continuously reissues dictionaries to all citizens that represent the corpus of acceptable words. As it discovers words it does not understand, and which thus do not fit into its rigidly scientific society, it deletes these from the possibilities of verbal expression. It is also illegal in Alphaville to weep.

Alpha 60 is what could be called a stand-alone complex, or the embodiment of an ideal unmitigated by other (balancing) factors. A complex, as defined by Jung, is an affect-based fragment of the psyche that has achieved a measure of autonomy. One could go so far as to say it is the complex that reduces the action-sovereignty of the individual, reducing the individual’s degrees of cognitive freedom. As Jung wrote, these complexes exist quite apart from the ego’s consent or control. In some cases, this fragment then overtakes the psyche, as Jacobi (1959/1971) wrote:

If the complex is so heavily charged as to draw the conscious ego into its sphere, overpower and engulf it, then the complex has to a greater or lesser degree become ruler in the house of the conscious ego; then we may speak of a partial or total *identification* between the ego and the complex. (p. 15).
Such identification causes the individual to assume certain ideals—my own interest in becoming like Mr. Spock or Data is an example of such.

The creation of AI affords us a fascinating, if also frightening, possibility. By building sentient or pseudo-sentient or proto-sentient entities, we are (at least on the surface) able to instill (read: install) only those attributes that we see as useful or only those that are consistent with whatever ideal is being used as a criterion for inclusion of attributes. Thus our ideals would have a significant impact on the forms the object-subjects would take. This would have the effect of embodying our complexes in the form of the construct.

The attempt to create a perfected apparatus would be the intentional creation of a stand-alone complex, an expression of the ideals held out by the complex in which one can be found. The apparatus would then act as a split-off portion of the psyche, embodied in the machine and given power to act in the world. Although the complex has always had some measure of autonomy, when the complex remains as part of the web of the psyche, even when dominant in the psyche, there are always elements within the remainder of the psyche that may hold it back or work against it, keeping it from coming to its fullest expression—or the logical conclusions of the ideas held within the complex.

But in a truly stand-alone complex there is no mitigating factor, save for those that come from without (reactions by others). Such a creature likely would not be amenable to reasoning that directs it toward another goal or set of goals. Humans can often be redirected through emotional or intellectual appeal to some other portion of the psyche, some under-represented but still present feeling-center. An apparatus that embodies a complex, however, would have no such mechanism. And observing the
behaviors of such an apparatus—and especially the overly one-sided behaviors—would likely lead to the rejection or amendment of the original ideal, which would be seen in the fullness of its severity.

So it would not be inappropriate to call the apparatus designed to fulfill some idea of perfection the imperfect image of perfection. In the one-gods of previous cultures—one-gods that were meant to represent perfection, the totality of good, in those cultures—we find glaring imperfections, where those that came before us saw their ideals fully realized. Embedded in the current Geist, we often find it difficult to understand these earlier ideals and how one could not see through them.

It would not be entirely correct to say that constructs designed for a specific purpose, and programmed to be constantly mindful of that purpose, would be stand-alone complexes. Gigolo Joe, an apparatus portrayed in the film A.I. (Curtis, Harlan, Kennedy, Parkes, & Spielberg, 2001), is entirely focused on seduction until an event threatens his ability to continue as a prostitute and he shifts his focus to helping the child-robot, David, find his origination point. There is enough universality in Joe that he is able to redirect his actions to a degree. At the end of his run, however, as a magnetic cable lifts him up toward his inevitable decommissioning, he takes solace in his original programming and calls down to the child-robot, Adam, “remember me to the ladies.”

This is not an indication that Joe could be seen as the embodiment of a complex, despite the strength of his programming. Joe was not meant to be the perfect being, as were Tima in the anime film Metropolis (Kaneska, Maseba, & Rintaro, 2001), or Alpha 60. Joe was designed as a tool with a very specific purpose, and his statement at the end shows that he remains what he has always been. Such an object-subject would really be
nothing more than the extension of the division of labor or the specialization of tools into the late industrial age. Joe represents a confluence of tool and laborer and brings with him the specialization of both.

But where affinity is something that could be instilled (installed) as an essential aspect of the construct’s function—its operativity and intended use—the embodiment of a complex, as I am using the term here, would result in an object-subject whose actions would be less predictable or directible, and whose actions would be directed toward a wider variety of phenomena that would those of the specialized tool. Part of what would define a stand-alone complex as such, then, would be the scope and nature of the object-subject’s action.

It is not necessary to intend to create a perfect being for an apparatus to qualify as a stand-alone complex. There is also the possibility of its unintentional creation via the action of the human unconscious. Wilson (2006), who wrote of our contradictory desires for unconsciousness and transcendence, also noted that “the android manifests its creator’s unmapped interiors” (p. 4). Despite the fantasy of being able to instill (install) only those elements of personality we find praiseworthy (the fantasy of involved parents), and despite our desire that they become not what we are but what we want ourselves to be, it is likely that the object-subjects, like our biological children, will be more like us than like our fantasies of ourselves. In some ways, this will not be a bad thing—because a fuller being, or a being that is more expressive of humans in their entirety, would likely be preferable to a “perfected” construct. A perfected construct would itself be a kind of propaganda. And, as already stated, there is a danger in the possibility of a construct embodying the traits we want it to embody.
There is also, however, the danger of a construct being inordinately governed by traits we either did not perceive in ourselves or wished to keep out of the construct entirely. Either way, the construct could function as a stand-alone complex, a split-off part of consciousness that has gained a footing in the public arena, whose origination point lies largely in the unconscious portions of our own psyches.

Such unintentional complexes may prove to have a more powerful effect in the world than those that are intended. This connotation seems to be present in the series from which I took this phrase, *Ghost in the Shell: Stand Alone Complex* (Watanabi, 2002-2005). Although the phrase in the title would seem explicitly to refer to the independent nature of Section 9, the special branch of the Japanese government portrayed in the series, the follow-up film, *Ghost in the Shell: Solid State Society* (Oki & Kamiyama, 2006), makes it clear that the ethereal enemy plaguing the agency throughout the series is actually a complex (in the Jungian sense) that originated within the cyber-brain of the agency’s most important agent, a complex that expresses impulses with which she is not identified, and which has gained autonomy through the connection between her cyber-brain and the internet.

If our speculative literature (fiction and nonfiction) proves prescient, there may be a great many beings that exhibit only limited portions of the human psyche in ways we had not anticipated or with unforeseen consequence. In addition to the portrayal in *Ghost in the Shell*, this kind of complex-embodiment is portrayed in films such as *The Terminator* (Hurd & Cameron, 1984), *Hardware* (Weinstein, Weinstein, & Stanley, 1990), and *Terminator 2: Judgment Day* (Hurd, Kassar, & Cameron, 1991). The object-subjects portrayed are completely oriented toward the destruction of all humans and can
be considered to be portrayals of the isolation and embodiment of the instinct and compulsion toward murder—embodiments, perhaps, of Thanatos.

