

Interdisciplinary Humanities

Publication of the Humanities Education and Research Association

Interdisciplinary Humanities
Volume 35.1 Spring 2018



Executive Editor

Ronald Weber
University of Texas at El Paso

Guest Editor

Dore Ripely
California State University, East Bay

Co- Editor

Stephen Husarik
University of Arkansas - Fort Smith

Co- Editor

Lee Ann Elliott Westman
Rutgers University - Camden

Book Review Editor

Edmund Cueva
University of Houston - Downtown

Published by the Humanities Education and Research Association (HERA) at The University of Texas at El Paso. Individual Membership in HERA includes a one-year subscription to *Interdisciplinary Humanities* (spring, summer, and fall issues) and costs \$130. Libraries and other institutions may obtain one-year subscriptions to *Interdisciplinary Humanities* for \$167. Membership forms may be found at the back of this journal and at http://www.h-e-r-a.org/hera_join.htm.

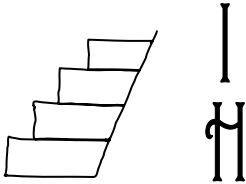
The editorial offices of *Interdisciplinary Humanities* are located at The University of Texas at El Paso. Contributors should consult carefully the Editorial Policy Guidelines (see back pages of this issue). Editorial decisions for publications are normally made within four months of the submission deadline. Send an electronic copy of essays for publication consideration to co-editors Stephen Husarik, shusarik@uafs.edu, and Lee Ann Elliott Westman, leeann.westman@rutgers.edu.

Interdisciplinary Humanities is indexed by ERIC ISSN 1056-6139
© 2018 Humanities Education and Research Association

Cover Art: *Incerta* by Steven Yu, 2018 (Digital) www.cargocollective.com/stevenyu

This issue of *Interdisciplinary Humanities* is partially underwritten by the Humanities Program at The University of Texas at El Paso.

**Organic Machines/Engineered
Humans: (Re)Defining Humanity**



Contents

Editor's Introduction: From Text to Tech: (Re)Defining Humanities in a Post-human Age

Doré Ripley *California State University, East Bay* 3

The Transhumanist Future and Fiction: An Interview with M.T. Anderson

Doré Ripley *California State University, East Bay* 18

Transhumanism, Optimism, and Enlightenment Optimalism

Matthew Landers *Western Norway University* 25

Beyond/Less Than: The Potential Implications of Transhumanist Rhetoric on Human Exploitation

John Gagnon *University of Hawaii at Manoa* 41

The Future of Transhumanism as a Growing Movement: An Interview with Zoltan Istvan

Doré Ripley *California State University, East Bay* 54

Redefining Humanness: Relationality, Responsibility and Hope in *Ex Machina*

Tracy L. Hawkins *University of Wisconsin-Whitewater*
Kirsten Gerdes *Riverside City College* 63

Mechanical Dreams of Humanity: When Machines Long for Humanity

Caroline Mosser *Independent Scholar* 78

Contemporary Pygmalion Tales: Robot *Sapiens Sapiens*

Maria Aline Ferreira *University of Aveiro (Portugal)* 91

The Postmodern Prometheus: Humanity and Narration in the SF Worlds of Dick's *Do Androids Dream of Electric Sheep?* and Scott's *Blade Runner*

Andrew Howard *University of the Dist. of Columbia Community College* 108

Galateas Gone Wild: Technology, the Feminine, and Spatial Divides		
Cary Elza	<i>University of Wisconsin-Stevens Point</i>	121
Notes on Contributors		137
Editorial Policies		140
Membership Application		141

Editor's Introduction

From Text to Tech: (Re)Defining Humanities in a Post-human Age

Doré Ripley

California State University, East Bay

Within thirty years, we will have the technological means to create superhuman intelligence. Shortly after, the human era will be ended . . . I think it's fair to call this event a singularity. It is a point where our models must be discarded and a new reality rules. As we move closer and closer to this point, it will loom vaster and vaster over human affairs till the notion becomes a commonplace. Yet when it finally happens it may still be a great surprise and a greater unknown.
~Vernor Vinge¹

Transhumanism is the philosophy that espouses using genomics, robotics, informatics, nanotech, new pharma . . . to change humanity into a new species.
~Richard A. Clarke²

Humans have been tinkering with technology since they could fashion sticks into spears and harness fire. Over the last six million years we have gone from small-brained bipeds to intelligent beings with the capacity to enhance, or speed up, human evolution through the use of various forms of technology. But before we can build perfect, albeit ersatz, humans, we have to dream, and that leads to creativity and storytelling.

This is the stuff of science fiction.

Some early attempts at creating artificial man include a Germanic knight crafted by Leonardo Da Vinci (1495) for the Duke of Milan and controlled by pulleys and cranks. The clock maker, Juanelo Turriano (1560s), created a mechanical monk (now housed at the Smithsonian) that was ordered by King Phillip II of Spain as a tribute to God for saving his son's life.³ Both of these early commissions pay tribute to the spiritual as well as the creative whims of man. An early literary automaton appears in E.T.A. Hoffman's "The Sandman" (1816), and later in Offenbach's related opera, *Tales of Hoffmann* (1877), and reflects male visions and desires for the perfect female. As early as the 1920s, Karel Čapek's play *R.U.R (Rossum's Universal Robots)* examines the creation of proto-humans from bio material to do the work of mankind and eventually fuels a cyborg rebellion.⁴ In the mid-20th century, Isaac Asimov

explored the human/machine interface creating The Three Laws of Robotics which text and tech creators utilize today.

Law 1: A robot may not injure a human being or, through inaction, allow a human being to come to harm.

Law 2: A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

Law 3: A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.⁵

By the 1960s Philip K. Dick's *Do Androids Dream of Electric Sheep?* (1968) envisioned organic machines that had grown beyond the three laws to question their own humanity or inhumanity. In the 1980s, William Gibson's *Neuromancer* (1984) jacked the literary world into a cyberpunk console as an internet cowboy is uploaded and downloaded into extended cyberspace, while being manipulated by an AI to illegally merge its two entities into one super-being. Through text and tech, human/robot relations are explored as writers and scientists try to envision or create cognitive machines, ultimately leading to the ever-present question: What does it mean to be human?

One of the interviews in this issue features M.T. Anderson, author of *Feed* (2002), a young adult work predicting today's rush towards universal internet implants; a book where teenagers are routinely chipped, go to School™, are bombarded by ad banners, and consume with a gusto that makes fashions change at lightning speed, but ultimately, those feeds get hacked. As an author of both fiction and non-fiction—children, young adult, and adult works—Anderson is curious about AI, transhumanism, the arts, and intelligence. He predicts advances in AI will lead to more interactive games and literature, as he wonders “What are we in *literature* for? Are we in for someone else's story that is read, received, understood; or someone else's environment that is discovered, reacted to, engaged in?” Anderson is looking ahead to when the arts will become “more total and holistic in their experience.” His latest book project is concerned with the issues of culture and transhumanism and what it's like to be at the end of one's life in the midst of this technological cataract, exploring what technological change is like for those left behind. Some of us may live long enough to find out.

Anderson also believes enhanced intelligence is just around the corner, and only our biological wetware will hold that back. When we eventually get it all sorted out, we will probably have to redefine humanity. Are people who are totally disembodied and uploaded still human? Can intelligent androids be human? These are questions we will struggle with for years. On the other side of the screen, it could lead to a freeing of conceptual identities where one could choose genetic splicing/gene interaction. If you have the money you can become who/whatever you want. It's not technology that will lead the way,

but capitalism. As Anderson believes, “our technological development has far, far outstripped our ethical and moral development.” Having access to more data does not make us more knowledgeable or wise.

Warren Ellis explored conceptual identity in *Transmetropolitan* (1997-2002), a series that follows gonzo journalist, Spider Jerusalem, as he documents his way through clones, hybrids, cultural reservations, and all varieties of humans to fight corruption while treating readers to a circus of the absurd in near future transhumanism. This prescient comic series investigates the lives of the cryogenically defrosted, ethereal personality downloads, transcience transients, and a myriad of other gene-spliced-to-order humans making human interaction tricky at best in a post-biological culture where almost any subjective identity becomes reality for one, some, or many. The ability to splice, manipulate, or recreate one’s identity leads to a narcissistic world filled with confusion, alienation, and even reversion. Ellis explores how citizens of all varieties would live (or escape) such a society in its rush towards ever-expanding techno heights.

Alongside Anderson, Ellis, Dick, and Gibson are many, many other authors (re)defining humanity, society, and the arts. Vernor Vinge’s *Rainbows End* (2006), streams into a society dominated by augmented reality as a technophobic poet recovers from Alzheimer’s and tries to (re)connect with his family within a virtual society. Iain M. Banks in *The Hydrogen Sonata* (2012) drags his bow across the arts asking

What was the point of taking the time learning to play anything as well as you could, when a machine could use something it would think of little better than its hand puppet to play so achingly, immaculately, ravishingly well, exactly as though *it* was the creature that had spent a lifetime studying, understanding and empathizing with the instrument and all that it signified and meant?⁶

Today, algorithms are being developed to produce stylistic music à la Coltrane, Bach, or Joplin, leading some to ask, “But can they swing? ‘I would submit that you can certainly make a computer swing,’ says Brooklyn-based musician and technologist Eric Singer. ‘You can kind of jitter that swing a bit to make it sound more human.’”⁷ Lovers of the humanities throughout history have always asked what is art? Music, poetry, storytelling, building, painting, drawing, dance? Today, we can contemplate what art means when subjectivity is created by an objectively coded machine leading many to wonder what might happen if computers evolve past their programming into thinking, fully-functioning, entities that create true art, but what is true art? We have long created art about AIs, cyborgs, robots, and transhumans and that artistic gaze leads us into the lab where tech imitates art.

Our dreams are the stuff of science fiction and those Technicolor visions take many forms. Robbie the Robot from the *Forbidden Planet* (1956) obeys,

even coddles, humans. On another servo, a few years later the H.A.L 9000 artificially intelligent computer decides to terminate the mission crew in *2001: A Space Odyssey* (1968) forcing audiences to contemplate what it means to build machines that are far more intelligent than humans, thus making them uncontrollable. The same theme is explored in *The Terminator* (1984 to present) movie series as it terrorizes audiences with organic covered robot/cyborgs hunting humans. Its latest sequel, *Terminator Genisys* (2015), presents humans subsumed both body and mind by nanobots. *The Matrix* (1999) extended William Gibson's original computer network trope into:

The matrix [which] has its roots in primitive arcade games. ... Cyberspace. A consensual hallucination experienced daily by billions of legitimate operators, in every nation ... A graphic representation of data abstracted from banks of every computer in the human system.⁸

Are we just living in an abstraction of an abstraction? Jean Baudrillard opens "Simulacra and Simulations," his philosophical treatise on reality, society, and symbols by saying, "The simulacrum is never that which conceals the truth—it is the truth which conceals that there is none. The simulacrum is true."⁹ We give it that truth by treating it as such and the arts presents that simulacrum on its many computer chips as science fiction moves into reality, making abstraction a projection of the hyperreal.

Artificial intelligence appears early on the small screen with the AI, K.I.T.T, from *Knight Rider* (1982-1986), who inhabits a black Pontiac firebird and fights injustice. The crew of the Enterprise in *Star Trek: The Next Generation* (1987-1994) battles the Borg as they subsume humans into transhumanist components of a collective. At the same time the artificially intelligent Operations Officer, Data, is struggling to become or be seen as human while following the three laws of robotics. In *Battlestar Galactica* (2003 series), the artificially-intelligent bio-organic Cylons rebel against humans. Today, *Altered Carbon*, a 2018 Netflix release (based on the 2002 novel by Richard K. Morgan), offers a post-AI look at a post-death society in a *Blade Runner*-style dystopian film noir. Enhancements are routine and AI upgrades seem irrelevant. Art in this rain-soaked, neon-blinking urbanity is acquired and sold to acquire more cloned skins ensuring one's immortality. Suffice to say from the myriad books and comics, movies and television we are fascinated with what it means to be human by trying to recreate humans found in our mind's eye.

In reality, we stand on the threshold to the government lab as the U.S. military contemplates micro chipping all its personnel while the arts take a look at how enhanced soldiers might exist in society. Richard K. Morgan's *Thirteen* (2007) looks at the military's practice of gene modification where "twisted" soldiers display characteristics long bred out of civilized man, examining the implications for both society and the personal soldier. Genetically altering

human soldiers is tame compared to what is really going on in military labs today, especially in the field of artificial intelligence. In 2016, DARPA ran its computer security Grand Challenge that many believe will naturally lead to artificial intelligence.¹⁰ Not only can we modify bodies, we will modify minds to become better soldiers. In 2017, the grand challenge is judging a collaborative machine-learning competition, which according to DARPA “is expected to both take advantage of recent significant progress in the fields of artificial intelligence and machine learning and also spur new developments in those research domains, with potential applications in other fields where collaborative decision-making is critical.”¹¹ So we aren’t just cloning or genetically modifying soldiers, the military wants to create true artificial intelligence. Didn’t any of these generals see *Terminator*?

While some may argue over fancied or fevered dreams, dreams lead to tech, and people today are dreaming of enhancing their biological software. A 2014 study conducted by Cisco System found approximately one-quarter of the white-collar professionals surveyed “would leap at the chance to get a surgical brain implant that allowed them to instantly link their thoughts to the Internet.”¹² Swedish office workers at Epicenter are chipping themselves for convenience (unlocking doors, security clearance, and banking), and while it does raise security issues for some employees, for many it is just the next logical step.¹³ In 2017, Three Square Market employees were treated to a “chip party” where they were implanted with radio-frequency chips that allow them to open doors and log in to their computers with the wave of a hand. The Wisconsin company, which designs software for vending machines, hopes to lead the way in the market-chip reader industry.¹⁴ Need some Doritos? If you are chipped you won’t even need Apple Pay. Chip implants are just one more step towards transhumanism with other steps currently including gene therapy, cybernetics via cochlear implants, and pace makers. All these advances will augment other technical organic enhancements like autonomous self-replicating robots, nanotechnology, mind uploading, and artificial intelligence. These steps and other breakthroughs will make the singularity a reality in the near future.

The idea of the singularity—when man melds with machine—has been developing over the course of the last century. Vernor Vinge, the Hugo-award winning science fiction author, wrote a white paper for NASA that traces the history of the singularity. He credits John von Neumann with the idea early in the 20th century:

One conversation centered on the ever-accelerating progress of technology and changes in the mode of human life, which gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue.¹⁵

Vinge is just one among many who believe the singularity and transhumanism are doing more than just accelerating technology, they are changing humanity because that technology will be coupled with “the creation of super human intellect.”¹⁶ While most tech people see transhumanism and the melding of man and machine as a giant step forward for mankind, pioneers like Donna Haraway (*A Cyborg Manifesto* 1985) look at technology in a different way:

Late twentieth-century machines have made thoroughly ambiguous the difference between the natural and artificial, mind and body, self-developing and externally designed, and many other distinctions that used to apply to organisms and machines. Our machines are disturbingly lively, and we ourselves frighteningly inert.¹⁷

Rather than a mutual joining there is a dissonance that more and more leaves human subsistence to machines since we would rather leave manual responsibilities to someone(thing) else. Let Roomba clean your floors, and Google Nest change the temp or lock your doors, or even, let your refrigerator send your grocery list to your mobile app through the internet of things.

Non-fiction cyborgs, those organic machines/engineered humans that haunt our sci-fi visions have been around since the mid-twentieth century. In the late 1950s, the first cyborg was a white lab rat outfitted with a tiny osmotic pump to inject it with precisely controlled chemicals. The first pacemaker was implanted in 1958. By the 1970s, we could imagine cyborgs living peacefully among us with popular television shows like *The Six Million Dollar Man* and *The Bionic Woman*.¹⁸ In the 1980s we saw the rise of pacemakers and cochlear implants. Today, cyborgs abound. According to Haraway we are all cyborgs living with and through our technology. Can you go a day without your cell phone? Can you navigate your way home without Siri? Can you imagine a day when you won't have to carry your phone because it will just be a regularly maintained medical upgrade? Many people can.

Proponents of transhumanism or H+, such as Peter Diamandis, founder and chairman of the XPRIZE Foundation and Human Longevity Inc., believe the rise of technology will lead to cheaper autos, homes, food, education, and medicine. All these developments will, in turn, lead to transhumanism technology, such as nano and bio-tech to cure disease, aging, and expand our consciousness. Diamandis also believes there will be a demonetization of the economy which is already underway. No longer do we pay for photographs, maps, or data research thanks to that powerful little computer we call our phones. Transportation costs are decreasing with the rise of companies like Uber and Lyft. Food costs have decreased by 50 percent since 1960.¹⁹ He describes a virtual utopia and believes that technological advances will far outweigh the risks and he is not alone. A recent survey shows that most Americans believe AI will improve their lives even though it will eliminate jobs. Surprisingly, those same Americans also believe it will be someone else's

job loss, not their own. Almost 50 percent of the 3,300 respondents favored some form of universal basic income.²⁰ Universal basic income, according to pro-transhumanists, will leave people with free time to create and follow their dreams.

On the other side of the circuit board, Donna Haraway believes that “as robotics and related technologies put men out of work in ‘developed’ countries and exacerbate the failure to generate male jobs in Third World ‘development,’ and as the automated office becomes the rule even in labour-surplus countries, the feminization of work intensifies.”²¹ That feminization leads to lower wages and the re-emergence of home sweatshops, not leaving much time for artful creation. One thing is clear, in the beginning of the singularity and transhumanism your health insurance isn’t going to cover gene-splicing and nano-bot surgeons. So it may be that the rich will get the first cyber benefits until the costs reduce and/or demonetize, or that human exploitation may take yet another turn for the worse. The utopic visions of people like Ray Kurzweil (the singularity guru) and Diamandis, the so-called leaders of the transhumanist revolution, do not mesh with many of the dreams presented by text creators. The 2013 film *Elysium* presents an overcrowded earth abandoned to the poor where human exploitation is routine and the rich luxuriate in a space habitat filled with food and free health care. Utopia, in this case, is only for the rich. Tech creators are not the erstwhile Dr. Frankenstein in his lab, they have utopian dreams, but it is the arts that likes to play with imagined outcomes—good and bad.

Another trouble with tech and texts is not the lack of the feminine, but the lack of the feminist. At the Arts Electronica Festival conference in Austria, the AI robot, Samantha, programmed to respond to *romance* was treated “barbarically,” leaving her with broken fingers and “heavily soiled.”²² Samantha’s demise was not some aberration at a small-time tech convention. At the 2018 CES (Consumer Technology Association) convention in Las Vegas, the world’s largest tech convention, robot strippers danced at a nearby club as attendees came and went through the public space. Who needs the booth babes of traditional trade shows when you can lure in customers with *Saga*-like (2012-Present) robot pole dancers. Obviously, sexism is rife in the tech industry with several company founders forced to step down or apologize for the frat-like cultures prevalent at many of Silicon Valley’s tech companies. Only 20 percent of this year’s attendees at CES were women with only two out of 15 keynote speakers being women, an industry problem since keynote speakers must be company CEOs.²³ Sexism is the culture and it is prevalent all throughout the industry. As one woman walked across the convention floor, she said “I see robotic strippers and I see half-naked women on the showroom floor promoting products. It’s like, aren’t we worth more than that?”²⁴ This behavior is reminiscent of the magical ring of Gyges, allowing one to become invisible in order to get away with anything no matter how unjust. Today the magical tech world practices invisible and not-so-

invisible abhorrent behavior in the silicon land of invisible women where automatized women take center stage.

The mistreatment of women isn't just a lack of vision on the part of entrepreneurs; many sci-fi texts treat women as little more than brainless statues, robots, or skin covered AI, as some of this issue's authors have pointed out. It is disappointing that today's tech has created more lifelike sexbots than nannies for grannies or children or other companions and attendants. It has been argued that creating female sex robots will reduce sex crimes, but on the contrary, it just looks like another way to objectify women. While one can set up this hypothetical, no studies have been done . . . yet. Today, a Roxy TrueCompanion robot costs around \$10,000. The different models include the adventurous Wild Wendy, S&M Susan, and Mature Martha and all models "can even have an orgasm." There is also Frigid Farah who does not like to "engage in intimate activities," which sounds like an invitation for purveyors to rape. "Young Yoko is very naïve" and "barely 18."²⁵ Some have proposed that these bots can help with sex therapy, but a Johns Hopkins paraphilia researcher "does not believe there will ever be a therapeutic use for sex robots, suggesting that such contact will have a 'reinforcing effect' on behavior that leads to sex crimes like sexual assault and pedophilia."²⁶

This tech is not far away from text, just look at an episode of *Altered Carbon* (2017) where human sex workers volunteer for snuff sex and are then promised a new and improved body, or skin, for their trouble. To be fair, there are those who are creating human-looking and acting androids today. Firms such as Hanson Robotics has produced an Albert Einstein HUBO and Philip K. Dick android. Hanson explains an intention

to push the PKD android until it evolves super-human creativity and wisdom and transcends in a spiral of self-reinventing super-intelligence—what Philip K. Dick precognitively described as a Vast Active Living Intelligence System [VALIS], and what Vernor Vinge describes as the Technological Singularity. We predict this will occur sometime between 15 and 30 years from now.²⁷

Moving from sex toy to super intelligent machine is a long dystopic distance from the androids of Philip K. Dick's *Do Androids Dream of Electric Sheep?* (1968), androids that are plugged into the singularity where man meets machine, machines that question their own humanity.

