

H O W   W E   B E C A M E  
P O S T H U M A N

---

*Virtual Bodies in  
Cybernetics, Literature,  
and Informatics*

N . K A T H E R I N E H A Y L E S

*The University of Chicago Press  
Chicago & London*

N. KATHERINE HAYLES is professor of English at the University of California, Los Angeles. She holds degrees in both chemistry and English. She is the author of *The Cosmic Web: Scientific Field Models and Literary Strategies in the Twentieth Century* (1984) and *Chaos Bound: Orderly Disorder in Contemporary Literature and Science* (1990) and is the editor of *Chaos and Order: Complex Dynamics in Literature and Science* (1991), the last published by the University of Chicago Press.

The University of Chicago Press, Chicago 60637

The University of Chicago Press, Ltd., London

© 1999 by The University of Chicago

All rights reserved. Published 1999

Printed in the United States of America

08 07 06 05 04 03 02 01 00                      3 4 5

ISBN (cloth): 0-226-32145-2

ISBN (paper): 0-226-32146-0

#### Library of Congress Cataloging-in-Publication Data

Hayles, N. Katherine.

How we became posthuman : virtual bodies in cybernetics,  
literature, and informatics / N. Katherine Hayles.

p. cm.

Includes bibliographical references and index.

ISBN: 0-226-32145-2 (cloth : alk. paper). — ISBN: 0-226-32146-0

(pbk. : alk. paper)

1. Artificial intelligence. 2. Cybernetics. 3. Computer science.  
4. Virtual reality. 5. Virtual reality in literature. I. Title.

Q335.H394 1999

003'.5—dc21

98-36459

CIP

∞ The paper used in this publication meets the minimum requirements of the American National Standard for the Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992.

CONCLUSION: WHAT DOES IT MEAN  
TO BE POSTHUMAN?

What, finally, are we to make of the posthuman?<sup>1</sup> At the beginning of this book, I suggested that the prospect of becoming posthuman both evokes terror and excites pleasure. At the end of the book, perhaps I can summarize the implications of the posthuman by interrogating the sources of this terror and pleasure. The terror is relatively easy to understand. “Post,” with its dual connotation of superseding the human and coming after it, hints that the days of “the human” may be numbered. Some researchers (notably Hans Moravec but also my UCLA colleague Michael Dyer and many others) believe that this is true not only in a general intellectual sense that displaces one definition of “human” with another but also in a more disturbingly literal sense that envisions humans displaced as the dominant form of life on the planet by intelligent machines. Humans can either go gently into that good night, joining the dinosaurs as a species that once ruled the earth but is now obsolete, or hang on for a while longer by becoming machines themselves. In either case, Moravec and like-minded thinkers believe, the age of the human is drawing to a close. The view echoes the deeply pessimistic sentiments of Warren McCulloch in his old age. As noted earlier, he remarked: “Man to my mind is about the nastiest, most destructive of all the animals. I don’t see any reason, if he can evolve machines that can have more fun than he himself can, why they shouldn’t take over, enslave us, quite happily. They might have a lot more fun. Invent better games than we ever did.”<sup>2</sup> Is it any wonder that faced with such dismal scenarios, most people have understandably negative reactions? If this is what the posthuman means, why shouldn’t it be resisted?

Fortunately, these views do not exhaust the meanings of the posthuman. As I have repeatedly argued, human being is first of all embodied being, and the complexities of this embodiment mean that human awareness

unfolds in ways very different from those of intelligence embodied in cybernetic machines. Although Moravec's dream of downloading human consciousness into a computer would likely come in for some hard knocks in literature departments (which tend to be skeptical of any kind of transcendence but especially of transcendence through technology), literary studies share with Moravec a major blind spot when it comes to the significance of embodiment.<sup>3</sup> This blind spot is most evident, perhaps, when literary and cultural critics confront the fields of evolutionary biology. From an evolutionary biologist's point of view, modern humans, for all their technological prowess, represent an eye blink in the history of life, a species far too recent to have significant evolutionary impact on human biological behaviors and structures. In my view, arguments like those that Jared Diamond advances in *Guns, Germs, and Steel: The Fates of Human Societies* and *Why Sex Is Fun: The Evolution of Human Sexuality* should be taken seriously.<sup>4</sup> The body is the net result of thousands of years of sedimented evolutionary history, and it is naive to think that this history does not affect human behaviors at every level of thought and action.