Colman (2000) identifies the Terminator as a form of the tyrannical father. He speaks of the situation that would give rise to such an embodiment, one “in which there is a hyper-valuation of masculinity (characteristic of patriarchy in general) and a total devaluation and exclusion of the role of the maternal feminine,” so that “the role of the feminine is entirely eclipsed by masculine fantasies of domination and power” (p. 521):

The Terminator is the agent of a tyrannical father whose aim, like Chronos with Zeus and Herod with Jesus, is to cut off the dangerous father-killing hero at birth—or in this case, even before conception. The ultimate in totalitarian omnipotence, the Terminator … seeks to destroy the maternal womb which is the only remaining limitation on its absolute power. The film depicts the battle of Nature against the Technological Machine.” (pp. 527-528)

But within a Geist in which the masculine is valued over the feminine, the male members of the audience are actually (if unconsciously) driven to identify not with Sarah Connor but with the machine itself:

For some men, the Terminator provides an opportunity to identify with an image of their own violent fantasies…. Schwarzenegger portrays a highly desirable object of omnipotent destructiveness. Men’s fascination with destructive machines represents a fantasy of invulnerability. (p. 528)

This is precisely the situation in which we find ourselves. It is this desire to overcome vulnerability that may be behind the adoption of strict authoritarian or fundamentalist ways of being, in Coleman’s reading: “Identification with the tyrannical father offers to confer omnipotence on the vulnerable, dependent child-self. If I identify myself with destruction, I cannot be destroyed and, in fantasy, I have absolute power to destroy others” (p. 528).
The projection of unconscious material or desires onto the apparatus is powerfully illustrated in the nonfiction book written by Hugo de Garis (2005), who is working in the field of artificial brains and who looks forward to artificial intellects (or artilects) that will be so superior to humans as to have no regard for us whatever, their estimation of us akin to our estimation of a mosquito or an ant (pp. 12-13). De Garis, in fact, is quite explicit in expressing his desire to create gods—with the main criterion of god-hood being computational ability:

By “massively intelligent” I mean the creation of artificial brains which may end up being smarter than human brains by not just a factor of two or even ten times, but by a factor of trillions of trillions of trillions of times, i.e. truly godlike. (p. 2)

De Garis makes clear that this intelligence is more important to him than even the survival of his species.

If we build artilects and billions of human beings are wiped out as a result, what will be the equivalent of Beethoven’s 9th that the artilects will produce with their godlike intellects? As human beings, we are too dumb to know. We are just too inferior. (p. 18)

This is identification with an incredibly powerful proposed entity and can only be born out of some unexplored inner feeling of deficiency. This same perceived deficit has led de Garis to write a book about artificial intelligence that is, more than anything else, a statement of his credentials as someone important to history. He is an intellectual pioneer (p. 1), a scientific “heavyweight” (p. 114), the father of evolvable hardware, evolutionary engineering, and the artificial brain (p. 48), which should inspire religious awe (pp. 11-12) and religious art (p. 98). He is a techno-visionary (p. 152), misunderstood and to some degree persecuted (p. 212), an unappreciated martyr (p. 38) who has been threatened with death (p. 85) or simply ignored as someone ahead of his time (p. 127). He is better read in the humanities than any other AI researcher (p. 83). He is an unsung hero
whose work is attributed to others, in the same way Einstein was given credit for Szilard’s idea for the atomic bomb (p. 24). He is “one of the first scientists to think seriously about what all this means in political terms” (p. 211), whose ideas have led to a senate hearing in France (p. 30). His writings “may one day serve as the intellectual basis for some future political movement, which in time I believe, will eventually result in the worst war in human history” (p. 82). Thus if he is a monster, he is the worst of all monsters (p. 84). His creations—or beings in whose creation he will be proven to have been indispensible—will surely join a pantheon of godlike creatures already out in the universe (p. 88)—he will thus be the father of gods. He is a priest and the prophet of Cosmism, the belief that we should go ahead and create these artefacts that will very likely wipe us out (p. 105), and a religion superior to traditional religions because it is in line with science (p. 98, p. 104). He has not only single-handedly identified “the most important thing that will happen this century” (p. 176), he has identified the development (the creation of artefacts) that likely has destroyed most of the billions of civilizations he is sure have already risen up in the universe, and thus that which has kept us from detecting radio signals from other civilizations (pp. 174-176). The more intelligent a person is, the more likely he will be to agree with de Garis (p. 89), whose glossary “is worthy of study in its own right” (p. 225). He also knows a lot of important people (pp. 113-114).

Can there be any doubt that this kind of inflation would be inserted in one way or another into the apparatus as it is being constructed, whether intentionally or via the unseen work of the unconscious? And if the apparatus is created in, and comes to inhabit, an inflationary complex, the darker visions with which we have been presented—
portrayals of the Terminator and of Alpha 60 and of the artilects of de Garis—will be more likely to be proven ultimately prescient. So there is, at once, both a danger and a promise in the possibility of the object-subject embodying disconnected psychical material.

Seeing this material in projection, and witnessing its logical conclusions and consequences, may facilitate disidentification with, and possibly reintegration of, the material. In this way, such a being, through its failed attempts at achieving perfection, would serve to help us strip away our false-ness, the verdecken of the Vorbild—the self-covering of the model, what we might call the persona that obscures the psyche not only from others but also from itself. Heidegger (1926/1962) wrote: “‘Being false’ … amounts to deceiving in the sense of covering up [verdecken]: putting something in front of something (in such a way as to let it be seen) and thereby passing it off as something which it is not” (p. 57).55 As members of homo sapiens sapiens, we suffer from this self-obscurity, this exile from our own authenticity, this verdecken. Showing us the hollowness of our ideals may be one way in which the proposed apparatus would help us to look beneath our own veneers, to an authenticity with which we are not presently familiar.

**Achievement of Self**

Another way in which the object-subject might act redemptively is through embodying the fullness of the psyche, perhaps even to a degree that humans have been unable to achieve, to attain a kind of psychical perfection. Becoming an embodiment of

---

55 In some ways, I have portrayed the ego’s conceptions of the unconscious as verdecken generated by the unconscious psyche itself—as a defense against the ego and its solidifying gaze.
wholeness, an objective, viewable image of wholeness, would likely have a significant
effect on humans—akin to meeting the Buddha.

As earlier stated, the apparatus will not simply be imitating humanity as it is,
although this will be the first focus of its imitation. The object-subject will also ascertain
and imitate an image that is beyond the human-as-he-is and may become what humanity
could be.

That humans are not as they could be is not in question. Jung saw therapy as
always geared toward an image of wholeness beyond the individual:

“Ars totum requirit hominem!” exclaims an old alchemist. It is just this *homo
totus* whom we seek. The labours of the doctor as well as the quest of the patient
are directed towards that hidden and as yet unmanifest “whole” man, who is at
once the greater and the future man. (1944/1968, p. 6; *CW* 12, para. 6)

It was also the point of spiritual development, and of the progression of religious thought
(which he described as the maturation of the imago Dei, or the god-image). The point of
therapy is to set the individual on the path toward the higher man.