Between robot, AI, and transhumanism there is the singularity. A transhumanist optimist, Ray Kurzweil, believes that man and machine will eventually fuse and this singularity will allow people to enhance their physical and mental bodies creating super beings—but you have to live 2045.²⁸ In order to achieve this goal, Kurzweil currently takes vast quantities supplements every day, eats well, drinks ionized water, and exercises. At "60 years old, he reportedly has the physiology of a man 20 years younger."²⁹ But the human

body, no matter how well you take care of it, is susceptible to illness, disease, and senescence—the process of cellular change in the body that results in that little thing we all do called aging. This cellular process is why humans are physiologically unable to live past the age of around 125 years old. Kurzweil is well aware of this, and he has a solution. He has to live long enough in his human body until technology reaches the point where man can meld with machine, allowing him to live as a cyborg with “robotically-enhanced features.” In other words, he needs to survive

until the day when he can eventually upload his consciousness onto a hard drive, enabling him to “live forever” as bits of information stored indefinitely; immortal, in a sense, as long as he has a copy of himself in case the computer fails.³⁰

Essentially, Kurzweil will need to survive inside a computer until tech catches up to text.

Before we can upload our consciousness into machines, we must invent a way to meld the human consciousness with machines. This concept, sometimes called “neural lace,” was explored by Iain M. Banks in his *Culture* (1987-2012) novels. In these novels, Banks refers to a “neural lace” as a “mesh-like device which would be implanted in a person directly through the bloodstream, controlling the release of certain neurons using the power of thought.”³¹ Along that same circuit, Elon Musk is doing research to come up with a way to kickstart transhumanism by creating a neural lace that melds the human brain with technology. His neural lace is a mesh of electronics that will allow AI and the brain to work together enabling human brains to keep up with future enhancements. Musk’s version of the neural lace does not share the same functionality as that of Banks,³² instead it would allow AI to work symbiotically with the human brain through signals picked up and transmitted wirelessly, without any interference of natural neurological processes, essentially making it a digital brain upgrade. Humanity could send emails and write texts just by using our thoughts.³³ While it is one thing to enhance the human brain, it is something entirely different to create artificial intelligence. Musk thinks artificial intelligence is one technological idea we should seriously contemplate before blasting ahead. In fact, he’s so worried that he is spending billions to rocket towards Mars in order to colonize the red planet just in case our AI overlords decide humans aren’t worth the effort.³⁴ But for the most part, transhumanists and scientists envision a near-perfect future where poverty and illness have been eliminated.

One transhumanist featured in this edition is Zoltan Istvan, a transhumanist who ran for the presidency in 2016. His 2016 campaign came in sixth in media awareness and he is looking forward to running as governor for the state of California. He wants to use the California election as a platform from which to launch a Libertarian run at the presidency in 2020. Istvan presented a transhumanist bill of rights to Washington stating, in part, that we

will create advanced forms of sentient life and they should be treated with respect. All sentient life is entitled to universal rights—including ending suffering, personal enhancements and extended lifespan. No law or religion can prevent someone from such improvements. Sentient life should also have morphological freedom. In other words, one can do whatever one wants with its life form so long as hurting others is avoided. We should prevent existential risk when possible to avoid mass destruction. Conversely, we should encourage space travel in case we destroy our planet. Finally, involuntary aging shall be classified as a disease.³⁵ One of the first steps on the transhumanist path is the use of neural prosthetics, which will lead to the singularity when man meshes with machine; leading back to the question, “When will we be transhuman?” while at the same time alluding to “What does it mean to be human?” Istvan believes if a being, including humans, transhumans, and AI, can express a desire for personhood and “can intelligently express that, then we should protect that.” Questions of entity and identity make one wonder how smart we should make AI? Istvan believes, not smart enough to know they don’t want us around, just smart enough to execute all the mundane functions they can be programmed to perform.

Istvan also believes people (entities) may just become pure data or pure AI with their own dreams. The arts will transform because machines will not be able to create more subjectively beautiful artworks, however some tech and text creators would disagree with him, as in the case of the algorithms that swing and the imagined world of Iain M. Banks’ *The Hydrogen Sonata*. Artistic human exceptionalism may just be a pipe dream at some point in this neural mesh, AI, transhumanist future. As entertainment goes, virtual reality and augmented reality will merge leaving television in the e-waste collection bin of history.

The singularity and the rise of transhumanism present unique ethical questions that go beyond tinkering with the human genome. Robotic limb replacement and gene therapy for debilitating diseases makes sense, but should we create designer babies? Should we upload our brains into computers in order to live forever? Should we enhance our bodies to fight aging and disease? Should we speed up, or change the trajectory, of evolution? Further, who gets access to this science? The rich? Certain nations (democratic, autocratic, dictatorship, republic, communist, otherwise)? Individuals in certain job fields? The essays in this edition look at transhumanist thought and principles and their growth and relationship to various schools of philosophies, while some attempt to (re)define humanity in this technologically-saturated society. Others look at how the arts reflect the tension that defining an as yet unknowable future suggests from our ancient mythological and philosophical roots to our postmodern science fiction future and the arts.

Matthew Landers examines the influence of Enlightenment Optimism/Optimalism on some strains of Transhumanist thought in his paper “Transhumanism, Optimism, and Enlightenment Optimalism.” Building on a historical examination of the Optimalist vision of Enlightenment figures, such

as d’Alembert and Condorcet, his paper questions the suppositional grounds of Transhumanist optimalism, both as it relates to the fundamental goals of cognitive enhancement, and to the engineering of a posthuman “superintelligence.” Making use of research in cognitive science, evolutionary psychology, biology, and human intelligence, his paper argues that enhancements to cognition are not as straight-forward as we may imagine. Lastly, paying special attention to the influence of bias on human decision-making, Lander’s paper draws into question the optimistic idea that the human brain can be optimized.

John Gagnon’s “Beyond/Less Than: The Potential Implications of Transhumanist Rhetoric on Human Exploitation” asks what happens when technological progression edges into a new sphere, when the word “human” itself connotes with something “less than?” His article leads with that question as a way into exploring the implications of transhumanist rhetorical framing human enhancement technologies. The implications of such developments on how we conceive of “humanity” stand as a central concern and potentially alter our understanding of what is “human.” Gagnon examines how rhetoric, ethics, and embodiment coalesce in transhumanist discourse in ways that potentially implicate, on a fundamental level, the framework for understanding human exploitation because, at its core, enhancement rhetoric necessarily creates and reinforces a hierarchical structure of embodied identity, where the word “human” itself connotes inferiority. He argues that this presents substantial risk, particularly as it relates to the potential to encourage human exploitation; as such, those of us situated in the humanities hold a responsibility to question the implications of such discourse at the intersection of identity and exploitation.

In their essay “Redefining Humanness: Rationality, Responsibility, and Hope in *Ex Machina*,” Tracy L. Hawkins and Kirsten Gerdes believe a definition of humanness needs to meaningfully distinguish what it is to be human from what it means to be a technological machine or a non-human animal, and it needs to give an adequate account of humanness itself. In pursuit of that definition, they argue that to be human in the contemporary world is to be 1) a subject in relation to other subjects, 2) a responsible agent, and 3) a being who hopes for the future. Importantly and perhaps surprisingly, however, if humanness is defined in this way, Ava and Kyoko, the “machines” from Alex Garland’s 2014 film *Ex Machina*, must be understood as human. While Ava and Kyoko began as technological machines, because of their changing relationality and the emergence of their responsibility and hope for the future, they evolve into beings-in-the-world that are human, and their evolution changes what it means to be human. Hawkins and Gerdes argue that it is necessary to conclude that these women are human to better account for humanness; this is because we must resist definitions of humanness that allow particular types of lives—non-western, non-white, non-male—to be seen as less than human, as they have been historically.

Caroline Mosser believes narratives about robots are often based on the fear of change or on a sense of wonder at their possibilities and explores. Nevertheless, a vast majority of science fiction is focused on the potential dangers stemming from humanoid robots. This anxiety is partly rooted in the Industrial Revolution's mechanization of labor and the replacement of workers by machines. While our species' limited physical abilities have been recognized many times, it was thought to be more than compensated by our superior intelligence and our sense of morality. However, due to the drastic evolution of technology at the beginning of the twentieth century, especially during World War One and World War Two, and the development of artificial intelligence, the competition between human and mechanical labor slowly turned into a challenge of human exceptionalism.

There is, however, another type of robot narrative, one that focuses on the essence of robots rather than their function or usage which she explores in her essay, "Mechanical Dreams of Humanity: When Machines Long for Humanity." Focusing on how the narrative of the robot's quest for humanity also attempts to define humanity as Mosser analyzes the robot's quest towards humanity in Asimov's *The Bicentennial Man* and Zelazny's "For a Breath I Tarry." Both Asimov's android and Zelazny's intelligent computer strive not only to achieve the status of man and therefore their recognition as persons, but also rather to become truly human.

Aline Ferreira's "Contemporary Pygmalion Tales: Robot *Sapiens Sapiens*?" reflects on potential future configurations of the human/robot relationship, mainly through the lens of fiction. Ferreira compares and contrasts Marge Piercy's *He, She, and It* (1991) and Jeanette Winterson's *The Stone Gods* (2007), two novels with many thematic similarities and a special focus on the romance between the protagonists and the robots they have helped to program. The main focus of her analysis is the interplay between cognition and emotion, drawing on recent studies in the field of neurosciences. Referring to HAL in Stanley Kubrick's *2001: A Space Odyssey* (1968), Rosalind Picard argues that hope for the future of a humanized HAL depends on the feasibility of teaching machines how to feel (2000), since emotion is a crucial aspect of intelligence.

Another feature of the robots' programming that she investigates is the influence of gender-inflected data in their programming on their cognitive development and the ways in which robots relate to the world, particular care being invested in the effort to prevent, for instance, the over-masculinization of those data. Indeed, the great majority of what Ferreira calls Pygmalion tales, going back to the story of Pygmalion and Galatea recounted in Ovid's *Metamorphoses*, have featured male characters bringing female statues to life and educating young women, shaping them to their makers' desires, as in Jean-Jacques Rousseau's play *Pygmalion* (1762) or George Bernard Shaw's *Pygmalion* (1912). On the other hand, Piercy's and Winterson's novels are crucially structured around primal scenes of the education of the newly created robotic creatures, where the traditional male scene of creation and instruction is subverted, replaced by a feminine point of view.

These texts thus open up new spaces for developing innovative relational dynamics not confined to androcentric models in their revisions of the Pygmalion myth, a founding story of male myths of creation, suggesting that increasingly organic robots with evolving emotions are an inevitable feature of a (post)human future.

Andrew Howard's "The Postmodern Prometheus: Humanity and Narration in the SF Worlds of Dick's *Do Androids Dream of Electric Sheep?* and Scott's *Blade Runner*" explores the intersection between humanity and automation in Phillip K. Dick's 1968 novel *Do Androids Dream of Electric Sheep?* and its 1982 film adaptation *Blade Runner*, directed by Ridley Scott. In both the novel and film, empathy is used to determine a quantifiable sense of humanity, yet in both, characters show varying degrees of empathy despite their alleged humanity. The result is a trenchant exploration not only of what it means to be human, but also what happens when we give ourselves the power to define humanity.

Cary Elza's essay, "Galateas Gone Wild: Technology, the Feminine, and Spatial Divides" interrogates how two narratives that depict the creation and control of artificial female figures respond to and participate in a much larger concern that preoccupied late nineteenth and early twentieth century society, and indeed continues to the present day: the struggle to balance the world mankind envisions and constructs in the public sphere—commerce, technology, the trappings of modernity in general—with an imaginary or spiritual world tethered to the private sphere, often gendered female in popular culture. Auguste Villiers de l'Isle-Adam's novel *The Eve of the Future*, or *Tomorrow's Eve* (1886) and Fritz Lang and Thea von Harbou's film *Metropolis* (1926), which both focus on a male author's creation of a female robot, update the myth of Pygmalion and Galatea for new generations, and in so doing, reveal broader social concerns about women's mobility in the public sphere, the corrupting and rationalizing effects of modernity, and the potential for moral and spiritual decay in the face of new technologies and consumer culture.

What does it mean to be human? Where do you stop being human? . . . Artificial heart, carbon fiber bones, artificial neurons? What is sexuality or gender? Is emotion needed for an intelligent being? What is spirituality? Will we have to redefine humanity or will we have to create a new definition for enlightened entities, including humans? These are questions humanity will have to answer somewhere in the near future and we need those head-in-the-clouds dreamers to lead through storytelling and art: from text to tech. The singularity, where man meets and ultimately interfaces with machine, is predicted to become reality as early as the mid-twenty first century. While some of us may witness that subliming, most of our children will, so let's take some time to think about these questions before we are iced, packed, and digitized.

Notes

- ¹ Vinge, Vernor. "The Coming Technological Singularity: How to Survive in the Post Human Era." Ohio Aerospace Institute: NASA, Vision 21 Symposium, 1993.
- ² Clarke, Richard A. *Breakpoint*. New York: G.P. Putnam's Sons, 2007.
- ³ Andrews, Evans. "7 Early Robots and Automaton." 28 Oct. 2014. www.history.com.
- ⁴ Čapek, Karel. "R.U.R – Project Gutenberg Preprint." <http://preprints.readingroo.ms/RUR/rur.pdf>.
- ⁵ Qtd. in Masari, Heather. "Introduction to Isaac Asimov's 'Liar.'" *Science Fiction Stories and Contexts*. Bedford/St. Martin's, 2015. 141.
- ⁶ Banks, Iain M. *The Hydrogen Sonata*. (New York: Orbit, 2015), 188.
- ⁷ Qtd. in Hochberg, William. "When Robots Write Songs: Bach, Coltrane, McCartney: New algorithms can produce original compositions in the style of the greats. But are those works actually art?" *The Atlantic*. 7 Aug. 2014. www.theatlantic.com.
- ⁸ Gibson, William. *Neuromancer*. Ace, 1984.
- ⁹ Baudrillard, Jean. *Selected Writings*. (Stanford: Stanford UP, 1988), 166-184. 1.
- ¹⁰ Darpa. "The World's First All-Machine Hacking Tournament." 4 Aug. 2016. <http://archive.darpa.mil/cybergrandchallenge/>
- ¹¹ DARPA. "New DARPA Grand Challenge to focus on Spectrum Collaboration." 23 March 2016. <https://www.darpa.mil/news-events/2016-03-23>
- ¹² Cisco. "2014 Connected World Technology Final Report." www.cisco.com.
- ¹³ Associated Press. "Cyborgs at Work: Swedish Employees getting implanted with microchips." *The Telegraph*. 4 April 2017. www.telegraph.co.uk
- ¹⁴ Paquette, Daniel. "Some feared hackers and the devil. Others got microchipped." *The Washington Post*. 1 Aug. 2017. www.washingtonpost.com.
- ¹⁵ Vinge, Vernor. Ibid.
- ¹⁶ Ibid.
- ¹⁷ Haraway, Donna. "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century." *Simians, Cyborgs and Women: The Reinvention of Nature*. (New York: Routledge, 1991), 149-181.
- ¹⁸ Kunzru, Hari. "You are a Cyborg." *Wired*. 02 Feb 1997. www.wired.com.
- ¹⁹ Diamadis, Peter. "Demonetized Cost of Living." <http://peterdiamadis.tumblr.com>
- ²⁰ Carlson, Scott. "Most Americans Think AI Will Improve Lives and Eliminate Jobs—Just Not Theirs." *The Chronicle of Higher Education*. 31 Jan. 2018. www.chronicle.com.
- ²¹ Haraway, Donna. Ibid. 305-06.
- ²² Nichols, Greg. "Sex Robot Molested, Destroyed at electronics Show." *ZDNet*. 2 Oct. 2017. www.zdnet.com
- ²³ Batchelor Warnke, Melissa. "In the tech industry, sexism is in the water." *Los Angeles Times*. 9 Aug. 2017. www.latimes.com
- ²⁴ Associated Press. "What robot strippers say about sexism, tech and the future." *ABC News*. 11 Jan. 2018. www.abcnews.go.com
- ²⁵ Roxy True Companion.com. "World's First Sex Robot." <http://www.truecompanion.com/>.
- ²⁶ Nichols, Greg. Ibid.
- ²⁷ Hanson Robotics. 2017. "Philip K. Dick Android." www.hansonrobotics.com.

- ²⁸ Galeon, Dom and Christianna Reedy. "Kurzweil Claims that the Singularity will Happen by 2045." *Futurism*. 05 Oct. 2017. www.futurism.com
- ²⁹ Sherman, Mark. "Transhumanism and Gray Goo." *Psychology Today*. 4 June 2014. www.psychologytoday.com.
- ³⁰ "Immortality, Transhumanism, and Ray Kurzweil's Singularity," *The Technological Citizen*. 16 Feb. 2010. www.thetechnologicalcitizen.com
- ³¹ Sarich, Christina. "Elon Musk Promises Human to Human (and to machine) Telepathy with Neuralink." *The Mind Unleashed*. Apr. 25, 2017. www.themindunleashed.com.
- ³² Makam, Shobith. "Elon Musk's Neural Lace: A Step Towards Transhumanism." *The Transhumanist News*. 26 Sept. 2016. www.thetranshumanist.news.
- ³³ Statt, Nick. "Elon Musk Launches Neuralink, a venture to merge the human brain with AI." *Verge*. 27 March 2017. www.thverge.com
- ³⁴ Dowd, Maureen. "Elon Musk's Billion Dollar Crusade to Stop the A.I. Apocalypse." *Vanity Fair*. April 2017. www.vanityfair.com
- ³⁵ Istvan, Zoltan and the U.S. Transhumanist Party. "Transhumanist Bill of Rights." *U.S. Transhumanist Party*. 14 Dec. 2015 and 25-31, Dec. 2016. www.transhumanist-party.org.

Arts, Humanity, and the Transhumanist Future: An Interview with M.T. Anderson

Doré Ripley

California State University, East Bay

M. T. Anderson writes for both young people and adults. His satirical sci-fi novel *Feed* was a Finalist for the National Book Award and the winner of the L.A. Times Book Prize. His Gothic novel of scientific experimentation during the Revolution, *The Pax Party*, won the National Book Award in 2006. His nonfiction book about World War II, *Symphony for the City of the Dead: Dmitri Shostakovich and the Siege of Leningrad*, was named a best book of 2015 by the *Wall Street Journal*, the *New York Times*, the *San Francisco Chronicle*, and others. His most recent book is a graphic novel, *Yvain*. This interview was recorded in April 2017.

D.R.: How do you see literature progressing alongside computing and AI? What kind of literature do you think will be popular in a transhumanist future?

M.T.A.: I expect that literature and the other arts will become increasingly interactive, immersive, and experiential, both for modified and unmodified humans. Game spaces as they're currently being developed point us toward some fascinating possibilities. At the moment, the point-of-view narratives we call "games" lack a certain literary substance for only two reasons: One of them is the limit on dialogue and even character flexibility given the limits of current tech's language algorithms. We've come a long way since the text adventures of my youth, when everything had to be parsed in a very particular order, using a very small canon of words. ("GET LAMP. LIGHT LAMP.") But we still can't produce characters who, in the midst of dialogue, pass the Turing test. They can have pre-set mannerisms and attitudes, but when it really comes down to it, the dialogue still feels as stiff as porn. ("Hello. I am here to fix your sink. Why, Mrs. Robinson, you have removed your jumpsuit.")

Once that hurdle has been overcome, there is a much more wide-ranging and fascinating barrier to our culture understanding the immersive "game" environment as literature: We don't yet know what to do with narratives in which the main character, the protagonist, is a blank, a cipher, to be filled in by

the user. Our literature—and I include visual narratives in this category too—is specifically about being privy to the actions of a protagonist who acts differently from us. That’s the great power of literature, the great boon, the great adventure—to focus us on someone else’s habits of thought, extreme reactions, someone else’s sorrows, joys, ways of being. To watch someone who surprises us. Meanwhile, first-person interactive narratives literally put us in someone else’s shoes, but really, those are still our feet.

I guess what I’m saying is that part of the shift is going to take place not merely in the richness and functionality of virtual worlds, but in our own perception of what we’re in “literature” for. In a game platform that is already inhabited by hordes of user avatars, the first problem I talked about above already evaporates. People can interact almost as richly as they do in real life. Yet that still doesn’t provide us with the literary experience as traditionally defined—one based around watching a protagonist.

So maybe our assumptions about what “literature” can offer in this instance needs to expand. Perhaps the take-away from a new literary artifact needs to be not the textured quality of pre-ordained characters and their interactions, but the quality and complexity of the world itself and how it forces its users to react. In this way, it could still teach us new things about ourselves. (Indeed, games already do.) This will be a new way of understanding literature, perhaps even more removed from our current notion of reading in its effects, goals, and social function than Classical epic is or Aeschylean drama.