Of course, the reflexivity that looms large in cybernetics also inhabits evolutionary biology. The models proposed by evolutionary biologists have encoded within them cultural attitudes and assumptions formed by the same history they propose to analyze; as with cybernetics, observer and system are reflexively bound up with one another. To take only one example, the computer module model advanced by Jerome H. Barkow, Leda Cosmides, and John Tooby in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* to explain human evolutionary psychology testifies at least as much to the importance of information technologies in shaping contemporary worldviews as it does to human brain function.<sup>5</sup> Nevertheless, these reflexive complexities do not negate the importance of the sedimented history incarnated within the body. Interpreted through metaphors resonant with cultural meanings, the body itself is a congealed metaphor, a physical structure whose constraints and possibilities have been formed by an evolutionary history that intelligent machines do not share. Humans may enter into symbiotic relationships with intelligent machines (already the case, for example, in computer-assisted surgery); they may be displaced by intelligent machines (already in effect, for example, at Japanese and American assembly plants that use robotic arms for labor); but there is a limit to how seamlessly humans can be articulated with intelligent machines, which remain distinctively different from humans in their embodiments. The terror, then, though it does not disappear in this view, tends away from the apocalyptic and toward a more

moderate view of seriated social, technological, political, and cultural changes.

What about the pleasures? For some people, including me, the posthuman evokes the exhilarating prospect of getting out of some of the old boxes and opening up new ways of thinking about what being human means. In positing a shift from presence/absence to pattern/randomness, I have sought to show how these categories can be transformed *from the inside* to arrive at new kinds of cultural configurations, which may soon render such dualities obsolete if they have not already. This process of transformation is fueled by tensions between the assumptions encoded in pattern/randomness as opposed to presence/absence. In Jacques Derrida's performance of presence/absence, presence is allied with Logos, God, teleology—in general, with an originary plenitude that can act to ground signification and give order and meaning to the trajectory of history.<sup>6</sup> The work of Eric Havelock, among others, demonstrates how in Plato's *Republic* this view of originary presence authorized a stable, coherent self that could witness and testify to a stable, coherent reality.<sup>7</sup> Through these and other means, the metaphysics of presence front-loaded meaning into the system. Meaning was guaranteed because a stable origin existed. It is now a familiar story how deconstruction exposed the inability of systems to posit their own origins, thus ungrounding signification and rendering meaning indeterminate. As the presence/absence hierarchy was destabilized and as absence was privileged over presence, lack displaced plenitude, and desire usurped certitude. Important as these moves have been in late-twentieth-century thought, they still took place within the compass of the presence/absence dialectic. One feels lack only if presence is posited or assumed; one is driven by desire only if the object of desire is conceptualized as something to be possessed. Just as the metaphysics of presence required an originary plenitude to articulate a stable self, deconstruction required a metaphysics of presence to articulate the destabilization of that self.

By contrast, pattern/randomness is underlaid by a very different set of assumptions. In this dialectic, meaning is not front-loaded into the system, and the origin does not act to ground signification. As we have seen for multiagent simulations, complexity evolves from highly recursive processes being applied to simple rules. Rather than proceeding along a trajectory toward a known end, such systems evolve toward an open future marked by contingency and unpredictability. Meaning is not guaranteed by a coherent origin; rather, it is made possible (but not inevitable) by the blind force of evolution finding workable solutions within given parameters. Although pattern has traditionally been the privileged term (for example, among the

electrical engineers developing information theory), randomness has increasingly been seen to play a fruitful role in the evolution of complex systems. For Chris Langton and Stuart Kauffman, chaos accelerates the evolution of biological and artificial life;<sup>8</sup> for Francisco Varela, randomness is the froth of noise from which coherent microstates evolve and to which living systems owe their capacity for fast, flexible response;<sup>9</sup> for Henri Atlan, noise is the body's murmuring from which emerges complex communication between different levels in a biological system.<sup>10</sup> Although these models differ in their specifics, they agree in seeing randomness not simply as the lack of pattern but as the creative ground from which pattern can emerge.