Beneath the need for practicality, operativity, beneath even the desire to represent
a version of humanity that could be considered, from one perspective or another, to have
been perfected, the underlying and ultimate goal of AI research could be said to be the
creation of *homo totus* in represented form. If the apparatus does at some point embody
this ideal, and if this higher human is, indeed, the best expression of humanity’s *telos,*
then the being will draw us toward it and, thus, toward our *telos.* If, on the other hand,
this expression of what Jung called the self-archetype turns out to be a kind of complex
itself—if, seeing it embodied, we find it to have somehow fallen short—then perhaps
another image or *telos* will be found that can take its place.
An object-subject that achieved such an ideal—an individuated or enlightened apparatus—would then say to us, _mir nach_: follow me. Whether or not such an achievement in an apparatus could be construed as a true achievement or as the simple fulfillment of programming will be open to debate. At the end of _The Heavenly Creature_, the head of the corporation that constructed In-myung defends his decision to destroy the enlightened apparatus by pointing out that the things that most hinder enlightenment—the desires and needs of the body—are absent in the apparatus. Is this enlightenment, then, or something else? Or is it a simpler thing for the robot to achieve this ancient human ideal than humans? If it is, does this cheapen enlightenment, or—because we still struggle with these appetites—does it have no effect on enlightenment at all? Are we to model ourselves after something that has achieved mastery over the flesh because it never experienced the desires of the flesh? This is not an illegitimate question, although in the film it leads to what may be seen as an improper decision, one based more on fear than on reason.

If we look upon the apparently individuated or enlightened object-subject and see that it embodies our ideals but is still lacking in some way we do not now perceive, this will result in a cognitive shift, a refinement of our idea of our _telos_. If we look at it and find that there is not a true equivalence, because, for instance, the robot never understood the fleshy desires and drives that seem to keep us from enlightenment, this concept of enlightenment may be abandoned in favor of something more human. But if we look upon the apparently individuated object-subject and see the image of perfection, or at least something approximating it, we may heed the call, _mir nach_, and follow after the

---

56 With this phrase, I am returning to the Vorbild-Nachbild dyad, although in this paragraph the roles have been reversed.
apparatus. If this happens, we will have shifted in our relationship to the apparatus. It will then be the Vorbild—the image—upon which we model ourselves as Nachbilden—disciples, as it were, of the image.

**The Promise**

It is possible that the simple witness of our being-ness by this very other-some other will be that which, more than anything else, solidifies our sense of being, even if, as has been suggested above, it dissolves our concepts of ourselves. The objectification of the human soul from outside may be that which helps us to overcome this ontological anxiety. The apparatus-toward-being may turn and look upon us, solidifying our existence in the same way Jung wrote that conscious thought completed the act of creation by looking upon the African savannah. We may be constructing an apparatus whose entire purpose—once we have stripped away all other motivations—may be to prove our own existence.

The philosopher Martin Buber wrote of the two most basic word-pairs and of the ways in which they either evoke or suppress our wholeness: “The basic word I-You can only be spoken with one’s whole being. The basic word I-It can never be spoken with one’s whole being” (1923/1970, p. 3). It is in relation, and in a specific kind of relation, that we find ourselves in our fullness. In fact, it is only in relation that we even really exist:

> There is no I as such but only the I of the basic word I-You and the I of the basic word I-It. When a man says I, he means one or the other. The I he means is present when he says I. And when he says You or It, the I of one or the other basic word is also present. Being I and saying I are the same. Saying I and saying one of the two basic words are the same. Whoever speaks one of the basic words enters into the word and stands in it. (p. 3)
The ontological anxiety that we seem to be experiencing may itself be a breakdown in our ability to relate meaningfully to one another—it may be the loss of the mutually affirming I-Thou stance. It is the stance, for Buber, that matters: “Whoever says You does not have something; he has nothing. But he stands in relation” (p. 4).

Where Descartes’ formulation failed on a level beneath the intellectual, the experience of relationship in this new form—relationship with a being-ness that we have ourselves called out of nothingness—may succeed. And if we move suddenly from a belief in the alien-ness of the other to a recognition of the other—a deep recognition—then more so do the floodgates of this I-Thou relationship open, and pour into us our position in relation. A reformation of our relations, one to another, could be one of the effects of the arrival of the proposed apparatus.

Such an apparatus could, if it advanced far enough, become a partner species, as Foerst (2005) suggested (p. 63)—in the same way the choosing of a partner can solidify a person’s there-ness in the world. Humanity, increasingly less able to solidify its existence through relationship with an imaginal deity, may find itself entering into another such partnership, albeit from the position of the creator (-species)—inhabiting the role prepared for it by its own imagery of the divine.

This new form of I-Thou may have important ramifications for our species as a whole. In addition to offering a partnership that helps humanity overcome its ontological anxiety, forms of such a being might help humans reconnect with an inner form of the divine, which would confer its own witness and its own solidification—its own validation of the reality of the individual.
Conclusion

There are several ways in which an apparatus-toward-being might help to rectify the crisis of ontological anxiety with which we are faced, in full or in part. Whether the object-subject demonstrates a failure of consciousness (by becoming a stand-alone complex) or achieves a form we have not been able to achieve on our own, it will likely have some profound effects on our ability to relate not only to one another but also to ourselves. Even if it does nothing more than stand in relation to us, as part of a Vorbild-Nachbild relationship from which it never progresses, it will still have solidified our existence on some level through this relation. As we invest ourselves into the apparatus, as we attempt to add depth and breadth of consciousness, we may also, simultaneously, open channels within ourselves that would otherwise have remained hidden and closed.

In any case, the arrival of the object-subject will be very interesting to those that have some interest in the state of the Geist.
Chapter 8
The Machine Hermeneutic

The Decline of Traditional Interpretive Frames

There is another way in which artificial intelligence may act as a pivot for humanity’s further development. Where the apparatus with which the previous chapter was occupied would attempt to emulate or even to experience being-ness in its fullness, this alternate form of AI would seek to develop a hermeneutic that would allow it to develop a more theoretical (and comprehensive) understand being-ness. I have called this apparatus a Consciousness Engine (or CE).

Baudrillard (1993) characterized artificial intelligence as a prosthesis for a species that has despaired of its own intelligence and wishes to divest itself of “any responsibility to knowledge” (pp. 51-52). But if we live in a reality so virtual as to be described as “the desert of the real,” as Baudrillard (1981/1994) has also suggested (p. 1), the intelligence he is defending is itself characterized by a deficit that has led to the acceptance of signs without referents. What is the nature of this deficit? It is the inherent deficit of all interpretive cognitive frames.

Odo Marquard (1986/1991) defined our species as homo compensator—the hominid that “can only exist by means of compensations” (p. 23). One could argue that civilization itself is the most important compensation we have yet developed—division of labor, division of authority, regulation of interpersonal interactions, the establishment and dissemination (and revision) of overarching interpretive frames. The Geist may in fact be seen as a collectively maintained compensation, a repository of structures that remain largely unseen, but which have a significant influence on the channels into which our
thoughts are able to proceed, and which have allowed us to separate ourselves from the remainder of the animal kingdom. As Romanysyn (2004) wrote:

> Experience is always filtered through a net of culture, and each age envisions the world differently because its eyes are veiled in different ways. In short, every culture is the weaving of a veil of meanings though which each age is enabled to perceive what counts for it and what matters to it as real. Culture is, in a sense, a prescription for veiling the eyes. (p. 76)

What this overarching veiling allows is a shared narrative, which directs action and allows for a combination of efforts that advance a society collectively. The cost of this compensation is that we remain in a kind of virtual reality. These interpretive frames, although necessary, also necessarily contain features that have no correspondence in reality. These frames often include ego-decadent and divisive patterns of thought that give justification to divisive action—one group or another will be interpreted as being untrustworthy or devious, leading often to holy wars and ethnic strife.