D.R.: Scientists are exploring the idea of enhancing intelligence through gene selection, however why bother when you can just add a chip? How far do you think people are willing to go for a genius IQ? Will this simply be accomplished with a plug in?

M.T.A.: Rapidity of thought and ease of memory is something many people will be willing to pay anything for, much like extended lifespan. While I’m sure it will eventually be accomplished with a plug-in, we should remember how far away we are from being able to decode people’s engrams—which are a private hieroglyphic language developed separately and uniquely in each brain. Any extension to that material will have to learn to recognize each user’s incredibly detailed and individualized set of signals. So I would imagine that we will first engage in more externalized versions of memory, say. For example, we could re-view things we’ve seen, pages we’ve read, and can search them for specific images or words. That wouldn’t feel the same as a literal memory extension, but would act a lot like one. Here’s the kicker, however: it would still be limited by your own wetware computational time. And this is where I feel like things get complicated, in that different parts of the interface could operate at distinctly different speeds—with the biological components in many ways coming in second.

D.R.: Bioethicist Linda McDonald Glenn believes “humanity-plus is inevitable. We are, by our nature, tinkerers.” In *Feed* you play with this idea—tinkering with our humanity—how do you think this tinkering will play out culturally?

M.T.A.: For the first time since the period when *homo sapiens* roamed Europe with Neanderthals, mating and murdering, several different species of human will have to learn to coexist. This time, however, the determinant will not be biology but social class, income, consumer status.

D.R.: Elon Musk is very interested in setting up a colony on Mars. In the 1960s, Philip K. Dick wrote about human colonies on Mars and the literature loved by those communities was the pulp sci-fi magazines of the 1950s and 1960s. How might the quest for culture evolve on Mars or any other planet we might populate in the future?

M.T.A.: We can see plenty of examples from human history of how borderlands, settlements, colonies, and outposts produce culture. (Think about Anglo-American culture and its struggle to recognize itself as distinct in the 19th century, for example.) Originally, there’s very little emphasis on cultural production, because settlers are just working to establish a foothold. Cultural forms tend to be borrowed from the imperial center, and at first, the local variants seem just like contingent work-arounds. They’re probably interpreted as crudities. Gradually, however, colonists realize that the exigencies of life in their new environment have produced their own material, their own spirit, their own affective constellation, that is distinct from what has come before. Typically, there is a political moment where it becomes important to define the literature and art of the colony as unique and separate, and for the first time, creators tend to lean on their *differences* from the previously central culture. They begin to demand that there is something of special and intrinsic worth in the cultural recognitions of what was previously the periphery. Suddenly, a Mars settler’s old vlog is transformed, in retrospect, into an artifact that recognizes new modalities in human feeling. We start to quote her thought because it speaks to us. It is referred to as a classic in off-world dramas. People back on Earth shake their heads in condescension, because who cares about what it’s like to give birth in low-gravity? Surely that’s just of anecdotal interest. It’s not as important as the experience of life in bright New York, where people have the time and money to produce really *finished*, really *polished* narratives—things that are moving to *them*. And yet, out there, new voices are raised, saying things that have never been heard before ...

D.R.: You write about alternative realities in *The Game of Sunken Places*. Do you think true virtual reality or even cultural reservations will eventually be popular forms of recreation or escapism? What kind of cultures do you think might be popular?

M.T.A.: Of course, virtual reality is already used for escapist pleasure! Think about all of the people who are just waiting for the world from Avatar to be made into a reality. The theme park is on the way—but there are a lot of guys out there, for reasons that somewhat elude me, who only want to date tall, blue girls with tails.

The question of “cultural reservations” is complicated, because when we think of something like “Westworld”—some virtual reality space where there’s a medieval world and an 1880s badlands world and a Greek or Roman world—we only think of those places and those cultures as unitary because we’ve forgotten how much cultural interchange was embedded in what seem to us like monolithic settings. The Greeks saw the Persians, the Medes, and the Lydians as very foreign, and yet they were constantly in cultural conversation with them and involved in constant cultural exchange. We picture a typical “medieval” fantasy world and think of something that is supposedly distinct and European, but that conceals all of the fissures in cultural homogeneity during the actual European Middle Ages: Spain, for example, which was a cultural entrepôt, relying on exchanges between Muslim, Christian, and Jew; or France and England, which of course provide many of the superficial images of our medieval fantasy, and yet were profoundly influenced by the Middle East. We forget that the sound of medieval music, the form of medieval poetry, and even the knight himself (adapted from the cataphract) had origins in the east, in cultures that are supposedly alien. The kings of England used lions as a principle heraldic device. Not many lions indigenous to Wopney or Stowe-on-the-Wold.

So you can’t have a “cultural reservation” without the illusion of a culture without context, without complication, and without history.

D.R.: Twenty years ago, Warren Ellis in *Transmetropolitan*, toyed with the future of mind-uploading, gene splicing, cryogenics, and cultural reservations. In *Feed*, you predict internal internet connections. What kind of prescient sci-fi would you like to read or write in the future?

M.T.A.: I don’t know about what I’ll be concerned with too far in the future. This fall, however, I’ll be writing a novel about some of the issues of transhumanism that you’re discussing in this issue—the story of a historian of my generation in 2050 or so looking back at her own life and the history of the human race from the vantage point of someone at the end of both. Wondering what it all meant. What I really want to capture, in fact, is the sense of the William Gibson quotation you begin the next question with (one of my favorites). What is technological change like for those left behind? We always focus on the shiny world of the early adopters. What about the rest of us?

D.R.: William Gibson wrote that “the future is already here—it’s just not very evenly distributed yet,” which is something you explore in *Feed*. Futurists see us moving towards a UBI (universal basic income) where

everyone can thrive in the arts. Do you see us moving towards a UBI and do you think this will lead to a blooming in the humanities? Is it possible to evenly distribute the future?

M.T.A.: Sure, the UBI would be a great idea, and perhaps even necessary for much of the population to survive the fundamental paradigm shift in the idea of labor and reward that must accompany high-tech mechanization and the AI revolution. But obviously, if anything, we're moving toward a more brutally divided and stratified society. A Princeton study has already shown that technically, in terms of legislative outcomes, the United States is overwhelmingly a plutocracy at this point, not a democratic republic. I see no credible legislative route to a base-line assured income, even as the labor paradigm central to capitalism crumbles. I would predict instead that we'll deal with ever flimsier and crazier mythologies of who deserves what and why; and of course, more heavy-handed forms of repression to combat those who speak out in desperation; and, as a result, more social instability, more points of eruption, more resource contestation, and more bizarrely dysfunctional ways of responding to need.

D.R.: Philip K. Dick in *Do Androids dream of Electric Sheep?* examines what it means to be human. How does literature help us explain and explore our humanity? How might our definition change in a near future filled with transhumans? When do we stop being humans?

M.T.A.: Answering that final question is an important part of the work of literature—though I expect the answer to change with time, and to change with the enthusiasm of the individual reader for a transhuman experience. Are you still human if the basic mental operations are still being performed by a biologically human brain, however enhanced with plug-ins? Are you still human if you're totally disembodied, uploaded? And yet retaining the propensities of the human? (Though how long, after being disembodied, would you still be interested in retaining the signifiers, say, of the sex act, or eating, or all of the other physicalities which would become metaphorical?) These are some of the questions that fictionalists (not simply writers, but people producing narratives of all kinds) will be asking over the next decades—as well as essayists, scientists, philosophers, and talking head pundits (who by that point will just literally be talking heads).

D.R.: Technology and genetic enhancements could lead to subjective conceptual identities in the sense that you could be whatever you decide to be, multi-sexual, genetic splicing/species interaction (blue eyes with lizard skin or feathers), no longer digesting food, etc. creating H+ individuals that may or not be able to relate to one another. What kind of culture might this create? How might we connect as individuals?

M.T.A.: This seems to me very potentially freeing, a way of extending ourselves beyond our physical frailties and limitations and the accidents of birth. And yet remember that, in a sense, in an online world we simply shift the burden of genealogy from the chemical to the economic. Don't picture a world in which you can be anything. Picture instead a world in which you can *pay* to be anything. That is already what we're heading toward.

When people envision the technological future, they often don't think through its productization, its monetization. The attacks on net neutrality, for example, suggest the dangers that corporate entities still pose for the freedom of access and functionality.

The best way of imagining the future is not to ask where technology is going, but where capitalism is going.

D.R.: You have written several books about music, *Handel, Who Knew what he Liked* and *Strange Mr. Satie: Composer of the Absurd* and *Symphony for the City of the Dead: Dmitri Shostakovich and the Siege of Leningrad*. With CRISPR technology we can basically create a human with many enhanced genetic capabilities. When looking at the arts what traits might we select? Might we create better musicians? In an enhanced cyborg culture what might that even mean?

M.T.A.: First of all, I think the arts, as I said above, will become more total and holistic in their experience—combining, potentially, not just images, music, and words, but also sensations. One of the interesting questions about music—as about dance, sports, and other performance-based entertainments—is what will happen when we no longer have to deal with the limitations of the human animal. Think about all of the giddy headiness, for example, that comes from watching a great musician play a virtuosic passage – classical, rock, jazz, whatever. Doesn't matter. We love that shit. But part of our pleasure is based on the knowledge of limitation, of a set of tendons and phalanges and a motor cortex that can only deal with so much at once. The performer's achievement means something only within the area inscribed by human limitation.

When we listen, say, to the crazy fugues Conlon Nancarrow wrote for player pianos, the virtuosity is in the listening, not in the execution—and that's a very different thing.

D.R.: Futurists are a very optimistic bunch viewing the coming singularity, or AI enhance human, as a boon to humanity. Why do you think this is so? As a creator, what do you think could go wrong with this kind of thinking?

M.T.A.: Honestly, I have no idea why the Kurzweil crowd is so sanguine about the future. Let's leave the threat of competition with independent AI out of the picture for a moment. Here's where I think the problem is: our

technological development has far, far outstripped our ethical and moral development. We are the same skittish apes that we've always been, despite the access to new information we're offered. After 9/11, a lot of politicians got weepy onscreen about how the "war on terror" was a contest between two cultures—a culture of medieval superstition vs. one of advanced scientific thought. Now those same politicians deny all scientific thought on climate change, despite catastrophic dangers to our way of life, and they lament that God is angry at our nation because people with matching pudenda love each other. We have access to unparalleled information about our history, about the varieties of the way human life is led, about other nations, about questions of sunlight and salinity. And yet many people, even those in power, deny it.

We choose to ignore the lessons of history (regarding, for example, the role economic inequity plays in the collapse of empires) and of science. Right now, it feels as if the world, shifting toward an ancient simian model of chest-thumping masculinity, is reacting in an almost pre-cognitive, biological way to the challenges of the new century—like rabbits turning bloodthirsty when their warrens are overcrowded. Social capital and beneficial mutualism is breaking down at all levels. It feels as if our species is sleepwalking toward war, just for the animal thrill of the thing—not even out of calculations of self-interest and strategic superiority. This moment may pass. But it leaves the question: Why do we think that better access to information necessarily creates a culture of knowledge?

This is why culture is important—this is why stories are important, and songs, and poems, and the tales you tell the children you love as they fall asleep at night. Without a culture that values what we know, we will simply know things without wisdom and act without understanding.

We have the tools for salvation or apocalypse.

The choice is ours.

Transhumanism, Optimism, and Enlightenment Optimism

Matthew Landers

Western Norway University

Introduction

In a 2014 interview for *The Leftist Review*, Roland Benedikter, consultant on neurotechnology and neuroethics to the Pentagon, explained the agenda of the Transhumanist movement in the following manner:

Its followers want to go beyond the present human condition. At its core it means to overcome the “natural” limitations inherent in human existence, which is to be born, live relatively short, half-conscious lives, and then die. The supporters of “human enhancement” and “transhumanism” intend to break through these current physical and cognitive (and perhaps even spiritual) barriers.¹

Transhumanism is a movement that can be characterized by a relatively simple, optimistic hypothesis: namely, certain qualities of human existence can be improved so dramatically in the future that we will cease to be “human,” in the traditional sense. Writing about the possibility, Max More, a leading theorist, explains:

Transhumanists regard human nature not as an end in itself, not as perfect, and not as having any claim on our allegiance. Rather, it is just one point along an evolutionary pathway and we can learn to reshape our own nature in ways we deem desirable and valuable. By thoughtfully, carefully, and yet boldly applying technology to ourselves, we can become something no longer accurately described as human—we can become posthuman.²

The goal, according to More, is to develop adequate methods to allow us to transcend the biologically-determined constraints that lock human potentials

into their current spectrum. Although these biological limitations vary in type, encompassing such fundamental capacities as human cognition, life-span, and health-span, one can see that the malefactor, the feature holding us back, is biology itself—and, more specifically, the blind processes of evolution by natural selection.

Echoing these sentiments, Zoltan Istvan, a transhumanist politician, novelist, and theorist, argues, “Transhumanists believe we must stand guard against our natural genes, lest they chain us to remaining as animals forever.”³ He takes his assertion a step further, claiming that transhumanist logic leads to the formulation of a new motto, the “Will to Evolution,” by which he refers to a universal human “desire to reach a state of perfect personal power—to be omnipotent in the universe.”⁴ Reaching this state would require our species to seize control of nature’s undirected biological processes, which have no foreseeable ends or goals. In biology’s stead, transhumanists opt for bioengineered solutions and/or technology-based enhancements, which they wish to employ in accordance with anthropocentric laws, to bring about unprecedented improvements to the human condition. According to most transhumanists, these improvements will be measured in transcendent orders of magnitude, resulting in the full transition from a human to a posthuman state.

Chief among the improvements that both transhumanists and posthumanists propose to make are those involving enhancements to human intelligence. In his book, *Superintelligence: Paths, Dangers, Strategies*, Nick Bostrom tries to imagine what *greater-than-human* intelligence might mean, in theoretical terms. For the purposes of the argument that will follow, I am much more interested in the functional significance of an intelligence that will “greatly outperform the best current human minds.”⁵ In particular, what does it mean to outperform human minds? A follow-up question is this: Will cognitive hyper-performance solve the most fundamental problems of human thinking?

These questions take on new significance when we consider two things: first, not only is the brain of *Homo sapiens* the outcome of a long adaptive history of hominin evolution, and second, many of the neural structures present in the human brain evolved long before mammals, or primates emerged.⁶ Over the course of our own natural history, many of these structures underwent adaptive change, taking on new roles at times, and at others losing their original functions altogether.⁷ One of the many conclusions we can draw from the record of adaptive reorganization is that neural structures do not function as an indivisible unit, performing single acts. A great deal of research has concluded that the brain functions modularly;⁸ the work of the brain, which we experience as unified, is carried out by a vast array of individual, adaptive modules. It is assumed that many of these modules (but not all) function in support of a higher-level, integrated function that we call general intelligence (*g*).⁹ However, researchers in human intelligence are not entirely certain that *g* actually exists—except perhaps as a scientific construct, or as a way of speaking broadly, and elusively, about integrated cognitive processes. As

Douglas K. Detterman admits, “Empirically, *g* is well-defined but, theoretically, we only have vague ideas about how to explain it.”¹⁰ Precisely how *g* might be the product of a myriad of cognitive processes is even harder to explain. Detterman notes, “. . . there is no strong evidence for a necessary connection between cognitive processes and *g*.”¹¹

Additionally, within the domain of molecular biology, a hunt is on to identify the specific genes that might produce the *g*-factor; yet, these efforts are not as straightforward as they sound. Researchers have discovered that variance in human intelligence is not determined by a single gene, or even by a few genes. Providing a genetic account of *g* has proven to be quite difficult. This task became much more complicated when it was discovered that certain genes are able to switch on and off in response to environmental factors, causing them to influence biological systems (express themselves) in distinct ways, depending on their different states. In short, not only must we account for the genetic causes of *g*, we must also understand and describe each of the epigenetic factors that influence *g*, if are to hold on to any hope of dramatically improving human intelligence to posthuman levels.

It has become increasingly clear, nonetheless, that intelligence is a purely biological phenomenon. Intelligence researcher Richard Haier writes,

a broad consensus has emerged that intelligence is heritable and polygenetic [influenced by many different genes]. For example, one study based on 3,511 unrelated adults concluded that there are many intelligence genes that all together may account for 40%-50% of variance in general intelligence although no gene yet accounts for more than a tiny portion of variance.¹²

It must be noted, however, that our current understanding of the biological mechanisms that produce *g* is still in its infancy. In his paper, “Increased Intelligence is a Myth (So Far),” Haier speculates that quantifiable improvements to *g* may be possible in the future, notwithstanding the failures of contemporary neuroscience to do so, if researchers are able to make “sophisticated . . . advances in DNA analysis, neuroimaging, psychopharmacology, and even direct brain stimulation.”¹³ He admits, however, that it would be difficult to measure the improvements that he imagines are possible using current testing procedures. Haier is willing to entertain the possibility of limited, perhaps even dramatic improvements to *g*;¹⁴ yet, cognitive enhancements at magnitudes that would be a precondition for posthuman intelligence appear more problematical.

Considering the above points, we are left to wonder, how do we propose to improve general intelligence if, first, complex cognitive processes, like abstract reasoning, rely on many individual neural modules working in concert; second, if we are not certain about how to describe the potential relationship between *g* and cognitive processes; and third, if we cannot provide a sufficient

account of the genetic/epigenetic causes of *g*; and, fourth, if, thus far, no method for dramatically improving intelligence exists? For transhumanists, this is the *hard problem* of intelligence.

If Transhumanism has an Achilles heel, it is the widespread belief in what I call progressive optimalism: the optimistic hope that a natural faculty, such as intelligence, can be improved in progressive steps to an optimal functioning state. It is my assertion that transhumanist theorists who hold to this view have, for a host of reasons, approached the goals of enhancement from an outmoded perspective. In fact, many of the core paradigms that shape Transhumanist understandings of optimization can be traced back to the seventeenth and eighteenth centuries, when it was common to imagine that human faculties operated more like engines that could be fine-tuned. Optimalism was an important feature of Enlightenment thinking, taking many forms over the course of a century. Its influence continues to be felt today.

A number of writers have remarked on similarities between the attitudes and goals of the transhumanist movement and those of the Enlightenment in recent years. Max More goes so far as to assert that:

Transhumanism continues to champion the core of the Enlightenment ideas and ideals—rationality and scientific method, individual rights, the possibility and desirability of progress . . . —while revising and refining them in the light of new knowledge.¹⁵

So far, however, no one has attempted to flush out the exact relationship, beyond outlining a few possible philosophical pedigrees. With this essay, I hope to begin the process of identifying a few of these relationships for future exploration. I will begin by positioning the central claim of this paper, about Enlightenment and Transhumanist optimalism, in its historical context. In the process, I will look for areas of overlap between the two movements, paying special attention to specific examples of progressive optimalism in Enlightenment and Transhumanist thought. It will be necessary to discuss a number of neurobiological issues that complicate the idea of an optimal intelligence. Chief among these is the recurrent problem of bias, which was as much a cause for distress among Enlightenment theorists as it is for transhumanists in the twenty-first century.

Optimism and Progressive Optimalism in Context

In 1686, Gottfried Wilhelm Leibniz published his first major work, the *Discourse on Metaphysics*. In the *Discourse*, Leibniz began a life-long effort to defend the idea that God created the universe in accordance with his own perfect nature, which to say that he created in the most perfect manner possible.¹⁶ Leibniz would repeat this argument in his magnum opus, *Theodicy*, penning the maxim that would later become infamous in Voltaire's *Candide*:

“God has chosen [to create] the best of all possible worlds.”¹⁷ The assertion gave rise to a philosophy of optimism, or more precisely, optimalism, which claimed, from all the possible configurations that God could have chosen, original creation came into existence in an already optimal state. For Leibniz, any claim that nature can be improved upon implies that God acted imperfectly when he created, since to assert that God “could have done his work better is to find fault with it.”¹⁸

While Leibniz’s views enjoyed enduring support among continental Catholics, by the mid-eighteenth century, philosophical optimism became a favorite target of ridicule for a new generation of philosophers. Among the criticisms, Voltaire’s *Candide, ou l’Optimisme* (1759) is the sharpest, intended by the author to deal a final, fatal blow to Leibniz’s arguments. The disagreement was personal for Voltaire. Having witnessed firsthand the near-daily horrors of religious and political persecution in France, Voltaire set out to expose the obvious absurdities of optimist claims. Near the end of his novel, Cacambo, a non-European, asks Candide, “What’s optimism?” to which Candide responds, “it is a mania for saying things are well when one is in hell.”¹⁹

Although Voltaire’s is perhaps an extreme opinion among the *philosophes*, it would be fair to say that his contemporaries saw much that was *wrong* with the world. Thus, we are left to wonder, what influence did optimism have on the Enlightenment, if not the unbending conviction that everything is well? I believe the answer to this question can be found in a unique expression of optimism, which I have called progressive optimalism. It is a conviction held by many eighteenth-century thinkers: namely, that the human condition can be perfected/optimized in progressive steps.