Indeed, it is not too much to say that in these and similar models, randomness rather than pattern is invested with plenitude. If pattern is the realization of a certain set of possibilities, randomness is the much, much larger set of everything else, from phenomena that cannot be rendered coherent by a given system's organization to those the system cannot perceive at all. In Gregory Bateson's cybernetic epistemology, randomness is what exists outside the confines of the box in which a system is located; it is the larger and unknowable complexity for which the perceptual processes of an organism are a metaphor.<sup>11</sup> Significance is achieved by evolutionary processes that ensure the surviving systems are the ones whose organizations instantiate metaphors for this complexity, unthinkable in itself. When Varela and his coauthors argue in *Embodied Mind* that there is no stable, coherent self but only autonomous agents running programs, they envision pattern as a limitation that drops away as human awareness expands beyond consciousness and encounters the emptiness that, in another guise, could equally well be called the chaos from which all forms emerge.<sup>12</sup>

What do these developments mean for the posthuman? When the self is envisioned as grounded in presence, identified with originary guarantees and teleological trajectories, associated with solid foundations and logical coherence, the posthuman is likely to be seen as antihuman because it envisions the conscious mind as a small subsystem running its program of self-construction and self-assurance while remaining ignorant of the actual dynamics of complex systems. But the posthuman does not really mean the end of humanity. It signals instead the end of a certain conception of the human, a conception that may have applied, at best, to that fraction of humanity who had the wealth, power, and leisure to conceptualize themselves as autonomous beings exercising their will through individual agency and choice.<sup>13</sup> What is lethal is not the posthuman as such but the grafting of the

posthuman onto a liberal humanist view of the self. When Moravec imagines “you” choosing to download yourself into a computer, thereby obtaining through technological mastery the ultimate privilege of immortality, he is not abandoning the autonomous liberal subject but is expanding its prerogatives into the realm of the posthuman. Yet the posthuman need not be recuperated back into liberal humanism, nor need it be construed as anti-human. Located within the dialectic of pattern/randomness and grounded in embodied actuality rather than disembodied information, the posthuman offers resources for rethinking the articulation of humans with intelligent machines.

To explore these resources, let us return to Bateson’s idea that those organisms that survive will tend to be the ones whose internal structures are good metaphors for the complexities without. What kind of environments will be created by the expanding power and sophistication of intelligent machines? As Richard Lanham has pointed out, in the information-rich environments created by ubiquitous computing, the limiting factor is not the speed of computers, or the rates of transmission through fiber-optic cables, or the amount of data that can be generated and stored. Rather, the scarce commodity is human attention.<sup>14</sup> It makes sense, then, that technological innovation will focus on compensating for this bottleneck. An obvious solution is to design intelligent machines to attend to the choices and tasks that do not have to be done by humans. For example, there are already intelligent-agent programs to sort email, discarding unwanted messages and prioritizing the rest. The programs work along lines similar to neural nets. They tabulate the choices the human operators make, and they feed back this information in recursive loops to readjust the weights given to various kinds of email addresses. After an initial learning period, the sorting programs take over more and more of the email management, freeing humans to give their attention to other matters.

If we extrapolate from these relatively simple programs to an environment that, as Charles Ostman likes to put it, supplies synthetic sentience on demand, human consciousness would ride on top of a highly articulated and complex computational ecology in which many decisions, invisible to human attention, would be made by intelligent machines.<sup>15</sup> Over two decades ago, Joseph Weizenbaum foresaw just such an ecology and passionately argued that judgment is a uniquely human function and must not be turned over to computers.<sup>16</sup> With the rapid development of neural nets and expert programs, it is no longer so clear that sophisticated judgments cannot be made by machines and, in some instances, made more accurately than by humans. But the issue, in Weizenbaum’s view, involves more

than whether or not the programs work. Rather, the issue is an ethical imperative that humans keep control; to do otherwise is to abdicate their responsibilities as autonomous independent beings. What Weizenbaum's argument makes clear is the connection between the assumptions undergirding the liberal humanist subject and the ethical position that humans, not machines, must be in control. Such an argument assumes a vision of the human in which conscious agency is the essence of human identity. Sacrifice this, and we humans are hopelessly compromised, contaminated with mechanic alienness in the very heart of our humanity.<sup>17</sup> Hence there is an urgency, even panic, in Weizenbaum's insistence that judgment is a uniquely human function. At stake for him is nothing less than what it means to be human.