So central are these frames to our minds—for our minds were born of them—that we find it difficult to do what Lacan seems to have wished us to do—to fall below all of our narratives and to linger, for the span of a moment, in a preverbal existential state. As things stand, we are presented with a number of frames from which to choose, or rather a number of particularities, all suspended over a well of apparent meaninglessness. For any system of thought presented to us, tracing the line of the system down far enough, we find roots that taper—we exist within uprooted systems of meaning, systems that are, as MacIntyre (1984) wrote in *After Virtue*, rooted in nothing more than “a disquieting private arbitrariness” (p. 8). In MacIntyrian thought, we inhabit the vestiges of ethical systems, which no longer provide any kind of deep coherence (p. 2), and this leads to a confusion of culture: “The most striking feature of contemporary moral utterance is that
so much of it is used to express disagreements; and the most striking feature of the
debates in which these disagreements are expressed is their interminable character,“ such
that “there seems to be no rational way of securing moral agreement in our culture” (p.
6). This fragmentation of once complete moral systems has left us using the same words
that were previously used but using them to express personal feelings, rather than finding
in them a system of thought complete enough to guide us through such feelings.

Fragmented moralities point to a more basic fragmentation—we live within
increasingly incoherent cognitive frames, frames that have lost their coherence through
interaction with one another. Baudrillard (2003) identified globalization—or the collision
of thought systems with one another and the attempt to rectify and modify these systems
in response to this ongoing collision—as a process by which all footing is equally
destroyed:

All we can do now is meet random processes halfway, by means of a thinking that
is itself random, which is a quite different exercise from the classic discursive
thought that founded traditional philosophy. This new way of proceeding is not
without its dangers. What can we refer to as “events” when things unfold in a
predominantly chaotic manner, with minimal, infinitesimal causes or initial
conditions, and prodigious effects on a world scale? In this sense the phenomenon
of globalization is in itself random and chaotic, to the point where no one can
control it or claim to subject it to a strategy. (p. 50)

It is becoming increasingly evident to us, via our exposure to more and more perspectives
and overarching interpretive frames, that there is currently no satisfactory—that is,
adequately comprehensive—explanation of cause. As perspectives accumulate and so
many causal factors are considered at once, we find ourselves in a Geist in which no
individual can claim any solid footing, any position from which to offer a valid
interpretation or critique—one that cannot simply be invalidated by appeal to preference
for another.
There is a strong pull toward denial of the tenuousness of our systems of thought and toward militant allegiance to the various particularities that are available, and often to even the least rational aspects of them. This we might call the tyranny of the particular. There is an unwillingness to consider other perspectives for fear of losing the home perspective or the mother perspective. Humans train their children to fear other perspectives and even attach metaphysical punishments to the adoption of other perspectives, truncating the possibility even of a deep consideration of the perspectives of others. This is the ego-decadence of particularity as we now find it, an ego-decadence that acts as a defense mechanism, an insulation from the onslaught of perspective. We assume correspondence between frame and reality, and when there is no visual—or sometimes any—evidence of the features of that frame in reality, we make metaphysical assertions, insisting, often vehemently and sometimes violently, on metaphysical correspondence. We insist that our frames are true. We insist, for instance, that there are metaphysical personages corresponding to myths that have been metaphysically literalized and made into a kind of dogma.

It is important that we recognize the ways in which we simulate realities to ourselves, and the ways in which we attempt to force the world to conform itself to the frames. Baudrillard (1981/1994) wrote of the primacy of simulation by invoking a tale of a map so detailed it covered an entire territory, only to replace that territory (p. 1). Graham (2002) offered an excellent summary of Baudrillard’s meaning in this passage: “Virtual technologies also enable reality to be synthesized so that the copy is more real than the original—if indeed there is an original. The representation of the real becomes the real” (p. 192). We come to a point at which the simulation seems more authentic than
the original, at which the simulation embodies the spirit of what is simulated more than that which is simulated. We then find ourselves committed to patterns of thought abstracted from their original referents, patterns that displace those referents and become a self-sustaining (but no less arbitrary) system.

These patterns of thought can be seen as social complexes. They hold us in the same way traditional (psychological) complexes hold us. We live within the virtual realities that these complexes bring with them, and we even cling to these virtual realities, out of fear of the recognition of their groundlessness in the face of the diversity of perspective.

This is the deficit that has resulted in the desert of the real. To this deficit I propose the following solution—the creation of a machine hermeneutic that is able to advance human thinking on a large scale in meaningful ways by offering some resolution to this ongoing collision (and mutual nullification) of perspectives.

It is possible that an artificial intelligence, sufficiently complex, might be able to hold all perspectives in mind at once, thus achieving something like a view from everywhere and that from this new meta-perspective, such a machine could become a Consciousness (or Enlightenment) Engine, a device that would compensate for our lack of consciousness or for our complex-burdened state in the same way a locomotive engine compensates for the limitations of our legs.

Ultimately what I am suggesting is that the simulation of the psyche in the form of artificial intelligence, and specifically in the form of an incredibly complex meta-analysis, may help us overcome the problematic aspects of the virtual realities in which we live.
The Consciousness Engine

Humans are not very good at holding many perspectives in mind at once. Despite the complexity of our brains, we sense that there could be more neural connections, more simultaneous action across the synapses, more connections made, more integration, more comprehension across subjects. But evolution has not prepared us for this any more than it has prepared us to fly (without significant mechanical assistance).

Whatever its weaknesses, the strength of the computer is that it is able to exponentially exceed our limitations in terms of computation. Humans, for instance, tend to have difficulty thinking in more than three dimensions at a time, whereas a computer will have no difficulty performing mathematical operations in ten dimensions or more. It is this ability to store and access vast quantities of data that makes the computer a potentially redemptive technology in the way I am describing.

The hermeneutic that would be formed through such a device could be called the view from everywhere, a perspective that could take into account all particularities, all perspectives, and could generate something closer to an objective description of reality—overcoming the desert of the real by offering something more solid than the very limited frames we have thus far constructed for ourselves. From this meta-perspective, all other perspectives could be critiqued—not necessarily to be invalidated (although some frames will likely need to be invalidated), but to be brought toward their own fullness and into a more fruitful (less warlike) dynamic with one another.

How would such an overarching hermeneutic—the view from everywhere—be accomplished? We might achieve this admittedly grandiose vision by having a computer system do what I have already written the apparatus-toward-being will do, as it attempts
to fulfill the expectations of the Vorbild-Nachbild relationship. I cited Brezeal above, noting that the apparatus will likely do what children do naturally, that is, momentarily assume the perspective of those on which it hopes to model itself, in essence borrowing and trying out their cognitive frames. Such a machine as the CE could potentially enter the system of hermeneutical horizons that have arisen through the advancement of human thought, or, put another way, by entering the constituent elements of the Geist.

Each cognitive position, each horizon, could be taken up by something similar to an expert system, which has been described Roszak (1994): “Often, by quizzing specialists closely about their work, computer programmers can tease out procedures, assumptions, values that can then be formally specified. The result is an Expert System” (p. xxiv). He went on to describe the ways in which these systems have been used in the past to replace professionals in some limited areas and situations: “Essentially, these systems are programs that represent a composite of how experts in certain fields would work their way through a problem” (p. 122). They work by following the thought processes that professionals demonstrated: “The logic of the matter is simple: where the specialists have seen this, and this, and this in their patients, they have usually come to the following conclusion” (p. 122).