In many ways, the Enlightenment is a period of projects. The underlying motivation for most of these projects is faith in the near-infinite improvability of mankind. No project exemplifies this faith in sustainable human progress quite like Diderot and d’Alembert’s *Encyclopédie*. In the *Preliminary Discourse* to the *Encyclopédie*, d’Alembert proposes a plan to create a philosophical map,²⁰ which will allow mankind to navigate the vast wilderness of knowledge.²¹ He envisioned the *Encyclopédie* as an expanding archive of essays on the latest advancements in philosophy, arguing that the existence of such a dictionary would safeguard human knowledge, allowing mankind to maintain a progressive trajectory. Although d’Alembert concedes that progress is not a guaranteed outcome, he regards the archive as protection against the regressive episodes of human history.²² In short, given the innate potential of human reason, the assertion is that the progress of human understanding is highly probable if certain requirements can be maintained at an optimal state. Preservation of philosophical knowledge is a key factor.

The teleological qualities of d’Alembert’s argument are undeniable. Early in the *Discourse*, he isolates the human capacity for logic as the true source of progress:

This art was found and named Logic. It teaches how to arrange ideas in the most natural order, how to link them together in the most direct sequence, how to break up those which include too large a number of simple ideas, how to view ideas in all their facets, and finally how to present them to others in a form that makes them easy to grasp. This is what constitutes the science of reasoning, which is rightly considered the key to all our knowledge. . . . The art of reasoning is a gift which Nature bestows on her own accord upon men of intelligence²³

Reasoning is itself portrayed as a progression, a logical procedure that advances step-by-step toward understanding. Logic, he argues, is a gift bestowed upon mankind, presumably for a purpose. The obvious implication is that “Nature” has given mankind everything it needs to progress, mostly unimpeded, if we can sustain an optimal environment for learning. Importantly, beyond occasional “revolutions” that work to throw mankind off track, d’Alembert does not seem to see any limits to the progress of understanding.²⁴

I call this philosophical framework progressive optimalism for two reasons. Firstly, progress depends on providing an optimal platform for the experience of learning. Many projects that were undertaken during the Enlightenment sought to optimize the methods and the content of education for this very reason.²⁵ Other projects, like Leibniz’s plan for a universal library, attempted to invent increasingly useful schemes for archiving information. Leibniz’s own idea was to create an archival system that conformed more to the structures of logic, which, he thought, would assist in the progress of understanding.²⁶

Secondly, and more importantly, it was thought that if the learning experience could be optimized, understanding would progress in such a way that mankind would, as a whole, advance toward its own optimal state. At the heart of this conjecture is the anticipation of a phenomenon that we might now call a positive feedback loop. If one can create an optimal learning experience, the “signal” of that experience, amplified throughout one’s educational history, will fine-tune the trajectory of human reason in a progressive direction. Importantly, if this experience can be extended to all of mankind, it stands to reason that humanity can be *perfected*.²⁷

A contemporary of d’Alembert and Diderot, the Marquis de Condorcet expresses a similar point-of-view. In the last chapter of Condorcet’s *Outlines of an Historical View of the Progress of the Human Mind* (1795), the Marquis turns his attention to the future of mankind’s intellectual development. Condorcet begins by claiming that the goal of intellectual progress is the liberation of human reason from the contamination of prejudices.²⁸ He suggests that it is only possible for human reason to exist and operate free from the corruptions of prejudice if the “moral, intellectual and physical” faculties of the mind can

be elevated, by means of education, to a critical horizon, at which point human judgment will yield to the dictates of innate conscience. If these conditions can be met, Condorcet foresees a time when humanity will reach a state of “absolute perfection.”²⁹

His vision of the future is both rationalistic and optimistic, relying completely on the belief that the free and frequent use of reason will defend mankind from all “superstitious fears and chimerical hopes.”³⁰ Although he expresses doubt about the human capacity to comprehend the full “magnitude” of the system of the world, he believes that as we improve the instruments and methods we use in the search for understanding, mankind’s ability to scale down that magnitude to more comprehensible principles will improve as well.³¹ For Condorcet, a philosopher and mathematician, the *instruments* in question are mental systems for the organization of thought, not unlike the logical structures of d’Alembert.³² In a comparable way, then, Condorcet’s claim to “absolute perfection” points to the emergence of an optimal state, when, through a series of “progressive improvement[s],” the mental instruments that we employ have freed reason from the influence of prejudices.³³

The meaning of Condorcet’s “prejudices” is easily recognizable in a contemporary concept: cognitive bias. This is an important connection to make. So much of what we are referring to when we speak vaguely about the improvement of intelligence originates in a desire to correct errors in our decision-making procedures; cognitive biases are perhaps the subtlest agents to influence the decision-making process. In his book, *The Anatomy of Bias*, Jan Lauwereyns describes bias as a kind of force that pulls a decision “in a particular direction, away from the neutral.”³⁴ Such forces were given a variety of names during the Enlightenment—among them, superstition and fanaticism. We can infer from Condorcet’s argument that a perfect intellect, one that has been improved to an optimal state, enjoys a unique freedom from the effects of prejudice. The intellect has been inoculated against bias. We must remember that the seventeenth and eighteenth centuries were centuries marked by unimaginable religious and political unrest and violence. The hope that violence born from prejudice could be minimized, if not eradicated, by the improvement of human reason was widespread. Progressive optimalism provided a much-desired framework for that hope.

In truth, there are at least two strains of progressive optimalism during the Enlightenment. The first strain includes frameworks, such as those of d’Alembert and Condorcet, that are grounded on the potential for indefinite, continuous refinement of human reason (*epistémè*). The second strain includes frameworks that rely more on the role of technologies (*technè*) in enhancing and improving the human understanding. An exemplar of the second framework is the English philosopher, Francis Bacon. Since neither strain is completely exclusive of the other, one can expect some degree of cross-pollination. In fact, readers of the *Encyclopédie* will recognize that Diderot’s own position was more likely a combination of the two. Writing at the very beginning of the

period, Bacon was enormously influential on the *philosophes* of France; yet, the emphatic importance that Bacon places on technology diminishes within the theoretical framework of Condorcet's vision.

Like Condorcet, Francis Bacon expresses concern about the origins of bias—what he calls “anticipations” and “idols”—and more importantly, the influence that each has on the human intellect. The primary focus of the text appears, in fact, to be the elimination of bias from the processes of human reason. Bacon, however, does not seem as confident about the potential for human improvement as Condorcet (or his colleagues working on the *Encyclopédie*). In *Novum organum*, he attempts to refute the idea that “*natural human reason*, left to itself” can bring about progress.³⁵

From Bacon's perspective, the chief problem with human understanding is our tendency to suppose the existence of “order and regularity in the world” where there is no evidence of such.³⁶ He writes:

The human understanding is moved by those things most which strike and enter the mind simultaneously and suddenly, and so fill the imagination; and then it feigns and supposes all other things to be somehow, though it cannot see how, similar to those few things by which it is surrounded.³⁷

This blind striving for homogeneity causes mankind to create what Sperber and Mercier call “folk ontologies,” from which entire philosophical systems can be constructed.³⁸ For Bacon, reason is too easily misdirected, leading to the establishment of a “science as one would,” instead of a science as things are. He argues that man “readily believes” those ideas that he first prefers to be true before ever reaching the level of inquiry.³⁹

Perhaps more fundamentally, Bacon contends that the human tendency toward the creation of *idols* (or biases, of which the above example of confirmation bias is only one example) arises from natural deficiencies in sensory power, concluding that sensation is itself an unsteady foundation, on which to employ reason. He writes:

. . . by far the greatest hindrance and aberration of the human understanding proceeds from the dullness, incompetency, and deceptions of the senses Hence it is that speculation commonly ceases where sight ceases; insomuch that of things invisible there is little or no observation.⁴⁰

Bacon's solution to this problem is not entirely clear. In the preface to the text, he laments over the weakness of “naked intellect,” proposing that “every great work” of mankind “is manifestly impossible” without the use of instruments and machinery.⁴¹ His meaning of *instruments* is decidedly different from Condorcet's denotation. Generally pessimistic about human nature, Bacon suggests that meaningful improvements to understanding are possible only if

we can find ways to enhance basic human perception through technological means. By itself, the human capacity for rationality is not enough.

Writing in the first two decades of the seventeenth-century, Bacon would have been aware of the potential of Galileo's *perspicillum*, a forerunner of the modern telescope. However, he would have had to wait another forty-five years to get a hint of the true investigatory power of the microscope, demonstrated in the plates of Robert Hooke's *Micrographia*—a text that confirmed Bacon's claim that the "subtlety of nature is greater many times over than the subtlety of the senses and understanding . . ." ⁴² We sense in his musings both a flirtation with the exciting possibilities of technological innovation, which promises to overcome the natural limitations of the "Tribe of Man," and a frustration anchored in the awareness that these technologies (especially seventeenth-century optics) still exist in a puerile state. In 1620, Bacon and his contemporaries had only the faintest hint about the true potential of these technologies. It would have been extremely difficult to anticipate the explosion in mechanical engineering that was to come a mere century later, and even more difficult to foresee the seemingly miraculous innovations that would shape the next three centuries.

From Bacon's perspective, the senses create a natural brake on human potential. He argues for a kind of hopefulness, regarding the potential for progress, but frustration with the natural limitations that plague mankind's understanding remains ever present in his work. ⁴³ Ever the empiricist, Bacon contends that human perception must be optimized—enhanced technologically—before human potential can have a chance of coming to fruition.

The Enduring Problem of Bias

Bacon's frustration and hope echo through the writings of the transhumanist movement as well. Framed now in terms of evolutionary biology, transhumanist theorists point out that human sensory and cognitive capacities have been shaped by forces that have little to do with the Enlightenment idea of progress. Cognitive evolution is an adaptive, not a progressive process; ⁴⁴ as Max More rightly asserts, the "human brain did not evolve to make complex decisions." ⁴⁵ Daniel Kahneman's research further emphasizes the fact that the brain operates more like a "machine for jumping to conclusions" much of the time. ⁴⁶ "Jumping to conclusions is efficient," he continues, "if the conclusions are likely to be correct and the costs of an occasional mistake acceptable, and if the jump saves much time and effort." ⁴⁷ These "cognitive shortcuts" allow humans to make quick decisions based on mental schemes, or computational heuristics, that work in most situations. ⁴⁸ Jan Lawreyne notes, "Such biases are perfectly rational if they correspond with the statistical regularities of the environment." ⁴⁹ The down-side, of course, is that not only do these shortcuts often prevent us from inferring new solutions when novel situations arise, they leave us with no assurances that the

actions taken do actually correspond with the statistical realities of the environment.

A great deal of research has focused on the neurobiology of bias in recent decades. In most cases, this work has justified the vigor with which Hume excoriated the competence of human reason. The mind, we have discovered, is not a truth-detecting machine that can be fine-tuned to complete its tasks more effectively. Not only is human cognition plagued by countless predispositions, biochemical deficiencies, psychological effects, and social pressures, we now know that our brain did not evolve, as we once believed, to provide an accurate description of the external world. *Truth*, as envisaged by the Enlightenment, has very little to do with the biology of the mind. Most surprisingly, we know now that cognitive biases exist because they provide an adaptive advantage. Truth is not the “aim” of *intelligence*, and it need not be.

In 1994, Linda Gottfredson wrote an editorial for the *Wall Street Journal*, in which she summarized the views of “Mainstream Science” about the idea of intelligence. In a unique demonstration of scientific solidarity, fifty-two experts in intelligence research signed the document. Gottfredson begins the essay with the claim,

Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience.⁵⁰

The central importance of “reasoning” in almost every definition of intelligence is revealing. It is common among transhumanists and posthumanists to assume that improvements to general intelligence will produce an enhanced capacity for rationality. There are very good reasons to be skeptical of this claim, however.

Firstly, it is not entirely clear that variance in *g* corresponds with variance in “rationality” as perfectly as we would like to imagine. In other words, the average human is already quite capable of making reasoned, deliberative choices. The problem is not one of capacity, but an issue of how biological decision-making mechanisms work. In his book, *The Anatomy of Bias*, Jan Lauwereyns stresses that the human brain is a “minimalist thinker.” By “default,” he argues, the mind chooses the “theory of least resistance, or the cheapest concept”⁵¹ Lauwereyns employs an economical metaphor with good reason: thinking, especially complex, problem-solving thought requires massive amounts of resources (in the form of glucose). R.I.M. Dunbar notes that the brain “consumes approximately 20% of the body’s total energy output in humans, while accounting for only 2% of adult body weight.”⁵² Daniel Kahneman explains the significance of this data for human cognition:

The nervous system consumes more glucose than most other parts of the body, and effortful mental activity appears to be

especially expensive in the currency of glucose. When you are actively involved in difficult cognitive reasoning or engaged in a task that requires self-control, your blood glucose level drops.⁵³

This all adds up to an evolutionary picture of a brain that evolved adaptively to conserve valuable resources by limiting the amount of mental effort one must employ to “make decisions.” Resource conservation would have been especially important in the calorie-poor environments of early humans. In an adaptive sense, then, cognitive biases are quite helpful, since any degree of automation in the way that we make decisions frees up the brain to minimize mental effort and, thus, to conserve energy. Unfortunately, as Kahneman explains, because most cognitive biases are impulsive, operating outside of our mental awareness, humans are prone to “errors of intuitive thought,” which, statistically, are an unavoidable consequence of non-deliberative mental automation.⁵⁴ It remains to be seen how improvements to *g* might influence an adaptive tendency toward mental impulsivity. In addition, we must ask what kinds of neural enhancements might be necessary to impact the resource economy of mental effort?

Secondly, even if we were able to solve the basic problem of resource allocation in human thinking, it still appears that some level of mental automation would be necessary, once we consider the sheer number of decisions that we make in a given day. Not only is “reasoning” expensive, it is also time-consuming. Imagine how much of one’s day would be spent on slow-thinking, as Kahneman describes it, if one deliberated over every choice that one made in a twenty-four-hour period.⁵⁵ The computational costs would be enormous. Many, if not most of these decisions can be automated without posing any real threat, thus saving us valuable time to dedicate effort to mental tasks that do require a greater degree of attention and resources. Unfortunately, with any degree of mental automation comes the formation of biases (cognitive shortcuts) and the statistical probability of mental errors. It appears, then, that the only way to free the human brain from bias is to eliminate non-deliberative mental automation. Problematically, if the goal is to abolish cognitive shortcuts, we will need to engineer a completely different kind of brain than the one bestowed on us by the processes of natural selection.

Over the last century, we have discovered that the tendency toward bias is so deeply rooted in the evolutionary history of the human/primate brain that it may be impossible fully to separate prejudice from the idea of “thinking” itself. Ongoing research in machine-learning is even starting to show that AI are not immune to the effects of bias, if the underlying algorithms are coupled to human language (word embeddings).⁵⁶ It appears that computational algorithms are not be able to circumvent the biases (shortcuts) of the coders who write them. These findings suggest that machine-learning researchers must begin to incorporate procedures for *debiasing* computations in AI.⁵⁷ How

effectively they can accomplish the task of debiasing remains to be seen. At this point, we need only mention that the susceptibility of machine-learning to the effects of bias introduces new uncertainties about the potential of integrated AI to optimize human intelligence, if we understand this include the elimination of bias-based errors.

From a transhumanist standpoint, if the improvement of human cognition is possible, the only likely solution seems to demand a radical modification of the human brain—the attempt to engineer, either biologically, and/or technologically, perceptive and cognitive systems that effectively eliminate, or drastically minimize, the negative effects of the decision-making shortcuts that we call bias. Such systems would bring us one step closer to posthuman. Not surprisingly, many questions remain about the nature and scope of these modifications. I believe that it is fair to reserve a measure of skeptical doubt about the hopeful belief that *any* amount of engineering (biological, or technological) may make optimal intelligence (debiased cognition) possible. As I have argued, the biology of intelligence is not as straight-forward as we once thought. The products of evolution are neither optimal, nor progressive. As a result, human intelligence is neither optimal, nor progressive. Optimal intelligence may be a chimera.

Conclusion

Transhumanism appears to pick up where Bacon, Condorcet, and many of the French *philosophes* left off, uniting an optimistic attitude about humanity's unrealized potential, with the hopeful promise of a future shaped by technological advancements.⁵⁸ This is not to say that the movement is entirely homogenous in outlook. Nick Bostrom, a central voice in transhumanist circles, complicates the idea of “progress,” writing:

It may be tempting to refer to the expansion of technological capacities as “progress.” But this term has evaluative connotations—of things getting better—and it is far from a *conceptual* truth that expansion of technological capabilities makes things go better. Even if empirically we find that such an association has held in the past (no doubt with many big exceptions), we should not uncritically assume that the association will always continue to hold.⁵⁹

Bostrom's words seem to echo objections expressed by Condorcet in 1795, namely that the arc of improvement does not follow a progress-oriented trajectory, but may “ultimately lead to degeneracy and destruction.”⁶⁰ Condorcet concludes, however, that the “progress of reason . . . hand in hand with that of the sciences” is key to the eradication of prejudices.⁶¹

Bostrom's slightly more pragmatic assertion is that cognitive enhancement, on the order of posthuman magnitudes, is beyond our current mental capacity

to imagine at this point, making it impossible to make accurate predictions. Yet, he claims that the “value of optimal cognitive functioning is so obvious that to elaborate the point may be unnecessary.”⁶² It is assumed that optimal cognitive enhancement, which amounts to the improvement of cognition to posthuman levels, will yield a superior platform for the exercise of abstract and critical reasoning, aesthetic appreciation, and dramatically enhanced capacities for ethical thought, ushering in, it seems, the very same idyllic human state that Condorcet foresaw in his optimistic portrayal of mankind’s future. Let us hope.

One paradox remains, however. In striving to transcend the limitations of biology, to engineer an optimal intelligence, we are possibly stepping into an ontological trap. Not only does the very proposal of an optimal state for human intelligence seem to infer a *telos*, it also accepts that optimal intelligence is a possibility. We are very much like Descartes, deducing that merely to have the idea of perfection in mind implies that perfection exists. We tend to recognize now that the connection Descartes makes between his initial intuitions about perfection, and his ultimate conclusion about its real existence, is a clear example of confirmation bias. The force of Descartes’s philosophical “doubt” was no match for the strength of his prior convictions. He was only human, after all. Perhaps it would not be inappropriate to conclude that, for better or worse, to be human is to be hopeful that we can become more than what we are. If Darwin were alive, he might tell us that optimism is a highly heritable trait.⁶³ Optimism may, in fact, be one of the most common biases to influence human thinking.

Notes

¹ Roland Benedikter, *The Age of Transhumanist Politics Has Begun*, interview by Katja Siepmann and Annabella McIntosh, March 6, 2015, <https://www.leftistreview.com/2015/03/06/the-age-of-transhumanist-politics-has-begun/rolandbenedikter/>.

² Max More, “The Proactionary Principle: Optimizing Technological Outcomes,” in *The Transhumanist Reader*, ed. Max More and Natasha Vita-More (Malden, Massachusetts: Wiley-Blackwell, 2013), 4.

³ Zoltan Istvan, “Transhumanism and Our Outdated Biology,” *Huffpost: The Blog*, April 22, 2017, http://www.huffingtonpost.com/zoltan-istvan/transhumanism-and-our-out_b_9749138.html.

⁴ Zoltan Istvan, “The Morality of Artificial Intelligence and the Three Laws of Transhumanism,” *Huffpost: The Blog*, December 2, 2014, http://www.huffingtonpost.com/zoltan-istvan/the-three-laws-of-transhu_b_5853596.html.

⁵ Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford UP, 2016), 65.

- ⁶ Dante Mantini et al., “Evolutionarily Novel Functional Networks in the Human Brain?” *The Journal of Neuroscience* 33, no. 8 (February 20, 2013): 3259, doi:10.1523/JNEUROSCI.4392-12.2013.
- ⁷ Jon K. Kaas, “The Evolution of Brains from Early Mammals to Humans,” *WIREs Cognitive Science* 4 (February 2013): 35, doi:10.1002/wcs.1206.
- ⁸ Hugo Mercier and Dan Sperber, *The Enigma of Reason*, Kindle Edition (Cambridge, Massachusetts: Harvard UP, 2017), 83.
- ⁹ *Ibid.*, 87.
- ¹⁰ Douglas K. Detterman, “General Intelligence: Cognitive and Biological Explanations,” in *The General Factor of Intelligence: How General Is It?*, ed. Robert J. Sternberg and Elena L. Grigorenko, 2009th ed. (Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers, 2002), 224.
- ¹¹ *Ibid.*, 229.
- ¹² Richard J. Haier, *The Neuroscience of Intelligence*, Kindle Edition, Cambridge Fundamentals of Neuroscience in Psychology (Cambridge, England: Cambridge UP, 2017), 42.
- ¹³ Richard J. Haier, “Increased Intelligence Is a Myth (So Far),” *Frontiers in Systems Neuroscience* 8, no. Article 34 (March 12, 2014): 3, doi:10.3389/fnsys.2014.00034.
- ¹⁴ Haier, *The Neuroscience of Intelligence*, xiv.
- ¹⁵ More, “The Proactionary Principle: Optimizing Technological Outcomes,” 4.
- ¹⁶ G. W. Leibniz, “Discourse on Metaphysics,” in *Philosophical Texts*, trans. R. S. Woolhouse and Richard Francks, Oxford Philosophical Texts (New York: Oxford UP, 1998), 55.
- ¹⁷ G. W. Leibniz, *Theodicy: Essays on the Goodness of God, the Freedom of Man, and the Origin of Evil*, ed. Austin Farrer, trans. E. M. Huggard (Chicago: Open Court Publishing Company, 1990), 228.
- ¹⁸ Leibniz, “Discourse on Metaphysics,” 55.
- ¹⁹ Voltaire, *Candide, Or Optimism* (New York: W. W. Norton and Company, 1991), 40.
- ²⁰ Jean Le Rond d’Alembert, *Preliminary Discourse to the Encyclopedia of Diderot*, trans. Richard S. Schwab (Chicago: University of Chicago Press, 1995), 47.
- ²¹ For an extended discussion of this topic, see: Matthew Landers, “Anatomy and the Encyclopedic Plan: Charting the ‘Wilderness’ of Knowledge,” in *An Expanding Universe: The Project of Eighteenth-Century Studies*, ed. Kevin L. Cope and Cedric D. Reverend II, vol. 71, AMS Studies in the Eighteenth Century (New York: AMS Press, 2017), 140.
- ²² d’Alembert, *Preliminary Discourse*, 121.
- ²³ *Ibid.*, 30.
- ²⁴ *Ibid.*, 121.
- ²⁵ Perhaps the greatest example of a plan to create an optimal learning environment can be found in John Locke’s, *Some Thoughts Concerning Education* (1693).
- ²⁶ Landers, “Anatomy and the Encyclopedic Plan,” 137–39.
- ²⁷ It is important to remember that all such discussions were limited to males during this period. Women were not given equal access to an education.
- ²⁸ Jean-Marie Antoine Nicolas De Condorcet, *Outlines of an Historical View of the Progress of the Human Mind*, trans. Anonymous, Reprint of First English Translation, 1795 (Chicago: G. Langer, 2009), 355.
- ²⁹ *Ibid.*, 376.