In the posthuman view, by contrast, conscious agency has never been "in control." In fact, the very illusion of control bespeaks a fundamental ignorance about the nature of the emergent processes through which consciousness, the organism, and the environment are constituted. Mastery through the exercise of autonomous will is merely the story consciousness tells itself to explain results that actually come about through chaotic dynamics and emergent structures. If, as Donna Haraway, Sandra Harding, Evelyn Fox Keller, Carolyn Merchant, and other feminist critics of science have argued, there is a relation among the desire for mastery, an objectivist account of science, and the imperialist project of subduing nature, then the posthuman offers resources for the construction of another kind of account.<sup>18</sup> In this account, emergence replaces teleology; reflexive epistemology replaces objectivism; distributed cognition replaces autonomous will; embodiment replaces a body seen as a support system for the mind; and a dynamic partnership between humans and intelligent machines replaces the liberal humanist subject's manifest destiny to dominate and control nature. Of course, this is not necessarily what the posthuman *will* mean—only what it *can* mean if certain strands among its complex seriations are highlighted and combined to create a vision of the human that uses the posthuman as leverage to avoid reinscribing, and thus repeating, some of the mistakes of the past.

Just as the posthuman need not be antihuman, so it also need not be apocalyptic. Edwin Hutchins addresses the idea of distributed cognition through his nuanced study of the navigational systems of oceangoing ships.<sup>19</sup> His meticulous research shows that the cognitive system responsible for locating the ship in space and navigating it successfully resides not in humans alone but in the complex interactions within an environment that includes both human and nonhuman actors. His study allows him to give an



excellent response to John Searle's famous "Chinese room." By imagining a situation in which communication in Chinese can take place without the actors knowing what their actions mean, Searle challenged the idea that machines can think.<sup>20</sup> Suppose, Searle said, that he is stuck inside a room, he who knows not a word of Chinese. Texts written in Chinese are slid through a slot in the door. He has in the room with him baskets of Chinese characters and a rulebook correlating the symbols written on the texts with other symbols in the basket. Using the rulebook, he assembles strings of characters and pushes them out the door. Although his Chinese interlocutors take these strings to be clever responses to their inquiries, Searle has not the least idea of the meaning of the texts he has produced. Therefore, it would be a mistake to say that machines can think, he argues, for like him, they produce comprehensible results without comprehending anything themselves. In Hutchins's neat interpretation, Searle's argument is valuable precisely because it makes clear that it is not Searle but the entire room that knows Chinese.<sup>21</sup> In this distributed cognitive system, the Chinese room knows more than do any of its components, including Searle. The situation of modern humans is akin to that of Searle in the Chinese room, for every day we participate in systems whose total cognitive capacity exceeds our individual knowledge, including such devices as cars with electronic ignition systems, microwaves with computer chips that precisely adjust power levels, fax machines that warble to other fax machines, and electronic watches that communicate with a timing radio wave to set themselves and correct their date. Modern humans are capable of more sophisticated cognition than cavemen not because moderns are smarter, Hutchins concludes, but because they have constructed smarter environments in which to work.