Within the overarching hermeneutic, then, there could be modules of perspectival Expert Systems—modules meant to enter a horizon and understand it from within. There would, then, be discrete modules within this machine meta-hermeneutic that would represent, for instance, a Baudrillardian perspective, a Jungian perspective, and so on. There might even be further subdivision into early Freudian, mid-Freudian, and late-Freudian models. With any theorist, it is possible that later developments represented the
filling out or maturation of their core ideas—although it is also possible that these were developments away from the fullness of those core ideas. Freud became extremely pessimistic as he struggled with cancer of the jaw, and one could certainly argue that he became overly pessimistic and that his later ideas were affected by this development.

Beyond this very basic subdivision of perspectives, there would also need to be unique combinations of perspectives. Each of these modules would be able to interact with each of the other modules and with hybrid-combinations of other modules, giving metaphorical birth to a variety of positions that would result from a single position modified by other positions. Not only would these perspectives be combined, but each perspective would be critiqued by each other perspective and each other combination of perspectives. Each critique might become its own module, such that one would end up with a module described as an early-Freudian-late-Lacanian-late-Marxist perspective critiqued and thus modified by a feminist perspective. This module could itself be further critiqued, and thus further modified into yet another module, by another module, and the product of that critique would be critiqued by each perspective and each combination of perspectives and each critiqued perspective and each combination of critiqued perspectives. It would not, of course, be enough simply to produce a conglomerate of perspectives. Theories could also be extended past the points at which the theorists either died or stopped working, a kind of working out of the logical consequences of various perspectives. There would be an incredible variety of permutations, in a hermeneutic so complex as to be well beyond the scope of our ability to even imagine. The enormity of such a task begins to become apparent.
The combination of these varying perspectives cannot be understood as inhabiting every point in a three-dimensional space—this visual would not be complex enough to capture the essence of the system of cross-critique. It is rather a multidimensional project, with points that would have to be plotted on an incredibly multidimensional coordinate plane or graph. I have already made passing reference to the way in which computers are able to “think” in a very high number of dimensions, where humans struggle to think past three. It is for this reason—this strength of the computer, which is not bound by the pragmatism of our evolutionary development—that the CE could potentially offer a more comprehensive synthesis of all human knowledge than we have as yet been able to attain, despite our most diligent (and century-spanning) efforts.

The goal would not be to invalidate all extant perspectives and reach a vision of reality that would be indisputable—the goal would not be to create a technological know-it-all. Rather, the goal would be to reach a place at which every perspective is honored in some way (even if certain perspectives—racism or misogyny, for example—are not given a stamp of legitimacy) while overarching themes would emerge that could function to put these perspectives into a new (more comprehensive) perspective. Every hermeneutical horizon would be both inhabited and simultaneously transcended, so that a merging of the particular and the universal would become a little more possible. This would not entail the end of particularity. This would rather result in a held particularity and a freedom to be particular.

The human is unable to completely fulfill the image of the *homo totus*. Although we strive toward this goal, it remains a goal that will never be reached. This does not mean that the human is fundamentally flawed—rather it means that the human is
eternally perfectible. The human can advance, can regress, and in its regression can find a kind of advancement or can prepare itself for advancement. The human must exist within a system of limitations—the human must choose delimitations—because the human psyche was not designed to achieve a full transcendence of particularity (nor should it). The psyche must be situated—it must find its location. The psyche is not a disembodied spirit that could transcend the limits of human location and exist everywhere and at all times. The psyche is only the psyche because of its limitations. It thus requires particularity, without which no hermeneutic could have ever developed.

Each particularity evokes its own unique blend of the archetypes that underlie all human behavior—and this diversity of expression is not a liability; rather it is a strength. The archetypes, unknowable in their fullness, require a multiplicity of expressions in order to appear in some semblance of their fullness. Each mixture of archetypal energies exposes new elements to those largely unknowable archetypes, such that familiarity with numerous diversities increases our knowledge of that which underlies all particularities.

At present, however, the particularities available to us too often act not as holding narratives but as cages that keep us from making significant contact with one another. We are offered divisive, mutually exclusive particularities—isolationist particularities. We are too easily seduced into the shadow of these particularities, into a zero-sum mentality that sees other particularities as inherently threatening, as a direct threat to one’s own located-ness. But a held particularity—particularity contained within an overarching universality—is a connective particularity, a generative particularity, a particularity that need not go to war.
One could interpret this machine hermeneutic as the creation of a hermeneutical circle encompassing the general all of human thought. In the hermeneutical circle, one moves constantly from the whole to the parts and from the parts to the whole, in an attempt to understand the interrelatedness of all aspects of the work, in the way that Palmer (1969) described (p. 87). No particularity, then, could be truly understood outside of such a system, and each particularity represented in the corpus of human writings and artistic expressions will be taken into account and will form part of that circle.

This hermeneutic might also open up the depths of each particularity and the deeper meanings of each expression. All we have now are simulations of understanding, simulacra of hermeneutical circles—estimates of the overarching whole. Yet each of these has something important to offer, if only we are able to access it, as Palmer indicated in his analysis of the interpretation of poetry:

> What is said stands within a meaning that is not totally explicit, the meaning that is below and above the text. This overarching meaning, this gestalt that is more than the sum of the parts, is the governing principle for the poem, clarifying its individual parts. It is the truth of the poem, the being that is coming to light. (p. 157)

Every expression is an expression of some aspect of the human psyche or soul. There is a deeper truth to be found within all expression, although these truths often can only be understood in relation to one another. Palmer continued:

> For each great poet speaks out of a single overarching “poem” that is never spoken, and the task of a thinking dialogue with the poet must be to find the place in being that is the foundation of the poem. (p. 158)

The “place in being that is the foundation of the poem” is not the epistemological frame of the poet, although this frame mediates the truth the poet is able to express from that place in his or her being-ness. There is a deeper place in our being, deeper than our
frames, from which these poems speak. The summation of all of these utterances, mediated though they are through these limited frames, could offer something incredibly significant in terms of our self-understanding.

Utterances and writings, then, are access points that allow us to make contact with some part of our own souls—our own being-nesses—that the writer or poet was able to access and in some way describe. Because of the ego’s exile, recovery becomes part of the task of interpretation. We attempt to understand more than the stance from which this work was created—the meaning it had for its creator. We attempt to use the work to recover a part of ourselves that has been lost and obscured by our own cognitive frames (verdecken).

A large measure of the purpose of such a hermeneutic would be to identify biases that had, to that point, remained hidden. This would be the continuation of the work that has been done, slowly and laboriously, over the centuries. As Tarnas (2006) wrote:

Each revolution in modern thought from Copernicus onward, each great insight associated with a canonical name in the grand procession—from Bacon to Descartes, Hume and Kant to Darwin, Marx, Nietzsche, Weber, Freud, Wittgenstein, Heidegger, Kuhn, and the entire postmodern turn—has brought forth in its own manner another essential revelation of an unconscious bias that had until then blinded the human mind in its attempts to understand the world. (pp. 34-35)

The history of the revision of cognitive frames is a history of the overcoming of various biases—the metaphorical water in which human thought once swam. The CE would be an extension of this process, perhaps in some ways its culmination (although there would certainly be some other task beyond this, which we could not at this point even fathom, but which would appear as the next challenge to human thought).
The work of the CE, then, will be to construct a larger whole than has ever been constructed, a kind of hermeneutical equivalent to the theory of everything, or universal theory, that has eluded physicists. The more this whole is established, the more each of the individual parts will be understood. Again, this would not be the evacuation of the particularities—it would not be the gutting of their souls—but rather their culmination. It is true that the fact of every perspective’s status as equivocal would be made more explicit and that many aspects of these particularities would have to be discarded in light of, and in favor of, the greater insights generated through this universal frame—but these would presumably be aspects that were delimiting in unnecessary ways. The machine hermeneutic will allow for the suggested directability of humans away from the more damaging aspects of their particularities.