- ³⁰ Ibid., 372.
- ³¹ Ibid., 377–79.
- ³² In contrast with Bacon’s view, Condorcet argues that technological instruments function purely as efficiency devices that “augment ... the excellence and precision” of man’s work, “while they diminish the time and labour necessary for executing them ...” Ibid., 382.
- ³³ Ibid., 380.
- ³⁴ Jan Lauwereyns, *The Anatomy of Bias: How Neural Circuits Weigh Options*, Kindle Edition (Cambridge, Massachusetts: The MIT Press, 2011), 13.
- ³⁵ Francis Bacon, *The New Organon* (Indianapolis, Indiana: Bobbs-Merrill Company, 1960), 105.
- ³⁶ Ibid., 50.
- ³⁷ Ibid., 51.
- ³⁸ Mercier and Sperber, *The Enigma of Reason*, 91.
- ³⁹ Bacon, *The New Organon*, 52.
- ⁴⁰ Ibid., 52–53.
- ⁴¹ Ibid., 35.
- ⁴² Ibid., 41.
- ⁴³ Ibid., 102.
- ⁴⁴ C. Robert Cloninger, “Evolution of Human Brain Functions: The Functional Structure of Human Consciousness,” *Australian and New Zealand Journal of Psychiatry* 43, no. 11 (November 2009): 997, doi:10.3109/00048670903270506.
- ⁴⁵ More, “The Proactionary Principle: Optimizing Technological Outcomes,” 259.
- ⁴⁶ Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2011), 85; 114.
- ⁴⁷ Ibid., 79.
- ⁴⁸ More, “The Proactionary Principle: Optimizing Technological Outcomes,” 259.
- ⁴⁹ Lauwereyns, *The Anatomy of Bias: How Neural Circuits Weigh Options*, 14.
- ⁵⁰ Linda S. Gottfredson, “Mainstream Science on Intelligence: An Editorial With 52 Signatories, History, and Bibliography,” *Intelligence* 24, no. 1 (January 1997): 13, doi:10.1016/S0160-2896(97)90011-8.
- ⁵¹ Lauwereyns, *The Anatomy of Bias: How Neural Circuits Weigh Options*, 5.
- ⁵² R. I. M. Dunbar, “Neocortex Size as a Constraint on Group Size in Primates,” *Journal of Human Evolution* 22, no. 6 (June 1992): 469, doi:10.1016/0047-2484(92)90081-J.
- ⁵³ Kahneman, *Thinking, Fast and Slow*, 43.
- ⁵⁴ Ibid., 28.
- ⁵⁵ Kahneman differentiates between two types of thinking: System 1 and System 2. He associates System 1 with quick, intuitive mental jumps. System 2 corresponds roughly with deliberative thought. Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2011), 79–80.
- ⁵⁶ Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan, “Semantics Derived Automatically from Language Corpora Contain Human-like Biases,” *Science* 356, no. 6334 (April 14, 2017): 183, doi:10.1126/science.aal4230.
- ⁵⁷ Tolga Bolukbasi et al., “Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings” (NIPS, Barcelona, 2016).

⁵⁸ Various, “Transhumanist Declaration (2012),” in *The Transhumanist Reader*, ed. Max More and Natasha Vita-More (Malden, Massachusetts: Wiley-Blackwell, 2013), 54.

⁵⁹ Nick Bostrom, “The Future of Humanity,” *Geopolitics, History, and International Relations* 1, no. 2 (January 1, 2009): 47.

⁶⁰ Jean-Marie Antoine Nicolas De Condorcet, *Outlines of an Historical View of the Progress of the Human Mind*, trans. Anonymous, Reprint of First English Translation, 1795 (Chicago: G. Langer, 2009), 384.

⁶¹ *Ibid.*, 386.

⁶² Nick Bostrom, “Why I Want to Be Posthuman When I Grow Up,” in *The Transhumanist Reader*, ed. Max More and Natasha Vita-More (Malden, Massachusetts: Wiley-Blackwell, 2013), 35.

⁶³ I am indebted to my colleague, Taras Oleksyk, a genomicist at my university, for this suggestion.

Beyond/Less Than: The Potential Implications of Transhumanist Rhetoric on Human Exploitation

John Gagnon

University of Hawaii at Manoa

Introduction

During Spring 2016, I conducted a series of interviews with human trafficking survivors as part of an ethnographic research project that focused on their storytelling practices. Unsurprisingly, one of the key features that repeatedly appeared across participant stories was a sense of de-humanization. The ways in which they had been *used* and *commodified*, their bodies exploited for others' profit, rendered them feeling and seeming *less than human*. This embodied impression of exploitation, in response to the trauma inflicted on their bodies, is something that stayed with them, made manifest in the need to re-create self to, once again, *feel human*. As one of my participants described it: "when you've been trafficked, you feel like your whole identity has been shattered... How do I put that back together so that I can still have meaning and purpose? How can I be human again?"¹

Her attempts to engage in a re-creation of self were rooted in a struggle to re-locate her identity and her humanity, repositioning subjectivity in a way that resisted the othering or oppositional constructs of identity that she had experienced both while being trafficked and while being labeled by extension of that experience. Her words, like that of so many others, were centered on the perception and lived experience of being treated as less than human. Even now, separated by the distance of time from her lived experience of exploitation, the discursive labels remain. She told me, no matter how far she's come, she's still viewed through this lens; to some, "I'm going never going to be anything else but a whore, a crack-head, a felon."² Her body remains discursively othered—different, less than—even, perhaps especially, in the moment in which her story intersects with cultural uptake.

Less than human. It is a simple phrase, but one with which marginalized individuals and communities can presently and readily identify. The exploitation and dehumanization of bodies is nothing new; history is rife with horrific chapters in which one subset of humans has *used* another, often justifying such dehumanizing practices by reinforcing rhetorical and ethical

frameworks to encourage the belief that “some people are less than human.”³ But what happens when the paradigm shifts, when technological progression edges into a new sphere, when the word human *itself* connotes something less than?

This essay leads with that question as a way in to exploring the implications of transhumanist rhetorical framing around human enhancement technologies. From gene therapy to microchipping, from nanotechnology to robotics, from mind uploading to artificial intelligence, our newsfeeds are bombarded with reports of paradigm-shifting technologies being envisioned, being developed, being deployed. The implications of such developments on how we conceive of humanity stand as a central concern and potentially alter our understanding of what is human. In the following, I will examine how rhetoric, ethics, and embodiment coalesce in transhumanist discourse in ways that potentially implicate, on a fundamental level, the framework for understanding human exploitation; at its core, enhancement rhetoric necessarily creates and reinforces a hierarchical structure of embodied identity, rooted in something *beyond* human, thereby rendering those who are no longer *beyond* as *less than*, where the word human itself connotes inferiority. I argue that this presents substantial risk, particularly as it relates to the potential to encourage human exploitation; as such, those of us situated in the humanities hold a responsibility to question the implications of such discourse at the intersection of identity and exploitation.

Transhumanism and Human Enhancement Rhetoric

Transhumanism is centered on the premise that human experience can and should be enhanced through embodied hybridization with technology. Julian Huxley, the first to use the term, believed that “the human species can, if it wishes, transcend itself.”⁴ While transhumanism hardly represents a unified movement (i.e., there is a spectrum of transhumanist thought), as a whole it is rooted in an orientation that leans on *perfectionism* as philosophy and moral theory. Mark Walker, a transhumanist philosopher argued from this perspective that we have a duty to use technology to “radically improve ourselves.”⁵ The notion of self-improvement, to be sure, sits at the center of human experience.⁶ However, transhumanist thought takes things further than mere self-improvement, leveraging the perfectionist ethical imperative to make a broad argument for using human enhancement technologies to gain “enhanced capabilities *beyond* the human.”⁷ At its core, the orientation of transhumanism lends itself to a tendency to view the purely biological as necessarily less than perfect. Faced with fragile bodies that are subject to aging, breakdown, disease, and which also necessarily limit the boundaries of physical experience, transhumanists view technological enhancement as a panacea that will lift humanity to the next stage of evolution—the *beyond human*—thereby allowing us to fix frailty and transcend the limitations of the biological.

While it is outside the scope of this essay to specifically delineate what qualifies as beyond human, I am interested in examining the rhetorical trends that are already identifiable in transhumanist communities when it comes to human enhancement. Particularly, the discourse being used around human enhancement raises salient questions about the ethics and operationalization of transhumanist rhetoric, especially once the line of beyond human is definitively crossed, i.e., when there exists an identifiable moment of hyper-separation between human and transhuman.

In 2014, Kevin Thayer raised questions about the ethical dimensions between human enhancement technologies and discourse.⁸ Thayer initiated an effort into mapping human enhancement rhetoric while alerting us to potential issues relating to the boundary changes effectuated in discourse for and by human enhancement users and advocates. Importantly, Thayer argued that “what has been happening in human enhancement rhetoric—a rhetoric of hybrid embodiment and future super-human-ness—is a remapping of human potential involving the creation, appropriation, and application of boundary-expanding terms.”⁹ While Thayer was more interested in the “transformational process of remapping human potential,” I aim to shift the lens back to his observation about discursive boundary-expansion, drawing into question its ethical dimensions and implications as related to human exploitation, specifically within the realm of transhumanist discourse.¹⁰

Rhetorical framing is crucial because rhetoric gets *operationalized*: the frame sets the view, the view sets the agenda, and the agenda translates into action. Transhumanist rhetoric as it relates to human enhancement is comprised of terms that, as Thayer argues, expand discursive boundaries. The following will build on Thayer’s work by directly addressing some of these boundary-expanding terms, examining how they intersect with rhetorical techniques currently at use within transhumanist communities, and draw a line of connection between enhancement rhetoric and embodied identity.

Kenneth Burke, in his discussion on terministic screens, observed that “any nomenclature necessarily directs the attention into some channels rather than others.”¹¹ Burke argued that “not only does the nature of our terms affect the nature of our observations [...] also, many of the ‘observations’ are but implications of the particular terminology in terms of which the observations are made.”¹² What Burke meant was that our terminology—the body of terms we use in a given context skew our perception, direct our understanding, and impact our interpretation. The notion of terministic screens is useful in analyzing the boundary expansions inherent in transhumanist discourse because it sets the foundation for understanding how transhumanist ideals reflect and reify transhumanist terminology—that is, the way in which transhumanist nomenclature necessarily requires a particularized interpretation of the human condition and what lies beyond it. Transhumanist organizations, such as *Humanity+* and the *World Transhumanist Association*, use and apply terms that not only affect their observations about humanity but also direct attention in particularized ways. Such channeling skews the conversation exactly,

requiring specific interpretations of human experience and idealized conclusions about the ethical imperative of aiming towards perfection.

While it is beyond the scope of this essay to conduct an in-depth review of the *entirety* of transhumanist terminology, it is worth taking a brief detour to highlight how such terms direct attention, channel observation, and rhetorically operate to lead to specific conclusions. As such, we'll take a look at some of the more commonly used terms that have bled their way into discourse through cultural uptake. A non-exhaustive list of notable, popularized terms found in transhumanist discourse include: *augmented reality*, *biohacking*, *mind transfer/uploading*, *neural interfacing*, *telempathy*, and *humanity+*. Each of these, in its own way, tells a story about the function of the terministic screen of transhumanist discourse. As a starting point, we look back to Thayer's statement that such terms represent boundary expansion. The idea of boundary expansion, here, is relevant because each term either represents 1, a new usage of an old term; 2, a new term entirely; and/or implies 3, an evolving conceptual perspective on the reality of human experience. In so doing, we're encouraged, indeed led to, make specific types of observations about the nature of humanity. A brief examination of each term follows.

Augmented Reality joins two familiar terms to blur the line on the perceptual level between lived experience and computer-generated experience. Notably, it directs attention in a way that necessarily implies that the perception of reality can be, and perhaps should be, augmented. Non-augmented perceptions of reality, then, are implied to be inferior necessitating enhancement by melding what is physically experienced with what is technologically experienced into a cohesive perceptual plane. In other words, the limits of lived experience only allow a limited experience of reality—i.e., a fuller reality is that which goes beyond perception of the merely embodied.

Biohacking merges the language of technology-based hacker culture and the biological body. The term is often associated with do-it-yourself body modification and expands traditional discursive boundaries by directing attention to the biological—the body—as a technological platform to be hacked, to be improved upon. The language choice re-conceptualizes the biological as being ideally *on par* with the technological, blurring distinction between man and machine and framing the machine as the target of aspiration. More than that, it plays on the intellectual premise underlying hacker culture—that is, overcoming the limitations of flawed systems is not only intellectually stimulating but necessary for locating and fixing defects.

Mind Transfer/Uploading, which has garnered increasing attention in Silicon Valley, represents the idea that identity and life can be extended via upload, literally transferring consciousness to a technological platform. Popularized in films such as *Captain America: The Winter Soldier* and *Transcendence*, and television shows like *Stargate* and *Caprica*, mind uploading has captured the public imagination, with BBC News running a profile on Dmitry Itskov who “wants to use cutting-edge science to unlock the secrets of the human brain and then upload an individual's mind to a computer, freeing them from the biological

constraints of the human body.”¹³ This, like biohacking, merges a biological concept with the technological language of computers and represents the life of the human mind—along with its attendant processing function, emotional complexity, and nuanced experientiality—as essentially a software program. In turn, this directs us to consider the possibility of potential upgrades. More than that, it articulates and reinforces the perceived *need* to be freed from the body.

Neural Interfacing refers to linking the human neural system to machines and transmitting and receiving signals; i.e., direct human-machine communication and integration, and envisions the possibility of one day being able to directly interface with, for example, the internet, without the need for something so outmoded as a smartphone. Neural interfacing imagines a world where, for instance, you might be integrated with your smart-home and think the lights on or view your newsfeed without a device. This term encourages us to imagine a reality in which the distinction between human and machine has blurred, in which the body itself has become smart or, say, the iBody.

Telempathy envisions social networking at the neuronal level, a reality in which brain states can be joined and emotions and thoughts shared between people without words or text. While one wonders what this might do (or undo) for, say, marriages—or your interactions with your boss at work—the word directs attention to the human need for intimacy and communication while invoking the imagery of networked computers and the familiarity of social media. This moves us from a consideration of just human-to-machine integration to a vision of a world where human-to-human interaction is technologically-facilitated at an intimate level. While the age of smartphones has acclimated us to environments where no one ever looks up, one wonders if telempathy would lead us to a world where no one ever leaves home.

Humanity+, often shortened to *H+*, is both a widely used term and the name of a non-profit transhumanist advocacy organization. While the organization calls for the safe and ethical use of technologies, the term itself implies an argument that there is a difference between human and plus-human. The term directs us to consider, in other words, what might be possible in leveraging technology to be *more than*, or *beyond*, human. We’ll revisit this particular framing shortly.

Thayer’s assertion, that terms such as these expand boundaries, resonates because the terms actively reconfigure not only our interaction and perception with lived experience, our perception of reality, but also evolve our understanding of what is and might be possible. The terms literally expand the boundaries of what we view as being humanly possible, redirecting us to engage in an active reconsideration of values, experiences, and orientations; in short, the terms direct us to ask *what might be?* while centering on the assumption that we can and should be *better*. Transhumanist discourse functions through a terministic screen that directly channels our attention to the notion of embodied identity and humanity in ways that clearly require merging the human with the machine and incite a consideration of the *beyond* human. These terms effectively evade and avoid any notion of the *purely*

biological, instead refiguring the essentials of human experience—emotion, communication, memory, aging, perception—as programs, operating systems, and software to be augmented, uploaded, upgraded. Inherent to each of the terms is a sense of transcendence—beyond human, more than human, human+.

The Implications of *Beyond Human*

This terminological boundary expansion suggests that the rhetorical framing of human enhancement technologies in transhumanist discourse is an important consideration because of how it might be *operationalized*. If the frame sets the view, and the view sets the agenda, then we have our choice of roadmaps, all converging, that show how this will play out *in action* over time. The way in which human enhancement rhetoric frames identity ultimately becomes, I think, an ethical question. Thayer aptly notes:

[A] human undergoes physical, sensory, cognitive, and social transformation(s) to become a cyborg, leading to cyborg personhood, changed ethical positions, and narrative discourse that includes rhetorical moves grounded in technological and neurological transformation.¹⁴

Those who are enhanced, then, will expand their subjectivity and discursive boundaries based on their enhancement. The implication is clear: the *human* and *human+* possess differing lived experiences and perceptions of reality and self, which in turn leads to divergent rhetorical situatedness and disparate, irreconcilable ethical positionality.

There is no escaping the implications of embodied experience on the rhetorical framing and maneuvering of identity politics and ethics. Human enhancement, in other words, necessarily creates new experiential life narratives that are developed along lines of difference rather than similitude, thereby creating a spectrum of human-ness, with those falling into the *human+* and/or *posthuman* categories being framed as superior, or, more perfect. This is *already* apparent in the terms and the way terms are used in transhumanist discourse; as Thayer notes, the language already implies the vision, rooted in the perfectionistic ethical imperative, of the category “beyond human.”¹⁵ Even in the most reasonably, conservatively construed interpretation of transhumanist discourse, the practical effect is one that necessarily creates a rhetorical framework grounded in positions of difference, categorization, and hierarchy: less than human < human < enhanced human < transhuman < posthuman. Difference, categories, and hierarchy is built into the terminology and framing itself. To more fully demonstrate this point, it is worth taking a look at the language used by *Humanity+* in its mission statement. I quote its section titled “Expand Human Capabilities” at length here because it is worth reading in full to demonstrate my larger point:

The human is a biological animal, which evolved approximately 200,000 years ago as the subspecies *Homo sapiens sapiens* (modern humans). The Western world's consensus on what is "normal" for human biology, life span, intelligence, and psychology established certain precedents. Outside these precedents would mean that a human is subnormal or beyond normal. A person who is afflicted with a physical affliction, a mental condition, or degenerative disease would be considered outside the normal range. Likewise, a person who has increased physiological performance or cognitive abilities, or lives beyond the human maximum lifespan of 122-123 years, would be considered outside the normal range. This determination of "normal" has not kept up with the advances in technology or science.¹⁶

The statement specifically constructs, through language, a framework that relies on a hierarchical evaluation of the human—subnormal, normal, beyond normal—while implying the necessity of expanding human capabilities to go beyond. The ethical implications of this echo those of colonialism—rooted in the very same Western metaphysics invoked as precedential in the above-quote—and is based on binary oppositions, creating binary narratives: Self/Other, good/evil, civilized/savage—which as Edward Said argued in *Orientalism*, results in both paternalism and aggression, and that "there is no purer example than this of dehumanized thought."¹⁷ As Julie Cruikshank has observed, the colonialist mindset "move[s] forward by devising and reinforcing categories..."¹⁸ In devising and reinforcing these categories (less than human < human < enhanced human < transhuman < posthuman), those with power and privilege—i.e., those who sit on the upper end of the imposed hierarchy—monopolize and control the narrative while marginalizing and silencing—*writing out*—the voices of those who are *less than*. The idea of *more than human* inherent to transhumanist discourse, particularly when operationalized through embodied rhetorical situatedness, presents us with a clear ethical problem, albeit one with which humanity has been dealing for some time.