Hutchins would no doubt disagree with Weizenbaum's view that judgment should be reserved for humans alone. Like cognition, decision-making is distributed between human and nonhuman agents, from the steam-powered steering system that suddenly failed on a navy vessel Hutchins was studying to the charts and pocket calculators that the navigators were then forced to use to calculate their position. He convincingly shows that these adaptations to changed circumstances were evolutionary and embodied rather than abstract and consciously designed (pp. 347–51). The solution to the problem caused by this sudden failure of the steering mechanism was "clearly discovered by the organization [of the system as a whole] before it was discovered by any of the participants" (p. 361). Seen in this perspective, the prospect of humans working in partnership with intelligent machines is not so much a usurpation of human right and responsi-

bility as it is a further development in the construction of distributed cognition environments, a construction that has been ongoing for thousands of years. Also changed in this perspective is the relation of human subjectivity to its environment. No longer is human will seen as the source from which emanates the mastery necessary to dominate and control the environment. Rather, the distributed cognition of the emergent human subject correlates with—in Bateson’s phrase, becomes a metaphor for—the distributed cognitive system as a whole, in which “thinking” is done by both human and nonhuman actors. “Thinking consists of bringing these structures into coordination so they can shape and be shaped by one another,” Hutchins wrote (p. 316). To conceptualize the human in these terms is not to imperil human survival but is precisely to enhance it, for the more we understand the flexible, adaptive structures that coordinate our environments and the metaphors that we ourselves are, the better we can fashion images of ourselves that accurately reflect the complex interplays that ultimately make the entire world one system.

This view of the posthuman also offers resources for thinking in more sophisticated ways about virtual technologies. As long as the human subject is envisioned as an autonomous self with unambiguous boundaries, the human-computer interface can only be parsed as a division between the solidity of real life on one side and the illusion of virtual reality on the other, thus obscuring the far-reaching changes initiated by the development of virtual technologies. Only if one thinks of the subject as an autonomous self independent of the environment is one likely to experience the panic performed by Norbert Wiener’s *Cybernetics* and Bernard Wolfe’s *Limbo*. This view of the self authorizes the fear that if the boundaries are breached at all, there will be nothing to stop the self’s complete dissolution. By contrast, when the human is seen as part of a distributed system, the full expression of human capability can be seen precisely to *depend* on the splice rather than being imperiled by it. Writing in another context, Hutchins arrives at an insight profoundly applicable to virtual technologies: “What used to look like internalization [of thought and subjectivity] now appears as a gradual propagation of organized functional properties across a set of malleable media” (p. 312). This vision is a potent antidote to the view that parses virtuality as a division between an inert body that is left behind and a disembodied subjectivity that inhabits a virtual realm, the construction of virtuality performed by Case in William Gibson’s *Neuromancer* when he delights in the “bodiless exultation of cyberspace” and fears, above all, dropping back into the “meat” of the body.<sup>22</sup> By contrast, in the model that Hutchins presents and that the posthuman helps to authorize, human

functionality expands because the parameters of the cognitive system it inhabits expand. In this model, it is not a question of leaving the body behind but rather of extending embodied awareness in highly specific, local, and material ways that would be impossible without electronic prosthesis.

As we have seen, cybernetics was born in a froth of noise when Norbert Wiener first thought of it as a way to maximize human potential in a world that is in essence chaotic and unpredictable. Like many other pioneers, Wiener helped to initiate a journey that would prove to have consequences more far-reaching and subversive than even his formidable powers of imagination could conceive. As Bateson, Varela, and others would later argue, the noise crashes within as well as without. The chaotic, unpredictable nature of complex dynamics implies that subjectivity is emergent rather than given, distributed rather than located solely in consciousness, emerging from and integrated into a chaotic world rather than occupying a position of mastery and control removed from it. Bruno Latour has argued that we have never been modern; the seriated history of cybernetics—emerging from networks at once materially real, socially regulated, and discursively constructed—suggests, for similar reasons, that we have always been posthuman.<sup>23</sup> The purpose of this book has been to chronicle the journeys that have made this realization possible. If the three stories told here—how information lost its body, how the cyborg was constructed in the postwar years as technological artifact and cultural icon, and how the human became the posthuman—have at times seemed to present the posthuman as a transformation to be feared and abhorred rather than welcomed and embraced, that reaction has everything to do with how the posthuman is constructed and understood. The best possible time to contest for what the posthuman means is now, before the trains of thought it embodies have been laid down so firmly that it would take dynamite to change them.<sup>24</sup> Although some current versions of the posthuman point toward the anti-human and the apocalyptic, we can craft others that will be conducive to the long-range survival of humans and of the other life-forms, biological and artificial, with whom we share the planet and ourselves.