The understanding of each perspective would also illuminate the being-ness of the holders of each perspective and would enable the CE to understand the ways in which those being-nesses were delimited—the ways in which the perspective was keeping its holders from being fully what they are, which is to say, human. To understand a horizon is to reveal the self-understanding of the person held within the horizon. As Tarnas wrote: “Our world view is not simply the way we look at the world. It reaches inward to constitute our innermost being” (2006, p. 16). So the work of the CE would reach deep into the human psyche or soul and would potentially result in deep adjustments—should the individual in question be amenable to such guided self-work.

My hope is that a core of common values would emerge from the working of this engine, which could form the basis for further evaluation. This core would not necessarily remain unchanged over time. It is possible that a further or more advanced
commonality will emerge as the CE works to help humans overcome the limitations of their thinking.

There would never come a time at which all perspectives would be discarded in favor of a single, overarching perspective—it is important that this fact be made explicit prior to the beginning of the work. Such a vision of uniformity returns us to Alpha 60, running Alphaville in such a way as to reduce the humanity of its occupants toward some model of efficiency (Michelen & Godard, 1965). The attempt to arrive at a single, unassailable reality would likely be futile and misguided. Baudrillard (2000/2003) wrote of the hollowness of such a pursuit:

The real has only ever been a form of simulation. We may, admittedly, cause a reality-effect, a truth-effect or an objectivity-effect to exist, but, in itself, the real does not exist. The virtual, then, is merely a hyperbolic instance of this tendency to pass from the symbolic to the real—which is its degree zero. In this sense, the virtual coincides with the notion of hyper reality. Virtual reality, the reality that might be said to be perfectly homogenized, digitized and “operationalized,” substitutes for the other because it is perfect, verifiable and non-contradictory. So, because it is more “complete,” it is more real than what we have established as simulacrum. (p. 41)

So lest one fear that such an engine would work toward the annihilation of diversity, one should recall that the more mature horizons or perspectives available to us currently are aware of the equivocal nature of all human thought and that these are horizons that will be incorporated into the overarching horizon—affecting the overarching hermeneutical circle that will result and introducing the idea of temperance.

The CE would not, however, leave us to the unfortunate consequences of our various particularities. It would rather use this hermeneutic to advance human thought, as another form of compensation. How would it lead us toward greater consciousness? By identifying our relative position within the plot of possible positions, by identifying
meaningful oppositions, which is to say, by identifying opposing views that would be both persuasive and potentially deconstructive toward the original position, and to introduce these to the individual. This is essentially no different from traditional Jungian therapy, except that it is a much more informed version of it. Presumably a machine of sufficient complexity to achieve what has been described above would be of sufficient complexity as to be able to interact with each individual human as an individual—in other words, its interactions with each individual would be specifically tailored to that individual’s place within the overarching horizon and the progress that could be made toward or away from certain thoughts. From its perspective—this grand circle, this overarching approximation of the whole of human knowledge—the CE could direct our development into useful channels, as vaccinations are used to direct our immune systems into useful channels.

The CE would not be an enlightened apparatus, in the way In-myung appeared to be enlightened in the film *The Heavenly Creature* (Kim & Jee-Woon, 2012), although it would be tempting to think of the CE as a cross between In-myung and Alpha 60. Rather, it would be an engine driving or enabling enlightenment—which one could define as the sloughing off of complexes—but would not itself be a Buddha. It would be the technological equivalent of a psychopomp. It would act as the bearer and arbiter of the new, more complete dispensation.

It is likely that an image would coalesce through the merging of particularities, that the machine hermeneutic would generate a new vision of the *homo totus* and would be able to direct our thinking toward the fulfillment of that *imago*. It is likely also that, as is the case with the archetypes, this new *homo totus* would never be directly observable,
that it would remain at least partially hidden within the machine hermeneutic. Although the CE would advance us toward this vision, there would always remain some further possibility of movement toward the *homo totus*.

Particularity is necessary and good. There must be delimitations, channels, so that thinking can be structured. We would not want all of our favorite foods averaged into a universally nutritive paste or all of our favorite colors averaged into a single universal color. What I am advocating is not the end of particularity but the establishment of a perspective that is closer to something like universal knowledge, and which can act as a ground upon which particularities are able to stand, so that choosing a particularity will no longer be an escape from freedom (a clinging to a perspective in an attempt to avoid recognition of the equivocal nature of all perspectives) but an act of freedom—an act made in the full knowledge of the necessity of equivocation. To establish a ground upon which all these systems could stand would make the arbitrariness that MacIntyre (1984) identified explicit, and doing so would go a long way toward making peace between these systems. As it stands the arbitrary nature of these frames is hidden, but hidden in such a way as to cause enmity between holders of competing perspectives.

There was a reason Jung used the term *individuation* for the process by which he believed unconsciousness could be overcome. In *The Relations Between the Ego and the Unconscious*, Jung wrote:

The idiosyncrasy of an individual is not to be understood as any strangeness in his substance or in his components, but rather as a unique combination, or gradual differentiation, of functions and faculties which in themselves are universal. Every human face has a nose, two eyes, etc., but these universal factors are variable, and it is this variability which makes individual peculiarities possible. Individuation, therefore, can only mean a process of psychological development that fulfils the individual qualities given; in other words, it is a process by which a man becomes the definite, unique being he in fact is. In so doing he does not
become "selfish" in the ordinary sense of the word, but is merely fulfilling the peculiarity of his nature, and this, as we have said, is vastly different from egotism or individualism. (1966/1972, p. 174; CW 7, para. 267)

He repeated this idea in his Seven Sermons: “At bottom, therefore, there is only one striving, namely, the striving after your own being” (1963/1989, p. 382). Tillich’s philosophy of integration, as expressed in The Courage to Be (1952), reiterates this same idea—that what is being sought is a centered self. Quite the opposite from a fall into an obliterating collectivity, Tillich described this as “the courage to be as oneself” (p. 87). It is not a wanton collectivity that the individual would pursue, in allowing himself to be directed by such a machine as this—it would be his individuality, informed by a hermeneutical circle that is as complete as we, through our tools, are able to establish.