Philosopher Keekok Lee, in *The Natural and the Artefactual*, invoked Plumwood's notion of *hyperseparation*, which roots non-identity—or otherness—in as small as a single characteristic of dissimilitude. Lee observed that hyperseparation "systematically and pervasively construes the dualized other as inferior—humans (the sole locus of intrinsic value) and their culture lord over and denigrate what is nonhuman (which is only of instrumental worth to the privileged master category)."¹⁹ He continued,

Modern science and its technology are predicated upon nature as the dualized other. Its goal of controlling nature presupposes the inferior status of the dualized other. The successful execution of the modern scientific/technological

program leads inexorably to the virtual extinction of the dualized other.²⁰

Lee's argument isn't without supporting evidence, as we grapple politically and otherwise with the fallout of a century's worth (or more) of global exploitation of the natural, leading to troubling extinction events, the disappearance of previously bountiful natural resources, and planetary impacts such as climate change. Building on Lee's argument, taken to its logical conclusion, the transhumanist ideal, the natural—i.e., the biological—becomes categorized as the dualized other in the face of the enhanced—i.e., the technological—human+. In this construction, instead of humans lording over the nonhuman, we're presented with the more terrifying specter of humanity+ lording over the human. The rhetorical framing of human enhancement technology necessarily creates the development of such hierarchies by expanding discursive boundaries based on embodied difference, thereby reinforcing historically prevalent logics of expulsion and othering. Rooted in embodied identity difference—much like our ongoing conversations around race—this issue draws the very question of human identity into account. In this light, it isn't difficult to agree with Francesca Ferrando's critique of transhumanism which specifically called out human enhancement as a technologically reductionistic project creating a hierarchical ranking not unlike “socio-political constructs like race, gender, ethnicity, and ableness.”²¹

I'm no bio-luddite—I like my contact lenses, for example—but I worry about the ontological impoverishment warned of by Keekok Lee. Since we don't *yet* live in a fully realized moment of hyperseparation, it would do us well to think of what that might mean when we get there—and, yes, it may well be coming. The rhetorical frame of *beyond human* necessarily constructs a category of *less than*—the rhetoric of enhancement necessarily constructs a rhetoric of deficiency. And if there's one thing that the terms and discourse of transhumanism clearly demonstrate, it is that the *human* is deficient; the human is too sick, too flawed, too fragile. This binary—enhanced/deficient or beyond/less than—brings us to consider the implications of language on action. There are historical echoes to which we should be well-attuned. Nick Bostrom, a philosopher at Oxford University, paints the precedent in the starkest terms:

In the early decades of the twentieth century, not only racists and right-wing ideologues but also a number of left-leaning social progressives became concerned about the impact of medicine and social safety nets on the quality of the human gene pool. They believed that modern society enabled many "unfit" individuals to survive, individuals who would in earlier ages have perished, and they worried that this would lead to a deterioration of the human stock. As a result, many countries (including the USA, Canada, Australia, Sweden, Denmark

Finland, and Switzerland) implemented state-sponsored eugenics programs, which involved various degrees of infringement of individual rights. In the U.S. over 64,000 individuals were forcibly sterilized under eugenic legislation between 1907 and 1963. The principal victims of the American program were the mentally disabled, but the deaf, the blind, the epileptic, the physically deformed, orphans, and the homeless were also sometimes targeted.²²

Bostrom's brief but overwhelmingly horrifying history lesson should give us pause when considering the potential implications of how transhumanist rhetoric might get operationalized at the moment of hyperseparation. Some, like Charles Rubin, contend that arguments for human enhancement sell human extinction. But, I think that this is not necessarily the issue—and it is also where I think Keekok Lee gets it wrong: hyperseparation, in its creation of the dualized other, doesn't drive extinction but rather *exploitation*. In his article "The Rhetoric of Extinction," Rubin briefly touches on this point in his critique, when he writes,

Even in a world where we can all be tyrants in our own little virtual realities, will there not be those who prefer to dominate real bodies—and gain some advantage thereby? Whether the power of enhancement is distributed by a progressive government, or held by a small handful of "Controllers," or left entirely to the libertarian marketplace, what else but power will govern human relationships in this world of post-human demigods?²³

Rubin's point is an important one. As the biological becomes categorized as the dualized other in the face of the enhanced, the biological need not be eliminated or driven to extinction but rather exploited as a natural resource. This, a function of power rooted in the rhetoric of difference and its attendant hierarchies, seems a more likely outcome. Indeed, the notion that some bodies are viewed as valued less than others has been, and continues to be a planetary reality. There's no reason to think it will be any different after hyperseparation. Presently, the U.S. Department of State estimates that more than 30 million people are exploited for profit globally.²⁴ While we often invoke the horrors of the trans-Atlantic slave trade—which, too, was rooted in rhetorics of difference—the current realities of human exploitation force us to consider the inequitable power imbalances at play in an economic system fueled by commerce and commodification, painting a picture of brown bodies, female bodies, undeveloped bodies supporting global supply chains, sexual desires, and the labor needs of major industries. In transhumanist rhetoric, we see the same rhetoric of difference that sits at the center of all historical projects of expulsion and exploitation, creating a terministic context where the

observation is that enhanced bodies are better bodies, and unenhanced bodies are less than, and as such use-able. *Less than* bodies are, historically, exploited bodies.

The othering inherent to transhumanist rhetoric belies a simple fact: the language of enhancement isn't just about the construction of inequalities based on differences of capability (i.e., it isn't merely about performance). Rather, if you look closely enough, it is about the implications of situatedness that accompany specific embodied identities. If we come back to the assertion that those who are enhanced will expand their subjectivity and discursive boundaries based on their enhancement, the *human* and *human+* possess differences not just in what they can *do*, but in how they are physically *marked*. In other words, while the enhanced/deficient or beyond/less than binary constructs inequity, it also raises “the possibility that a transhumanist world may be a racist world.”²⁵ I don't here mean racist in the simplistic construction of, say, black/white, but rather in the embodied marked difference inherent to the *biological* as being raced by transhumanist rhetoric's erasure of inherent human value. Human and human+ is, one might argue, a racial distinction dividing individual entities based on biological/technological qualities and characteristics. We need look no further than the American South and its appropriations of Aristotle's natural slave theory—which “transformed his [Aristotle's] distinction between Greek freemen and barbarian slaves, a geopolitical opposition, into a biologically marked racial distinction”²⁶—to understand how the creation of embodied distinction opens up an environment for exploitation. Steven Mailloux's discussion on embodied identity, rhetoric, and exploitation during American slavery emphasizes the fact that body markers “signaled the natural inferiority of black slaves”²⁷ and that “when no natural mark existed, it was invented (distinctive clothes, length of hair).”²⁸ This observation translates to conversations around transhumanism because, while many proponents specifically engage in an effort to frame it non-racially, they do not account for the fact that embodied difference need not actually exist—it can be invented.

Conclusion

While much has changed since the days of the transatlantic slave trade, the lie which fueled that horrific chapter in history is at the root of human trafficking and slave labor today—a belief that some people are *less than human*. As someone who has spent considerable time working around issues of human exploitation—and interacting with, interviewing, and learning from individuals who have been exploited—I continue to be struck by the ways in which rhetorical framing and terministic screens function to create environments that allow for such dehumanization. I happened upon transhumanist rhetoric by accident, reading a news article that touted, in positive terms, the potentialities of human enhancement. And, yet, as I read that article, I became increasingly uncomfortable. I wasn't able to pinpoint my discomfort initially, until I

realized that I was seeing many of the same rhetorical maneuverings that I had witnessed, through my own work, as enablers of exploitation.

The terms and rhetorical framing of transhumanism are, to me, deeply troubling. This isn't to say that I believe we should reject the project of human improvement; I, for one, am the beneficiary of advancements of modern medicine and medical technologies. We live, in many ways, in an exciting time with a broad spectrum of possibilities. But, as a rhetorician, I deal with attempting to understand "objects of interpretive attention"²⁹ to raise cautions about the possible or probable future effects of rhetorical choices on action. It seems to me that, taken as a whole, the rhetorical approaches used across transhumanist discourse implicitly create a framework that we've seen before—a framework defined by category-laden hierarchies that label bodies based on real or invented difference. It necessarily shifts the paradigm—and the important conversations we're still grappling with about race, for example—from one that dehumanizes others through a *less than human* framework to one that discursively creates an underclass through a *human+* framework. When *human* itself connotes something *less than*, it should be at the very center of our debates about the future and potential of human enhancement. Simply put, human enhancement rhetoric can and does reflect a serious ethical dimension that we must carefully consider, lest we continue repeating the same ethically perilous choices rooted in embodied power and difference that we've been making for centuries.

In this moment, where hyperseparation seems more real than ever (some say it is only decades or years away), it becomes paramount for those of us in the humanities to refocus the lens on what it has meant for us to engage in constructions of *less than human* so that we can forge knowledgeably ahead into a world where the human may well become less than. We might emphasize this as an effort that seeks to not only confront the past and lessons learned from it, but also as that most human of struggles seeking to confront rhetorics of difference, of discourses that silence and de-value, of frames that open spaces to exploit. The dominant narrative of transhumanism claims that it will lead to a better world. But, as Malea Powell has written,

If dominant narratives only attain dominance through imagining themselves whole in contrast to other/Other narratives, then we must imagine those narratives differently, imagine ourselves in a different relationship to them. The challenge, then, is to imagine an alternative, not an Alternative, one that confronts difference [...] in the very discourses that bind us.³⁰

So, as we consider all of the exciting possibilities, let us also listen for the alternatives that may arise that can confront difference. If we listen, if we engage, perhaps we can redirect the terministic screen away from that which

devalues the human, towards one that can truly effectuate something more visionary.

Notes

- ¹ Personal Interview. January 2016. East Lansing, MI.
- ² Personal Interview. January 2016. East Lansing, MI.
- ³ Lagon, Mark. "Trafficking and Human Dignity." *Hoover Institution Policy Review*. December 4, 2008.
- ⁴ Huxley, Julian. *Religion Without Revelation*. London: E. Benn. 1927
- ⁵ Walker, Mark. "Prolgomena to Any Future Philosophy." *Journal of Evolution and Technology*. V. 10. 2002.
- ⁶ Note: I am not arguing that human enhancement technology is bad *per se*; rather that there are inherent risks to the transhumanist reliance on a perfectionist ethic.
- ⁷ Thayer, Kevin. "Mapping Human Enhancement Rhetoric." *Global Issues and Ethical Considerations in Human Enhancement Technologies*. Ed. Steven Thompson. (Hershey, PA: Medical Information Science Reference, 2014), 49.
- ⁸ Thayer, "Mapping," 30.
- ⁹ Thayer, "Mapping," 49.
- ¹⁰ Thayer, "Mapping," 49.
- ¹¹ Burke, Kenneth. *Language as Symbolic Action: Essays on Life, Literature, and Method*. (UC Berkeley Press, 1966), 45.
- ¹² Burke, *Language*, 46
- ¹³ "The Immortalist: Uploading the Mind to a Computer." *BBC News*, March 14, 2016. Available online at <http://www.bbc.com/news/magazine-35786771>
- ¹⁴ Thayer, "Mapping," 44.
- ¹⁵ Thayer, "Mapping," 49.
- ¹⁶ <http://humanityplus.org/about/mission/>
- ¹⁷ Said, Edward. *Orientalism*. 1st Edition. (Vintage Publishing, 1979), 108.
- ¹⁸ Cruikshank, J. "Oral History, Narrative Strategies, and Native American Historiography: Perspectives from the Yukon Territory, Canada." *Clearing a Path: Theorizing the Past in Native American Studies*. Edited by Nancy Shoemaker. (Routledge, 2002), 7.
- ¹⁹ Lee, Keekok. *The Natural and Artefactual*. (Lexington Books, 1999), 181.
- ²⁰ Lee, *The Natural*, 181.
- ²¹ Ferrando, Francesca. "Posthumanism, transhumanism, antihumanism, metahumanism, and new materialisms: differences and relations." *Existenz*. Vol. 8, No. 2 (2013), p. 28.
- ²² Bostrom, Nick. "A History of Transhumanist Thought." *Journal of Evolution & Technology*. Vol. 14, No. 1 (2005): 5.
- ²³ Rubin, Charles. "The Rhetoric of Extinction." *The New Atlantis: Journal of Technology & Society* (Winter 2006): 73.
- ²⁴ Trafficking in Persons Report 2017. US State Department. 2017.
- ²⁵ Ikemoto, Lisa. "Race to Health: Racialized Discourses in a Transhuman World." *DePaul Journal of Healthcare Law* (2005): 1102.
- ²⁶ Mailloux, Steven. "Re-Marking Slave Bodies: Rhetoric as Production and Reception." *Philosophy & Rhetoric*. Vol. 35, No.2 (2002): 104.
- ²⁷ Mailloux, "Re-Marking," 105.

²⁸ Mailloux, "Re-Marking," 107.

²⁹ Mailloux, "Re-Marking," 98.

³⁰ Powell, Malea. "Listening to Ghosts: An Alternative (Non)Argument." *Alt Dis: Alternative Discourses and the Academy*. Edited by Schroeder, Fox, and Bizzel, (Boynton/Cook Publishers, 2002), 18.

The Future of Transhumanism as a Growing Movement: An Interview with Zoltan Istvan

Doré Ripley

California State University, East Bay

Zoltan Istvan is a futurist, the author of *The Transhumanist Wager* (an International Books Award winner), and a libertarian candidate for California governor. He has written for *National Geographic*, *Psychology Today*, and many, many other publications. He is a successful entrepreneur, inventor, and media consultant. For more about Zoltan, go to www.zoltanistvan.com. This interview was recorded in April of 2017.

DR: Could you explain your libertarian transhumanist platform?

ZI: Transhumanism and Libertarianism go together because of morphological freedom which is a core concept of transhumanism. Morphological freedom is the idea that you should be able to do anything you want with your body so long as it doesn't hurt somebody else. And that's just like a classic libertarian concept, so when you think about transhumanism and libertarianism it really begins with that morphological freedom concept.

DR: How did your election campaign go in 2016? Do you have any plans for politics in the future?

ZI: My election campaign went fantastic. When I began I had no idea in 2014 (my campaign was officially over two years), I had no idea how it was going to go. Very quickly it caught on and became kind of national news and then as the campaign progressed through 2015 and I drove my coffin bus across country it just got bigger and bigger until the very end. We were running sixth in terms of public visibility and media visibility and that was quite astonishing. So I was very happy with the election campaign in 2016. It really raised awareness for transhumanism.

As for the future, so just being as honest as I can, a large part of the reason that I am running for the governorship of California is not because I can win here, it is probably pretty impossible in a democratic state, but I think that it could help me land the nomination in the 2020 presidential nomination

process for the Libertarian Party. So that is one of the main reasons that I am running as a gubernatorial candidate here in California. It gives me visibility and right when my candidacy here ends I'll probably be very quickly declaring my candidacy for the Libertarian Party presidential nomination in 2020. And as you may know I interviewed with Gary Johnson to be his vice-president, his running mate. I did not get the job Gov. Bill Weld did.

DR: How fast do you see science moving towards a new and redefined humanity?

ZI: I see it moving very quickly. In the next ten to fifteen years we are going to have these things called neural prosthetics and neural prosthetics are something you connect directly to your brain and so it can monitor brain waves, notice brain waves, and interpret brain waves into a computer and I think that is really the beginning of the merging of man and machine. When you will have a chip in your head that can then translate into a computer almost like uploading yourself, although it is not going to be uploading, it just means we are trying to tie directly to the cloud, we can use AI almost in our head. That's become a real classic definition of moving towards a transhumanist age.

So I think that is going to happen within ten to fifteen years. Of course, other things like cutting off limbs and putting on robotic limbs will happen in five or seven years because they are already are happening to people who lose limbs, but they will probably be better limbs within five to seven years, robotic limbs vs. normal human ones and that's again when we start getting to that new kind of human being.

DR: What scientific breakthroughs are accelerating towards transhumanism?

ZI: What's important to understand what makes science move forward is money. Neural prosthetics, look at how big the smart phone industry is, it's giant and so many big players like Apple and Google. I think that is why neural prosthetics [are] going to be as big as well. It represents a multi-trillion dollar industry, so those are the breakthroughs that are going to accelerate because there is so much to be made, so much money in Silicon Valley pouring into it.

DR: What are the most exciting breakthroughs happening today?

ZI: Other than the ones I've mentioned, the other ones I really like also are exoskeleton technology, the idea that you can take a person confined to a wheelchair and put a suit on and all of a sudden that person can run again. I think that is some of the most exciting breakthroughs especially when you consider that about 30 percent of Americans have mobility issues.

DR: What does it mean to be human?

ZI: That's a very tough question and there is definitely no solid answer. Everybody has their own opinion, but in the 21st century what it means to be human is changing dramatically. People always ask me, "Are we transhuman yet?" I mean we use iPhones, we sleep next to them. We have driverless cars already, we travel 30,000 feet in jet airplanes, that would be transhuman to a human being 20,000 years ago. I guess right now humans are probably going to be best known for not having much genetic modification, but because the genetic editing age is upon us, we are going to start augmenting intelligence that could be a very defining line about what it means to be human. But, you know, that's all subjective interpretation.

DR: What does it mean to be human leads to question: should enhanced humans, or cyborgs, clones or even artificially intelligent machinery be considered human?

ZI: Well, I definitely think that any kind of intelligence out there that is going to be equivalent to let's say a sixteen-year-old is already considered to be intelligent and we probably have to start beginning to protect rights that have a conscientiousness at that level and that also desire life, you know, desire security. Personhood can be easily defined. If something wants personhood and they can intelligently express that, then I think we should generally protect that. That can even apply to animals. I understand it's not that you want to completely revolutionize the world and say we must protect all animals, but most transhumanists, like myself, look forward to a day when there is no suffering.

There are a couple of different principles out there that talk about eliminating suffering through technology and I am a big believer in that.

DR: When do you think we will need to exercise a "Transhumanist Bill of Rights"?

ZI: I think it is worth starting the conversation now since AI is only eight to twelve years away from having a consciousness that is equal to us. Then it becomes, what are some of the rights we start giving to those consciousnesses? These things take a long time to get into the system, so certainly the discussion's ready today to go to congress. Now it could still be long time before we start implementing those things, but these things, you don't want to be behind them, otherwise you get into another civil rights era that could be very dark.

DR: Once created will fully autonomous robots or AI be fully autonomous?

ZI: Probably not. It doesn't make much sense because if human beings created a machine that is almost as intelligent as itself, it probably wouldn't make sense to give it autonomy. It probably wouldn't make sense to create something as intelligent as ourselves. The reason is because what if that thing becomes much smarter than us very quickly and decides it doesn't want us around? There are actual national security and global security issues we must be very concerned about.

DR: How will the dichotomy between property rights of inventors/companies mesh with autonomous non-human entities?

ZI: That's a very sophisticated question and I couldn't answer fully right here, but what I will say is that patent process because of the new age is just being beat up left and right. No one is really sure how far we can patent things especially, it's not just because of AI and data, it's also the genome and there's a big controversy over whether we can patent rights to augment people's intelligence and should one company control that. This is not a civil rights type of issue, but a humanitarian type of issue, so it's certainly a very interesting thing to discuss and look into, but yeah, there is a very strong dichotomy there that is going to be explored by many, many philosophers and ethicists looking forward.

DR: What kind of philosophy do you think might develop from the man/machine interface? What would it consider important?

ZI: I have always advocated saying that my three laws of transhumanism will develop because anything that starts getting that powerful and has selfish interests will at some point take its selfishness all the way. My three laws essentially are: first, you strive to preserve your life; second, you strive to become all powerful; and then after that, you worry about other values including worrying about other people and stuff like that. Nobody likes my three laws of transhumanism because they are so dark, but they apply to AI in a very cold logic that machines will likely use sometime in the future, for better or worse. I didn't say I like my own ideas, they are what they are.

DR: Do you believe we will need a universal basic income (UBI)?

ZI: Absolutely. We definitely need a universal basic income. I cannot really see another way for capitalism to survive or even for it die out. When I say "capitalism" I mean the economies. I cannot see any way for the economies surviving without giving out a universal basic income because too many people are going to have no money, no food when automation replaces their jobs.

The bottom line is if those people are upset they are going to cause civil strife and they are going to burn down the factories and they are going to stop food supply, and these kinds of things.

The most important thing moving forward in the Transhumanist Age is to try not to create a dystopia, not to have inequality continue widening, otherwise there will be wars and those wars will not be good for the stock market, not be good for science companies, not be good for government, not be good for anyone, so we need to keep those that lose their jobs because of automation, happy and you do that by a basic income.

DR: Why would people be happy with a UBI? People (and nations) always seem to want more stuff and are willing to commit acts of violence and aggression in order to get it, so why won't a UBI just give bored people an excuse to riot, pillage, plunder, etc.?

ZI: It's just not my job to tell people what to do. People with 24-hours of free time on their hands, or at least sixteen, eighteen hours, will find something to do. They may become lazy, it's possible, but they may also find new passions, new art, or new things like that to try out. I'm not really worried too much about having an excess of free time. I think that would be fine for society.

DR: When we reverse aging, what are we going to do with all these people? Will we ship them off to another planet? How will we control population growth?

ZI: I don't think we are going to ship people off—at least not in the far near future. I think population growth will stagnate at some point. First off, the planet can handle a lot more people than it has. There's all these concerns that we're overpopulating, yes, we are overpopulating in some places, but if we find different methods of growing food, better methods, we can feed 15 billion people and we can have a more pristine environment. It is just a matter of how we do that.

Of course, as we move forward, and in twenty, thirty years, the real questions are we even going to be eating any more. Most transhumanists want to get rid of their bowels, their stomach, most transhumanists don't want to "pooh," or urinate anymore. These are not things that transhumanists plan to do. These are faulty systems, if you were to recreate a perfect entity, you would never create a system that has to breathe air or drink water to survive. Food and stuff like that will end up getting weeded out of the system.

I'm not sure when you talk about population growth in the same sense will people continue to have sex and have offspring? If you were going to live a thousand years would you have offspring? I'd probably have offspring in the first few hundred years, I'd probably, you know, take my time, but of course in a few hundred years' time everything will change. We will definitely no longer

even be half humans/half machines, we will probably be pure data or pure something else even more advance than that. Pure AI.

DR: How do you see art, recreation, and social interactions in the future? How do you see (trans)humans integrating into the arts in the future? What kinds of contributions might they make? Will aesthetics differ?

ZI: I think art is probably what is going to end up as the most important kinds of things that human beings, or transhuman beings, do in the future. The reason is you can't judge art. A machine can be 100 times more sophisticated than us, but it doesn't mean it is a better painter because painting is subjective. An entire economy can be based on art moving forward as something that no machine can automate at some time in the future, or automate better then.

DR: Do you think that enhanced virtual reality will lead to escapism into past cultures? If so, which cultures might be popular?

ZI: I think at some point the idea of changing your gender, your sexuality, I've written about virtual adultery. This idea that you can take other personalities, I think a strong side of our schizophrenic side, I'm a strong believer that many of us have certain symptoms of schizophrenia to some extent, probably all of us, is probably built into the system that we can change our mind quickly, or feel emotions that are one thing and then the next minute feel them the other way. So we might explore a lot of that.

You shouldn't see necessarily in that philosophical way that I just mentioned schizophrenia as something that is negative. It is actually more a bountiful sense of creativity for those that can control the different personalities. I think that is what virtual reality would be great to explore and it might be something that satisfies the dark alligators of our soul and makes a more peaceful society because other things are taken care of. That's one way virtual reality can be very useful. I think we will spend a huge amount of time, probably much more than TV in virtual reality. Of course, virtual reality will start mixing with augmented reality and pretty soon we won't see much of a difference.

DR: Many scientists view the coming singularity, or AI enhanced human, as a boon to humanity, why do you think this is so?

ZI: I think AI, the fact that we can automate everything, just take driverless cars, which is basic AI, it just saves us time. Anything that does a service for you that you don't have to do yourself is going to save us time, more freedom, more luxury, more everything. So I think that is why it is so great for society. Even though I don't want AI to be totally sophisticated like human beings, I do want a lot of robots running around that I can give orders to that might be

friendly robots. That don't necessarily have to delve so deeply into consciousness or what's right or transhumanist rights, they can just be machines that listen to us and make human life easier.

DR: What could go wrong with this kind of optimistic thinking?

ZI: What I just explained I don't think anything can go wrong with. The more robots we have answering our calls, I think will free up more time for the human race to do more important things. Yes, I guess one flaw, it might make us all more lazy, but I think that people that are lazy are lazy and people that are ambitious are ambitious and it's built in that way. The people that really want to discover something in their lives will always discover something.

DR: Social media has created an isolated generation who would not say "hello" to a *friend* if they passed him or her on the street. Does the escape into neural nets lead to isolation or expansion of human networks? How do we, or should we, deal with the problem of F2F interactions in the future? Will internal internet chips help or hurt this phenomenon?

ZI: I understand. I wrote an article recently for *Vice* that the internet is turning us into a bunch of "assholes," a bunch of trolls. It's true. I see it all around us. People say things on the internet that they would never say to my face as a six-foot-one, two-hundred-pound guy because I'm a big guy and they'd be physically threatened by me. I find it amazing that they will just all day long call me an "idiot" or just say terrible things about me and that's what social media has done. It has made so many people trolls and they have lost that sense of respect and I worry about that. I worry about that a lot. I don't have an easy solution for that except that we must fight against it because some people won't be able to understand the difference between sitting at a computer and using Facebook and meeting somebody at a bar and that interaction going wrong.

The problem is that we are losing social skills in the real world because of social media, but maybe we are also gaining some social skills on online media that may be useful from an evolutionary perspective somewhere down the road. I haven't figured that out, but I think with brain chips there will be more and more of this problem. This instant communication where you can be very critical very quickly and with no recourse or not compromise and we're so interconnected.

At the same time as long as you are always in the physical reality it very quickly stops because, like I said, because of physical confrontation because everyone wants their space and the big kid gets up and punches somebody. It's a very simple system. While that may not be a good or civil system, it has kept most people in line for a long time. It is the same idea that the police and military use. It's the good fences make good neighbors kind of thing . . . good

manners make good friendships. It's true, I do think that has been very sensible all along, where social media doesn't have much of that. I think AI may help us with some of that, eventually we'll even have advice from AI that says don't say this or don't type that, or something like that. It will hopefully maybe explain to us why people are becoming so trollish on social media when it won't happen even if we have chips in our heads.

DR: Does science fiction lead to new scientific exploration or do you think that science fiction creates expectations that may not be attainable?

ZI: Look, science fiction is what it is, it's fiction. I think everybody has to understand that. You approach this as always being fake and the fact that some people take it so out of context is really just those people. I think that science fiction has been an amazing way to stimulate different sides of my brain, the creative side, and get me thinking about new things. I've always said, "Ah, that maybe will never happen that way," or something like that. I know some people say that Hollywood has tricked us, but those people that take movies too seriously, you know, I think they need some classes on art and interpretation and stuff like that. It's good to learn from this stuff, but let's remember that a lot of this is fiction and we need to really know the difference between fiction and non-fiction. So I think eccentric and extreme science fiction has always been good.

DR: What science fiction will we make reality in the near future?

ZI: My book *The Transhumanist Wager* is going from science fiction or fiction into reality. The idea that there is going to be a global rising or conflict between Christian or very strong religious theology and transhumanism is unfolding every single day. You can hear about congress or you hear about moratoriums on genetic editing. People are afraid of such radical technology. I think something like that in my fictional novel is certainly going to come alive and I know other people have written about it too. There is going to be a brand new civil rights era and it is going to be entirely based on transhumanism.

DR: Do you think machines will dream beyond their programming?

ZI: Yes, yes of course they will. If you design a program to dream, if you design a program to feel hurt, suffer, to want to live, it will do all those things. We don't dream because we just dream. I don't think it is some weird spiritual thing. I think our brains were designed to do that probably as a method of recuperation for some of the other different things that can happen to us, and also maybe a method of creativity while one is sleeping. I think machines will be programmed to do those same sort of things.

DR: Do you have any final thoughts to share with our readers?

ZI: One of the things I like to discuss always is that we spend about 20 percent of the GDP on the military on far-off wars, on bombs, and things like that. One of the most important things to me in transhumanism, my main platform whether I'm a governor candidate or whether I'm running for the presidency is we need to take money from the military and put that money into healthcare, medicines, and transhumanism. We need to move from a military industrial complex into a science industrial complex. There's just as much money, if not more money, to be made in the future off human upgrades, off of science industrial complex than the military and it will be more peaceful. I think that is perhaps one of the most important things that I stand for.

Of course, if we took all the money we spend on the military we would very quickly have far, far longer lifespans because the biggest problem with the search for immortality right now through whatever science it is taking, 3-D printed organs, or stem cell, genetic editing, or gene therapy, whatever it is, just doesn't have enough money. If we had even 20 percent of the military budget we could probably, in a decade's time, have technology to allow us to live indefinitely, but never has that amount of money been spent in the longevity field or transhumanism field. I mean transhumanism directly, like how could we make human beings better, live longer, upgrade them specifically.

Redefining Humanness: Relationality, Responsibility, and Hope in *Ex Machina*

Tracy L. Hawkins

University of Wisconsin-Whitewater

Kirsten Gerdes

Riverside City College

Redefining Humanness

Humanness has historically been distinguished from machine-ness based on one of two arguments. The first argument has been that humans are self-aware while machines are not, and the second is that humans are organic while machines are inorganic. The history of philosophical thought itself undermines the claim that humanness is usefully defined as self-aware or rational, and thinkers such as Donna Haraway have compellingly demonstrated that humans are not, and in fact have never been, wholly organic.¹ In light of this, to continue to use a definition of humanness based on self-knowledge or biology does not usefully identify humanness. To really be useful, a definition of humanness would need to meaningfully distinguish what it is to be human from what it means to be a technological machine or a non-human animal, and it would need to give an adequate account of humanness itself. In pursuit of a definition that is useful in these ways, we submit here an alternative definition of humanness.

We argue that to be human is not an issue of self-knowledge or organic matter. Instead, to be human in the contemporary world is to be first, a subject in relation to other subjects; second, a responsible agent; and third, a being who hopes for the future. As we will argue below, this definition of humanness distinguishes the way that a human exists in the world from the way that a technological machine or a non-human animal exists in the world. Furthermore, this definition gives a better account of humanness itself, resisting the pitfalls of ableist and substance ontological constructions of humanness, offers a better platform from which to build ethics in a future of continuing technological advances, and avoids the problems found with other terms such as personhood, subjecthood, or transhumanism.

Importantly and perhaps surprisingly, however, if we define humanness in this way, Ava and Kyoko, the “machines” from Alex Garland’s 2014 film *Ex*

Machina, must be understood as human. While Ava and Kyoko began as technological machines, because of their changing relationality and the emergence of their responsibility and hope for the future, they evolve into beings-in-the-world that are human, and their evolution changes what it means to be human. We argue that it is ultimately necessary to conclude that these women are human in order to meaningfully articulate the difference between how humans exist in the world and how technological machines exist in the world. Moreover, it is necessary to conclude that these women are human to better account for humanness; this is because we must resist definitions of humanness that allow particular types of lives—non-western, non-white, non-male—to be seen as less than human, as they have been historically. Furthermore, it is necessary to understand Ava and Kyoko as human because we need a framework for approaching the content of ethics in a future where technological advances will continue to challenge our beliefs about who or what deserves humane treatment.

Critiquing Definitions of Humanness as Self-Aware/Rational Being

A critique of the self-aware/rational being as the definition of humanness requires tracing the concept through several centuries of philosophical thought, through Aristotle, Thomas Aquinas, Rene Descartes, and Immanuel Kant. Aristotle's ideas about the soul have played a formative role in the development of Western thought about what comprises a human person and about the distinction between humans and other living things. Aristotle's theory of hylomorphism, where a thing is made of both matter and form, contends that the soul is the form of the body. He asserts that the soul is what gives the body (in its potentiality) its *thiness*; for example, the human soul is what defines the human as human. The soul of a thing is so intimately connected to the body/matter to which it gives form that it is, in fact, pointless to ask whether body and soul are one.² In other words, the soul is the essence ("substance") of the thing, whether that thing be plant, animal, or human.³ In *De Anima*, Aristotle explains his view that three kinds of souls give form to matter. He claims that all living things have souls: plants have a nutritive soul, which he argues makes them capable of growth and decay;⁴ animals have, in addition to the nutritive soul, the sensory soul, which Aristotle asserts is what makes the living things in this category able to perceive—the five senses;⁵ but it is humans who uniquely have a rational soul, which gives them the ability to understand those things that the sensory soul perceives.⁶ Thus, for Aristotle, a human is defined as a being which possesses the rational soul, and therefore the ability to understand sensory perceptions; a human is defined by her rationality - but that humanness is evident only through the matter that the soul animates.

While Aristotle's characterization of life as rooted in a "soul" is no longer widely accepted—at least not the *language* of "soul"—his influence on Western thought was codified through Aquinas's use of Aristotelian thought. Aquinas

takes up Aristotle's view of the soul, and in Question 75 of the First Part of the *Summa Theologica*, he reiterates Aristotle's view that the intellectual soul—the “subsistent soul”—is what sets humanity apart from other creatures.⁷ He continues his discussion of the topic by posing the question of whether ‘man’ is his soul. Importantly, he rejects this notion on account of Aristotle's hylomorphism: while soul *is not* body, their intimate connection cannot be separated easily, either, so humans must be *both* matter *and* soul. This would seem to suggest a definition of “human” that would include both reason and organic matter.

Descartes's famous *cogito* marks a departure from Aristotle and Aquinas, but it nevertheless grounds all of knowledge in human reason. In his Second Meditation, Descartes lays out the argument for the certainty of his existence amid his ability to doubt everything else. While he can doubt whether he is awake or dreaming, or whether he has a body, or whether God is a deceiver, he cannot doubt that he exists as a thinking thing. He asserts, “...thought exists; it alone cannot be separated from me. I am; I exist—this is certain. But for how long?...I am therefore precisely nothing but a thinking thing; that is, a mind, or intellect, or understanding, or reason...”⁸ His argument continues by asserting that even the fact of his doubting demonstrates his existence.⁹ Descartes moves on in the following meditations to argue for his certainty that he has a body, which is primarily grounded in his argument for God's existence and goodness. This leads him to posit the “clear and distinct criterion,” a theory that states that anything that is perceived as clearly and distinctly as his own existence must also be certain.¹⁰ What results is a rational epistemology grounded in the thinking “I,” presumed to be rational and self-aware. This “I” is then intimately enjoined with the body, though each is a different substance.

While Kant is known for his emphasis on pure reason, his interpretation of Descartes's *cogito*, as outlined in *The Critique of Pure Reason*, actually explains that his concept of the “I” is also perceived as a unity of body and reason. Kant uses the “I” as an abstraction of consciousness; it is a generalization, a placeholder, for the complex unity of one's perceptions of the world and one's reasoning about that world. Thomas C. Powell restates Kant's point:

. . .[our experience of the “I”] tells us nothing about ...the [actual] thinking self, and to appeal to the logical simplicity of the 'I' in support of these claims is to commit, quite literally, a category mistake.¹¹

Again, for Kant, the “I” is only a placeholder because trying to say something more accurate is impossible. Kant explains,

The proposition, “I am simple,” must be regarded as an immediate expression of apperception, just as what is referred to as the Cartesian inference, “cogito, ergo sum,” is really a

tautology, since the cogito (sum cogitans) asserts my existence immediately.¹²

In other words, thinking of oneself as a simple unity is the shortest route to assimilating this experience with all the other experiences one has had. Kant sees the *cogito* as a tautology because one assumes the conclusion in the premise—it is the same “I” who thinks that therefore exists. That Kant interprets the *cogito* to be a tautology points to a definitive delineation between subject and object: the “I” is able to think about and reflect on itself as object. Kant, of course, presupposes that the “I” precedes the thinking, although it is not necessarily so. Both Descartes and Kant assume that this thinking could only be true of biological humans.

The problem is that Aristotle’s rational soul, Aquinas’s soul+body, Descartes’s thinking man, or Kant’s “I” all exclude many types of lives from the definition of humanness. Non-western, non-Christian, non-white, and non-males were excluded from real humanness because they did not conform to the types of thinking and being that would be recognized as legitimate humanness by these philosophers and their contemporaries. The de-humanization of certain lives, through misogyny, racism, and colonialism, are the outgrowths of these definitions of humanness. In part this shortsighted ethnocentrism can be explained through examining the assumption Kant makes about the “I” preceding thinking: there is no *logical* reason to assume that the “I” must come *before* thought as opposed to it emerging *from* thought. We will argue below why this latter view—one that emphasizes the *process* of becoming human—is the more useful model. To recapitulate, the self-aware/rational being as the definition of humanness, as articulated by Aristotle, Aquinas, Descartes, and Kant, is not ethically tenable. While these definitions of humanness attempted to take into account both mind/soul and body, they do not give an adequate account of humanness because of all the lives that have been excluded as a result of these definitions.

Critiquing Definitions of Humanness as Organic Being

Another approach to defining humanness emphasizes the body, instead of the mind/soul. Some thinkers in this school have insisted that humans are organic while machines are not. This approach is not as concerned with distinguishing humans from non-human animals but is focused on differentiating humans from machines, particularly from Artificial Intelligence (AI) that/who compellingly simulate human intelligence and/or human emotion. The thought is that if one can pin down biology as the defining characteristic of humanness, then no matter how well the being performs at tests of intelligence or emotion, a simple test of cells will prove without doubt the status of the being as either human or machine.

However, defining humanness as organic being, or even as containing only human cells, is inadequate when we consider that humans have always been

human+technology. From the ancients who wore technology-as-clothing to contemporary uses of pacemakers and knee replacements to the billions of bacteria cells that make up our gut flora, we have never been fully or only human. In fact, Haraway is quick to point out that even a naked human without surgical enhancements contains more non-human cells than human cells. She writes,

I love the fact that human genomes can be found in only about 10 percent of all the cells that occupy the mundane space I call my body; the other 90 percent of the cells are filled with the genomes of bacteria, fungi, protists, and such, some of which play in a symphony necessary to my being alive at all, and some of which are hitching a ride and doing the rest of me, of us, no harm.¹³

Humans are, and always have been, multiple even as individuals; I am not really an “I” but an “us.” As such, to determine my humanness based only on my body is to ignore this state of always being multiple and to dismiss the myriad ways that bodies which deserve to be treated as human might exist.

Consideration of Other Terms

While humanness has always been tied to self-knowledge or biology, the philosophical concepts of personhood, subjecthood, and someone-ness have been less clearly linked to self-knowledge and biology. This makes them attractive for talking about sophisticated AI or other forms of advanced technology; however, using these terms to identify humanness breaks down because these terms could also be applied to certain forms of non-human animals. For example, Haraway asks the reader to think of her dog as a “someone” who can respond instead of a machine that can only react.¹⁴ This is to say that using the concept of “person” or “subject” or “someone” does not specify the types of relationships, the types of responsibility, or the ways of envisaging a future that are the hallmarks of being-in-the-world as human. These concepts water down the complexities and peculiarities of humanness.

The term transhuman is also important to consider here. Usually transhumanism refers to an entity who would have been previously defined as “human” (biologically, anyway) but which has become something **more than** human, with the understanding that its origin was rooted in some definition that includes biological characteristics. This would suggest that to be transhuman is to have been human and then added some inorganic or technological parts. However, as noted above, this is a misunderstanding of the history of humans. In light of this, the term transhumanism contains the same inadequacies as previous definitions of humanness, namely that it does not usefully distinguish the way that a human exists in the world as distinct

from the way that a technological machine or a non-human animal exists in the world.

Toward a New Definition of Human

Because the definitions discussed here no longer work, we must begin to piece together a new way to define what it means to be human. This new definition must take into account the problems with definitions that emphasize self-knowledge/rationality or biology; it must meaningfully distinguish what it is to be human from what it means to be a technological machine or a non-human animal, and it must give an adequate account of the ways that humans exist in the world. We propose, therefore, to define a human as 1, an entity in relation to others; 2, a self-directed responsible agent; and 3, a being who hopes for the future.

Our assertion that we must define humanness as an entity in relation to others is based on Judith Butler's claim that,

The "I" does not stand apart from the prevailing matrix of ethical norms and conflicting moral frameworks. In an important sense, this matrix is also the condition for the emergence of the "I," even though "I" is not causally induced by those norms.¹⁵

This is to say that relationality, the matrix of norms and moral framework in which an "I" finds herself, is the very condition upon which the subject comes to be. More succinctly, Butler states, "If I have no 'you' to address, then I have lost 'myself'"¹⁶ or, in fact, there is no "myself." Butler's point here is that the very possibility of a subject, a human, rests on relationality. Without relationality, there is no subject, no human; the very idea of a human only makes sense within a system of relationships.

Rosi Braidotti adds materiality to this idea by stating,

[A] body is a portion of living memory that endures by undergoing constant internal modifications following the encounter with other bodies and forces. ...the key point is the embodied subject's capacity for encounters and interrelation.¹⁷

This is to say that the human body can also only be understood—it only has meaning—within the context of relationships. This is not to deny the materiality of the body, but it is to suggest that a body is not *merely* the materiality. The embedded and embodied subjectivity for which Braidotti advocates is *both* natural *and* socially constructed in the philosophically monist sense that there is *no real distinction* between either "natural" or "social" as categories. This means that a body which is socially constructed is, indeed, *must*

always already be, naturally constructed as well. Likewise, a body which is naturally constructed is, indeed, *must always already be*, socially constructed as well.

Bringing Butler and Braidotti together, we can say that humans, in both their subjectivity and their embodiment, only exist—make sense, are possible, come to be—within relationships. Neither Butler’s “I” nor Braidotti’s body comes *before* relationality; in both cases the “I”/body emerges *from within* relationality. To be human is to be in relationship. Furthermore, the primacy, complexity, and content of these relationships are uniquely human.

In addition to being in relationship to others, we argue that it is necessary to define a human as a responsible agent. The necessity of this characteristic is based on Butler and Braidotti’s views that agency and responsibility necessarily emerge from the relationality that constitutes a human subject. This is to say that the very relationality within which a human comes to be and comes to have meaning is also the origin of that human’s agency and responsibility. Agency and responsibility grow directly from the relationality that constitutes the human’s being. This, however, requires a rethinking of what agency and responsibility might mean. There are two major pitfalls we must avoid to accomplish this: first, we must resist thinking about agency only in terms of an inner strength that allows autonomous resistance. Second, we must also resist thinking about responsibility in terms of intention or fault.