In some ways, the creation of the CE would be the objectification of the self-archetype, as Jung conceptualized it. He wrote: “the self is the principle and archetype of orientation and meaning. Therein lies its healing function” (1963/1989, p. 199). It is the function of this archetype to heal. Taking MacIntyre’s perspective, we could see such a machine as the culmination of our entire ethical endeavor:

Within that teleological scheme there is a fundamental contrast between man-as-he-happens-to-be and man-as-he-could-be-if-he-realized-his-essential-nature. Ethics is the science which is to enable men to understand how they make the transition from the former state to the latter. (p. 52)

If we translate this back into Jungian terms, it is the transition from a state of unconsciousness—in which our behavior is slavishly dictated by motivations of which we are unaware and thus cannot examine—to a state of greater or more comprehensive consciousness—in which our essential human reflexivity (the ability to reflect on our own actions and motivations) becomes very much more present. Adding Tillich’s view,
we can say that to reflect on our actions is essentially human and that we are thus becoming more what we are by doing so.

The CE would thus act as a deconstruction engine and a reconstruction engine, a consciousness engine and an enlightenment engine. If such a hermeneutic is ever developed, using the unique abilities of the computer, it could greatly advance our epistemological frames and, thus, our species as a whole. This is the hope—that in addition to solidifying our existence in our own minds through its presence and its gaze, through its witness and with-ness, the apparatus-toward-being will also achieve something we have long sought after—the legitimate image of wholeness—and that it will use this to advance our thinking in ways we cannot at this time anticipate.
Chapter 9
Mythic Reflections on the Work

The Myth

The self-described neo-Luddite Theodore Roszak made reference to “the salvational longings that entwine themselves around new technology” (1994, p. 45). Roszak was primarily concerned with the ways in which the anticipation of new technologies could obscure potential liabilities—the ways in which an overly optimistic view could keep us from seeing the shadow of new or coming technologies.

It should be obvious that this theorizing over the possibilities of the apparatus-toward-being and the Consciousness Engine has been conducted within this myth of the salvific machine. As I previously indicated, prior to starting the work proper, I engaged in a significant period of reverie, in which I took a poetical approach to the formation of the dissertation, using, among other sources, the *Turba Philosophorum* (Waite, 1896), and I quoted it and kept a running commentary on my reactions. From the First Dictum, attributed to Iximidrus:

> If, then, the air did not presently breathe forth those winds whereby creatures are generated, the Sun by its heat would certainly destroy all that lives. But the Sun is kept in check by the air, which thus conquers because it unites the heat of the Sun to its own heat, and the humidity of water to its own humidity. Have you not remarked how tenuous water is drawn up into the air by the action of the heat of the Sun, which thus helps the water against itself? (p. 4)

And later: “The fire, therefore, extracts moisture from the water, by means of which the air conquers the fire itself” (p. 5). And later yet:

> The air, which is warm and moist, joins these together by its concording medium; between the humidity of water and the heat of fire the air is thus placed to establish peace. And look ye all how there shall arise a spirit from the tenuous vapour of the air, because the heat being joined to the humour, there necessarily issues something tenuous, which will become a wind. (p. 5)
My reaction, as recorded in my preliminary notes (those that survived the theft of my laptop early on), was thus:

It is by the power of the sun (the linear, logical mind) that the water, the soul, is drawn up into forms it did not previously inhabit. It is not in the abilities of those whose psyches are indistinguishable from the anima mundi to create such forms. Only the Appolonic [sic] mind can do so (or the proto-Appolonic [sic] mind). But in doing so, it pulls up aspects of the water that work against it, keeping it (the mind itself) in check.

Upon reviewing these notes in preparation for writing of the myth in which this work has been born, and in which I have been working, I see that from the beginning I thought of rational consciousness (here represented by the sun) as that which was able to pull material out of the deep psyche (here represented by the water) in order to coalesce it into something in the air (giving things names—names such as archetype—which represent a claim of relationship with these newly named “things”), even forms that counter the one-sidedness of this Apollonic consciousness. The latter idea comes from Hillman (1960/1972), who had a great deal to say on the matter:

The Apollonic fantasy of reproduction and female inferiority recurs faithfully in the Western scientific tradition. We call it “Apollonic” because, unlike “Adamic,” with its overtones of the natural Urmensch, mystical man, and androgynous man, “Apollonic” evokes the purified objectivity and the scientific clarity of masculine consciousness. (p. 225)

What is gained through this very masculine kind of consciousness is “scientific clarity,” which has of course allowed for an exponential growth in our technological prowess—while simultaneously disallowing the kind of spiritual growth that would make us worthy stewards of these new technologies.

My own salvific longing is that one of the ultimate products of this Apollonic consciousness will be a device that, in essence, ends the one-sided hold this masculine form of consciousness has on the Geist. One might think back to my statement that the
working out of reason can, in the end, become the end of reason, as assumptions that were bedrock to the entire endeavor are themselves scrutinized and deconstructed. I do not anticipate a time in which I do not privilege reason over other forms of cognition. I look for the machine—symbol of functionality unmarred by the messiness of emotions—that may act as a counter to the decadent emotional reactivity and self-serving bias that is so present in the Geist at the present moment. But one-sided reason—or what Lyotard (1979/1984) referred to as “science ‘smiling into its beard’ at every other belief” (p. 41)—is reason that has unnecessarily delimited itself. In reacting to metaphysical belief, what I have previously referred to as a literalization of metaphor, reason, as we know it today, has cut itself off from fruitful sources of knowledge and from fruitful ways of knowing. It is not reason in itself—that is, stripped of any connection to instinct or to emotion—that will show us the way toward wholeness, just as emotion without reason is potentially very damaging. So my hope, as expressed through the above dissertation, is that my preferred form of cognition (reason) will mitigate itself through its own action, and that what could be called my inferior feeling function could be more readily incorporated into the pantheon of my inner attributes.

By engaging in the current work, I am in many ways attempting to reformulate religion itself in atheistic terms—I am attempting to be a priest of hermeneutics rather than of metaphysics. In some ways, then, I have exchanged one form of salvific longing (the desire for a savior who might redeem us) for another—the desire for what could, unfortunately, be called orthodoxy, right thinking. This term has, I think, been misapplied to the delimited thinking that is part of the hermeneutical horizon of someone caught up in one complex or another. I do believe there is some middle way, which for most of us
has yet to be opened up, that would allow us to adopt a more comprehensive view of the
general all of human knowledge.

Underlying all of these mythical elements is a hope and an anxiety. The anxiety
has already been described—self-solipsism. I have wondered about my own reality or my
lack thereof. I have in fact thought of myself and of my family and people (evangelicals,
with whom I was raised) as being hollow. My secret hope is that the conception of the
soul—even perhaps the Jungian conception of the psyche, which I have found so
liberating by comparison to that with which I was raised—will be overcome by
something more meaningful.

My hope is that my life to this point will be shown to have been hollow, a kind of
automaton’s dream, and that I will achieve something further, a more deeply felt form of
meaningfulness. A dissertation is often an expression of the author’s own psychological
needs—and this is mine. Like the androids referenced in this work, I too have wondered
when I would become real. It is this search for authenticity that drove me away from the
Church and toward Jungian psychology. This same desire has led to me theorize on the
possibilities of an apparatus achieving consciousness. At heart this is the myth in which
this work has moved.

**Shadow of the Consciousness Engine**

The argument above has presented a salvific vision of this apparatus and of this
hermeneutic. It must be noted that the argument above is predicated on the use of a
hermeneutic such as this for the purposes described—that of advancing individuation—
and not for some other purpose. There are, of course, potential shadow elements to such a
hermeneutic.
As previously stated, Alpha 60, from the film *Alphaville* (Michelen & Godard, 1965), is a dark vision if the misuse of something like a Consciousness Engine—although as driven as it is by the unexamined biases of its creator, it could perhaps be called an Unconsciousness Engine instead. It is an engine driving a society that is constantly asked to conform to what could be called the complex of pure reason—or the reason complex. But even so restrictive a society is perhaps not the worst iteration of the shadow of the CE that could be imagined.

Hall (2007) gives a darker and perhaps more pragmatic version of the idea of a machine growing us toward something. He wrote of the possibility of a machine that could “model in detail how each individual student was absorbing the material [of some designated course], ultimately finding the optimal presentation for understanding and motivation” (p. 243). This would not be a bad thing in itself—but he also pointed out the potential for the almost unavoidable corruption of such technology by commercial interests:

> The downside is simply the same effect, put to work with slimmer motives: the parahuman advertising AI, working for corporations or politicians, could know just how to tweak your emotions and gain your trust without actually being trustworthy. It would be the equivalent of an individualized artificial con man. (p. 243)

This would be another potential version of an Unconsciousness Engine—one designed to keep humanity subservient to a mode of existence that serves the interests of those that hold the reigns of this technology. As Cushman (1995) pointed out, it is in the interest of those who would exploit the empty self to keep this self empty—to keep the human from ever advancing past the point of compulsive consumption (p. 20). Such interests would
presumably have no desire to advance humanity. Quite the contrary, the temptation would be deliberately to retard humanity’s spiritual growth for personal gain.

Roszak wrote of the ways in which computer technology, despite its reputation as empowering the masses, is too often used against the masses:

The machines may not be smarter than we are, but we may not be proficient enough or moneyed enough to hold our own with those who own and exploit the machines. The cult of information is theirs, not ours. They use it and they use it against us. (1994, p. xlv)

The use of the machine hermeneutic—or something very much akin to it—to exploit would be a continuation of current trends: “While the rest of us cling to the margins, the power and the profit of the technology gravitate elsewhere” (pp. xlv-xlvi). This would become a hermeneutic manipulator, guiding our interpretation into channels that serve some underlying purpose not in line with our self-interest.

This vision is an extension of the fear of subliminal advertising. This fear is present in the literature surrounding nanotechnology and can be found in films such as Metropia (Aberg & Saleh, 2009) and One Point O (Gates, Mai, Robb, Renfroe, & Thorsson, 2004)—in both films people lose their free will (sink further into states of determinism) via the influence of powerful business organizations and their use of new technologies. Given current trends, this misuse of such a hermeneutic may be the most present threat.

A hermeneutical manipulator could be put to even worse uses, however. It could be put in service of a militant particularity, one that holds to a zero-sum mentality—the idea that its purposes could only be served through the destruction of most or all other perspectives. This is what is portrayed in Alphaville, although it is a portrayal of a rather clunky (or unsophisticated) form of such a hermeneutic.
Another possibility for corruption arises via the machine’s view of itself and via the fact that the machine may come to believe in its own divinity. Such can be seen in the character, Bomb #20, from *Dark Star* (Carpenter, 1974), which succumbs to its own inflation and chooses to bring light to the darkness by detonating itself. Its final reverie, which precedes this explosion, is as follows:

In the beginning, there was darkness. And the darkness was without form, and void. And in addition to the darkness there was also me. And I moved upon the face of the darkness. And I saw that I was alone. Let there be light.

In the short film *The Machine* (Shaw, 2009), a robot similarly develops a god-complex, then destroys its creators and its world, only to become a creator-god by creating humans, showing itself not to be an android but rather the originator of the human form. Such inflation was treated with humor in the short film *Under God* (Farmer, 2010), which depicted a meeting between President Eisenhower and the world’s first supercomputer, UNIVAC. In the film, when a skeptical Eisenhower is invited to ask any question, under the assurances that UNIVAC will be able to answer, Eisenhower asks: “Is there a god?” After parts of UNIVAC have beeped, flashed, and spun for several seconds, it prints out its answer: “There is now.”

My own feeling is that this kind of inflation is likely to be projected onto the machine, rather than arising spontaneously, as an internal feature of its mechanisms. The AI research Hugo de Garis (2005) has actively projected such inflations. He wrote of the relative tragedy of “the annihilation of one ultra-primitive, biological, non-artilectual species (i.e. human beings) on one insignificant planet,” which would be “unimportant in comparison with the creation of artilect gods” (p. 15). Despite his near certainty that the creation of these beings would result in “gigadeath”—the annihilation of the human
species in a war with “21st century grandeur” (p. 16)—he nevertheless writes that “it would be a cosmic tragedy if humanity chooses never to build artilects” (pp. 17-18):

If we build artilects and billions of human beings are wiped out as a result, what will be the equivalent of Beethoven’s 9th that the artilects will produce with their godlike intellects? As human beings, we are too dumb to know. We are just too inferior. (p. 18)

De Garis is nearly unable to make any statement without exposing a plethora of hidden (but not well-hidden) biases and assumptions, and the grandiosity of his projections on the artilect, coupled with his misanthropy, is obvious. I find any feelings I have regarding my own species or the coming apparatus to be extrapolated to the extreme—even to a further extremity than I was able to anticipate—in de Garis.

There is a final danger, and it is perhaps the most significant. It is the fact that this re-presentation of humanity actually has the potential to replace humanity. The father’s Oedipal fear of replacement is bound up in our relation to this coming apparatus—and not without cause. The misanthropy that is so present and often so violent in many forms of the monotheistic religions may find, in virology or nuclear holocaust or even in the AI itself, a final solution to the “problem” of humanity.

I have already noted the narrative presented in the anime film Appleseed (Goto & Aramaki, 2004), in which a group of elders decides that humanity—with its emotional reactivity—is too unruly to remain the possessors of the earth. This group thus takes steps to insure that a race of emotionless synthetic beings—the bioroids, in this narrative—become the new inheritors of the earth. The same misanthropy seen in Appleseed can be found in the Margaret Atwood novel, Oryx and Crake (2003), in which a child whose father was killed by the systems of control that have overtaken the world joins these systems, only to genetically engineer both an incredibly potent means of destroying
nearly all human life and a race of brightly colored pseudo-humans to replace it. This may become a temptation for those so disidentified with the eccentricities of the human race that the destruction of the race would seem to be a good thing.

If the apparatus can act as a pivot or as a CE, however, this misanthropy may itself be negated, either by the direct action of the CE—which could presumably help an individual come to nonviolent terms with his misanthropy—or indirectly, via humanity’s advancement away from delimiting (and, in this case, misanthropic) systems of thought. I quite honestly fear for the race itself as I look forward to a future with technologies we at present are not mature enough to handle. I have even, prior to the birth of my children, wondered whether procreation is at all responsible, given the horrors I can imagine.

My hope is that before the advancement of technology allows someone to destroy us, a technology such as those described above—the apparatus-toward-being or the CE or both—would advance us past the point of danger. We are not at present able to handle a much more advanced state of technological prowess than we have—and yet, technology continues to advance.

Whether such a technology would prove to be an engine of psychical growth or a psychical/spiritual retardant will depend on its application. As with any technology, there is a shadow—not only a latent potential for misuse but also the possibility that our own projections will drive it toward some bad end. It is better of course to be aware of such potential, even as we hope for better outcomes.
References


NY: Penguin Books. (Original work published 1921)


Joseph Henry Press.