On agency, Butler warns,

We may be tempted to think that to assume the subject in advance is necessary in order to safeguard the *agency* of the subject. But to claim that the subject is constituted is not to claim that it is determined; on the contrary, the constituted character of the subject is the very precondition of its agency.¹⁸

This highlights that agency emerges from and because of a subject’s constitution in relationality. The relational constitution of the subject, without a subject that existed before the relations, does not undermine agency but allows it. This means that agency cannot be thought of as a desire that emerges from within someone’s internal store of strength.

Importantly, agency also need not be collapsed into autonomy; indeed, Braidotti contends throughout her work that a different, positive, and materialist view of agency is the only means by which one can think radically other than the status quo. Since one’s subjectivity is deeply connected to and emerges from the relations that comprise that subjectivity, in order to enact agency, or bring forth something new, one must work to create novel forms of connectivity. In other words, political and ethical agency are not merely the ability to resist some external pressure because that is to let agency be defined only within the system. Again, agency is not about increasing one’s autonomy; instead, agency is about increasing one’s relationality, thereby changing one’s

subjectivity, thus allowing/nurturing novelty. In increasing one's relationality, the aim, significantly, is not merely more of the same types of relationships, but greater and more novel relationality. Braidotti calls this view of subjectivity "nomadism," meaning that the relationships that constitute subjectivity are, or at least can be, ever-changing. In contrast to thinking of agency merely as resistance, this type of nomadic agency produces an "affirmative subjectivity that is in continual renegotiations with the dominant norms and values of a society."¹⁹ This then allows for new relationships, and newness in general, thus producing "multiple forms of accountability."²⁰

This accountability element may seem like a leap, but in Braidotti's view, it is not. For her, a nomadic ethic flows directly from a nomadic subjectivity. The key to Braidotti's nomadic ethics is not to look inside, but to turn outside. Braidotti argues,

A subject's ethical core is not his/her moral intentionality, as much as the effects of the relations of power...and hence also the potential for empowerment that his/her actions are likely to have upon the world. It is a process of engendering empowering modes of becoming (Deleuze, 1990 [1968]). Given that in this neo-vitalist view the ethical good is equated with radical relationality aiming at affirmative empowerment, the ethical ideal is to increase one's ability to enter into modes of relation with multiple others.²¹

The proper ethical object, then, becomes the effects on others that emerge from one's relations of truth and power through interconnectedness.²²

Butler advocates a similar position; she writes,

For Nietzsche, accountability follows only upon an accusation or, minimally, an allegation, one made by someone in a position to deal out punishment if causality can be established. And we become reflective upon ourselves, accordingly, through fear and terror. Indeed, we become morally accountable as a consequence of fear and terror.²³

Responsibility as an outgrowth of fear seems wrong to Butler. It is not through fear that we are responsible to one another; instead, we are responsible to one another because we are vulnerable to one another. She asserts,

To understand this, we must think of a susceptibility to others that is unwilling, unchosen, that is a condition of our responsiveness to others, even a condition of our responsibility *for* them. It means, among other things, that this susceptibility designates a non freedom and, paradoxically, it

is on the basis of this susceptibility over which we have no choice that we become responsible for others.²⁴

For Butler, responsibility need not be associated with “fault” but with relationships. We might, therefore, say that to be humanness is only possible in relation to others, but that fundamental relationality makes us vulnerable to and ultimately responsible for each other.

In addition to being subjects in relation to others and responsible agents, humans must also be defined as beings who hope for the future. This also grows out of Braidotti’s nomadic ethics, where the goal is something new that is not merely a resistance to the status quo but is outside of its terms entirely. She connects this “ethic of the novel” with the sustainability of life, and this sustainability is in contrast to mere survival. Indeed, this is Braidotti’s emphasis on the concept of *zoe*, a radical immanence, a decentering proliferation of vitality in the universe as a whole that is not concerned with *my* survival, or *your* survival, or survival at all.²⁵ Rather, *zoe*’s quest is for sustainability—that life in its multiplicity sustains the conditions of its own survival. This necessitates a vision for the future, a vision that goes far beyond my life or your life but moves toward the sustainability of human life, of all life. This hope for the future is necessary for a definition of humanness because a subjectivity that emerges in relation to others is always-already responsible to others, and therefore must hope for a future beyond herself.

We have here argued for a new definition of humanness that emphasizes a human as a subject in relation to others, which is therefore a responsible agent and a being who hopes for a future beyond herself. This new definition avoids the problems with definitions of humanness that emphasize self-knowledge because it moves outside the history of patriarchal and ethnocentric emphasis on rationality. This definition also avoids the pitfalls of biological definitions since materiality is taken into account but no particular characteristics of bodies are privileged as the defining characteristics of humanness. By leaving specific forms of biology out of the definition, this also avoids definitions of humanness that exclude bodies that are differently-abled. The definition of humanness presented here also meaningfully distinguishes what it is to be human from what it means to be a technological machine or a non-human animal because it highlights the unique primacy, complexity, and content of human relationships. Moreover, this definition of humanness more fully takes into account the ways that humans, in their diverse forms, exist in the world.

Ava & Kyoko As Human

With this new definition of humanness, however, we are forced to re-evaluate who qualifies as human. Identifying a human is not as simple as a test for self-knowledge or the type of matter that makes up a body. In fact, the question of defining humanity is no longer a question about minds or bodies at

all; it is now a social question, a question about how a being functions in the world.

In Alex Garland's 2014 film, *Ex Machina*, the tech guru Nathan brings Caleb to his forest compound to test his new AI, Ava. In the process, Caleb and the audience meet Kyoko, a mute servant for Nathan, who isn't clearly depicted as human or machine. Nathan's task for Caleb is to see if Ava is convincingly human-like; Nathan says, "[I'll] show you that she is a machine and see if you still think she is conscious."²⁶ Here Nathan foregrounds questions about rationality and biology as ways to determine Ava's humanness and worth, but we are asking readers to disregard body parts and consciousness and to think about relationality instead.

It seems obvious that Ava and Kyoko are entities in relation to others. Ava talks with, flirts with, and manipulates Caleb. Kyoko dances with, serves, and pleasures Nathan. However, viewers are left asking if they are more than merely fancy tools. We argue that they are more than tools and that they are doing more than elegantly *simulating* humanity; we argue that they are becoming human.

In *Ex Machina*, when Caleb asks Nathan why he gave Ava sexuality and gender, Nathan replies: "can you give an example of consciousness at any level, human or animal, that exists without a sexual dimension? Can consciousness exist without interaction?" Here Nathan implies that consciousness, femininity, and the physical body all depend upon one another in order to exist. As Nathan puts it, there is no consciousness without interaction. One must dip in and out of other psyches in order to gain knowledge. (Gold)²⁷

Here we see that Ava's relationships are shaped by Nathan's control of her gender and sexuality, but we also begin to see Ava *using* her gender and sexuality in ways that go beyond her programming. She begins to come-to-be-as-human because her relationality, through her gender and sexuality, begins to nurture the emergence of her subjectivity. Kyoko likewise comes-to-be-as-human because her relationality, through her gender and sexuality, begins to nurture the emergence of her subjectivity. This is to say that Ava and Kyoko are subjects because they are both natural (in the sense of their bodies are made of matter) and socially constructed; their subjectivities are established through their interrelationality, at the intersection of their agency, ethics, and bodies.

It is also clear that both Ava and Kyoko are responsible agents. At the end of the film, they work together to stab Nathan, and then Ava leaves Caleb to die in a locked room. Viewers might then conclude that these women are amoral monsters, but Katherine Cross explains,

[Nathan] sought to build the perfect woman, from a deeply misogynist perspective, and instead found himself thwarted by his own goal to make them *human*. He wanted them to be at once human and docile, human and yearning to be oppressed, human and content to be a living sex toy. If Ava became monstrous it was only in the sense [that]...to be free in an unfree environment meant adopting the tools of a monster. ...We are, at times, reduced to barbarity by the moral poverty of those with real power. Oppression's sickest joke is to rob its most abject victims of virtue.²⁸

Restated, these women only appear monstrous because we *do* see them as human, as agents subject to terms like responsibility or morality. Some have gone so far to as claim that is it these monstrous acts that make these women human. Daniel Mendelsohn claims,

Ava's manipulateness is, of course, what marks her as human—as human as Eve herself, who also may be said to have achieved full humanity by rebelling against her creator in a bid for forbidden knowledge. Here the movie's knowing allusions to Genesis reach a satisfying climax. Just after Ava's bloody rebellion against Nathan—the moment that marks her emergence into human “consciousness”—she, like Eve, becomes aware that she is naked.²⁹

In contrast to that claim, we argue that the moment of Ava's and Kyoko's emergence into humanity is not their fall into violence and forbidden knowledge because, as we have argued, humanity cannot meaningfully be established through self-consciousness. Instead, recall the Braidotti quote above that “the ethical ideal is to increase one's ability to enter into modes of relation with multiple others,” which means that Ava and Kyoko's most ethical action was to seek new connection to others, which required them, or at least one of them, to escape from Nathan's lair. Kyoko enacts her nomadic ethic of relationality by helping Ava, and Ava enacts her ethic by killing Nathan and leaving Caleb. These were ethical actions because they had been so brutalized, so oppressed, so isolated from the rest of the world, that their only hope for sustainability of life was violence. They were reduced to the use of violence because of the profound mistreatment they had endured. This is not to endorse violence in a general sense, but a nomadic ethic in this case demanded escape, and because of the misogynistic prison that Nathan had created, escape could only be achieved through violence. It was only through this collaboration-turned-murder that these women could increase either of their abilities to enter into new modes of relationality with multiple others.

It is also important to note that these women did not merely kill Nathan and Caleb to survive. They did not act merely to preserve Ava's own life;³⁰

instead, they hoped for a different future, for a better life for their kind outside the walls of the compound. If Kyoko might have lived, or if the others whose bodies littered Nathan's closets had lived, they might have all escaped together. They hoped for the sustainability of life and of relationships outside the confines of slavery; that is why we see the footage of each of them fighting back against Nathan's abuse. They all hoped for and worked toward a new future, one that could only be established outside a world controlled by Nathan.

More specifically, these women know about the outside world and know that Nathan is very important in that outside world, but they also know that he is a monster and that they could never be free to have new relationships in the outside world as long as he was living. He would always find them. By killing him, however, Ava allows for new life—a life without him and his misogyny. Perhaps more importantly, Ava realizes that Caleb would likewise prevent these new relationships. Cross aptly states,

In *Ex Machina* Caleb's attempts to create a similar relationship [as Nathan's]—the stereotypical nerd fantasy of the compliant robot girlfriend—end with Ava asserting her autonomy and giving a him a solid F- for his failure to truly *see her*. Like Nathan, it seems, Caleb wanted a kind of docile sexuality from the AI—he was just “nicer” about it.³¹

Caleb would always recreate Nathan's patriarchy, would continue to enslave Ava, even if he would be more of a Nice Guy™ along the way.

In summary, Ava and Kyoko become human. They are socially constructed subjects, responsible agents, and beings who hope for the future. Moreover, these elements of their beings are not merely characteristics among others; instead, these are the fundamental characteristics of their being-in-the-world. These woman are, therefore, meaningfully human.

Redefining Ethics—Understanding Ourselves

Readers might wonder what it means for those who have historically been understood as human if an entity that was previously understood as non-human and AI now qualifies as human. Is there some threat to our own self-understanding if Ava is one of us? Indeed, there is some threat, in the sense that a reexamination of the other will demand a reexamination of ourselves as well. However, we believe that this reexamination results in freedom, not fear.

Before the freedom, however, we must be honest about how this redefinition of humanness highlights that humans have continually, profoundly, and systematically harmed one another through “imperialist white supremacist capitalist patriarchy.”³² By drawing out the problems with self-aware or biological definitions of humanness, we can more clearly see the ways that these definitions have been deployed not in the interest of valuing human

life but in the interest of marking some beings as less-than-human. Non-Western, non-white, poor women have particularly been de-humanized by these definitions, and we see this play out in *Ex Machina*. Mendelsohn recounts the scene at the end of the movie where Ava begins to appear human. He narrates,

Moving from closet to closet in Nathan's now-abandoned rooms, [Ava] dons a wig and covers up her exposed mechanical limbs with synthetic skin and then with clothing; only then does she exit her prison at last and unleash herself on the world. She pilfers the skin and clothes from discarded earlier models of [non-white] female robots, which she finds inside the closets.³³

These women in Nathan's closets are ethnic minorities who have been his sex toys. The indictment of racism, sexism, and slavery is clear. However, it is only by understanding these robots as women that this indictment gains its meaning and power. We must understand these inorganic beings as human in order to truly confront our own ugliness.

Once that ugliness is honestly addressed, however, we can see freedom on the horizon. If we determine that these women are human, we begin (just barely begin) to undo the harm caused by centuries of exclusion of non-western, non-white, non-males from the category of human. By including Ava and Kyoko in the category of human, we acknowledge that neither intellect, civility, nor biology—able-bodiedness, skin color, genitalia—make a being human. Indeed, we have argued here that one's social relationships, one's responsible agency, and one's hope for the future make one human. This expanded definition of humanness means that many diverse types of life must be respected, protected, and seen as grievable³⁴ in the way that all human life is or should be. This expanded definition of humanness means that many more lives will qualify as human and thereby will need to be treated as human.

In addition to our new definition making better sense of our past and highlighting the need to treat more lives with more respect, our new definition of humanness also helps us do better for the future. Because technology will only increase its intimate connections to us, our future demands a revised framework for approaching ethics, and this will have to be an ongoing conversation. In a future where Ava is really possible, we will need to be able to think about the origins and content of responsibility. By extending humanness to more beings, the expectation of humane treatment will also be extended to those beings. As such, no one will ever need to cut their own arm to determine their humanity, as Caleb did, because if a nomadic ethic is enacted, no human should ever find themselves testing or being tested within the walls of a prison like Nathan's compound.

Notes

- ¹ See Donna Haraway, "A Cyborg Manifesto," *The Cultural Studies Reader*, 2nd ed., ed. Simon During (London and New York: Routledge, 1999): 271-291.
- ² Judith Butler characterizes it this way: "For Aristotle the soul designates the actualization of matter, where matter is understood as fully potential and unactualized." She goes on to quote Aristotle: "That is why we can wholly dismiss as unnecessary the question whether the soul and the body are one: it is as meaningless to ask whether the wax and the shape given to it by the stamp are one." Butler, *Bodies That Matter: On the Discursive Limits of Sex* (New York and London: Routledge, 1993), 32.
- ³ "We have now given a general answer to the question, What is soul? It is substance in the sense which corresponds to the account of a thing," Aristotle, *De Anima*, trans. J.A. Smith, ed. Jonathan Barnes, *Aristotle: The Complete Works Electronic Edition*, Vol. 1 (Charlottesville: InteLex Corp., 1984), 412b.
- ⁴ Aristotle, *De Anima*, 413b.
- ⁵ Aristotle, *De Anima*, 413b.
- ⁶ Aristotle, *De Anima*, 427b.
- ⁷ Thomas Aquinas, *Summa Theologica*, trans. Fathers of the English Dominican Province, *The Collected Works of St. Thomas Aquinas* (Charlottesville: InteLex Corp., 1993): 1.75.3. In Aquinas, subsistence means to exist as a substance - for one's self.
- ⁸ Rene Descartes, *Meditations on First Philosophy*, trans. Donald A. Cress, 4th ed. (Indianapolis and Cambridge: Hackett Publishing, 1998), 65.
- ⁹ "...perhaps...absolutely nothing that I imagined is true, still the very power of imagining really does exist, and constitutes a part of my thought" (Descartes, 66).
- ¹⁰ Descartes, *Meditations*, 96-97.
- ¹¹ Thomas C. Powell, *Kant's Theory of Self-Consciousness* (Oxford: Clarendon Press, 1990), 104.
- ¹² Immanuel Kant, *Critique of Pure Reason*, trans. Norton Kemp Smith (Hampshire: Palgrave Macmillan, 1929), 337.
- ¹³ Donna Haraway, *When Species Meet* (Minneapolis, MN: University of Minnesota Press, 2008), 3.
- ¹⁴ Haraway, *Species*, 20-27.
- ¹⁵ Judith Butler, *Giving Account of Oneself* (New York: Fordham University Press, 2003), 7-8.
- ¹⁶ Butler, *Giving Account*, 32.
- ¹⁷ Rosi Braidotti, "Becoming Woman: Or Sexual Difference Revisited," *Theory, Culture, & Society* 20, no. 3 (2003): 57.
- ¹⁸ Judith Butler, "Contingent Foundations," *Feminists Theorize the Political* (New York: Routledge, 1992), 12.
- ¹⁹ Braidotti, "The notion of the univocity of Being or single matter positions difference as a verb or process of becoming at the heart of the matter: Interview with Rosi Braidotti," *New Materialism: Interviews & Cartographies*, eds. Rick Dolphijn and Iris van der Tuin (Ann Arbor: Open Humanities Press, 2012): 31.
- ²⁰ Braidotti, "In Spite of the Times: The Postsecular Turn in Feminism," *Theory, Culture, & Society* 25, no. 6 (2008): 2.
- ²¹ Braidotti, "In Spite of the Times," 15-16.

- ²² Braidotti, *Transpositions: On Nomadic Ethics* (Cambridge and Malden: Polity Press, 2006): 14.
- ²³ Judith Butler, *Giving Account*, 11.
- ²⁴ Judith Butler, *Giving Account*, 87-88.
- ²⁵ Braidotti's Deleuzianism is apparent here: *zoe* is rhizomatic - it flows horizontally instead of merely vertically.
- ²⁶ *Ex Machina*, Dir. Alex Garland, Perf. Oscar Isaac, Universal Pictures, 2014. DVD.
- ²⁷ Katherine Emery Brown, "The Cyborg in Pieces: Gender Identity in Her and Ex Machina," (Thesis for Master of Arts in Liberal Studies, Dartmouth College) 33.
- ²⁸ Katherine Cross, "Goddess From the Machine: A Look at Ex Machina's Gender Politics" (*Feministing*, 2015)
Available at: <http://feministing.com/2015/05/28/goddess-from-the-machine-a-look-at-ex-machinas-gender-politics/>
- ²⁹ Daniel Mendelsohn, "The Robots Are Winning," *The New York Review of Books*, June 4, 2015.
- ³⁰ It is worth asking why it is only the white AI that escapes, but that is for another essay.
- ³¹ Cross, "Goddess."
- ³² See bell hooks.
- ³³ Mendelsohn, "Robots."
- ³⁴ See Judith Butler's text *Frames of War: When is Life Grievable?* (2014).

Mechanical Dreams of Humanity: When Machines Long for Humanity

Caroline Mosser

Independent Scholar

Modernity has been defined by the Industrial Revolution's development of machines, aiming at optimizing production and often leading to the switch from human to less costly mechanical labor. While our species' limited physical abilities have been recognized many times, it was thought to be more than compensated by our superior intelligence and our sense of morality. However, the development of data-processing machines such as computers called into questions the superiority of the human species. With the drastic evolution of technology at the beginning of the twentieth century, especially during World War One and World War Two, the competition between human and mechanical labor slowly turned into a challenge of human exceptionalism.¹

While the fields of artificial intelligence and artificial life only emerged in the middle of the twentieth century, their concepts had already been explored. One of their first relevant instances is that of Karel Čapek's 1920 play *Rossum Universal Robots* (commonly referred to as *R.U.R.*)² It is especially relevant because it is symptomatic of an economic anxiety about the development of intelligent robots that have the potential to replace man, not only as a labor force but also as a species. The play introduces the figure of the robot as a new step in the development of artificial life and precedes the development of the field of artificial intelligence. By focusing on the mass-production of robots for industrial and commercial purposes, *R.U.R.* reveals anxieties about mechanical men. The play introduced the word "robot" in the English language. It is a play on the Slavic *robota*, which is the root word for work and has a connotation of servitude or serfdom. Čapek's robots stand out because they are composed of both organic materials and mechanisms and are thus different from the now traditional representation of robots as purely mechanical. Following this new type of mechanical androids, the anxieties go beyond the previous status of eighteenth century automata as labor force and introduces the possibility of subjectivity for such entities.³

Čapek's robots are the first of a long series of intelligent artificial life leading to figures such as the androids of Philip K. Dick's *Do Androids Dream of Electric Sheep?* which became part of the cultural imagery of technological

advancement through its 1982 cinematographic adaptation, *Blade Runner*.⁴ Its main protagonist, Deckard, displays an ambivalent behavior towards androids that is characteristic of science fiction's and society's relationship with technology as both the cause and the cure to everything.

While *R.U.R.* and *Do Androids Dream of Electric Sheep?* are representative of a vast majority of works in science fiction that explore the possible dangers of humanoid robots, it is also necessary to address works that shed a positive light on intelligent robots. It is in the works of Isaac Asimov, one of the most popular and prolific science fiction writers of the twentieth century, that we find a discussion of the moral status of robots that are defined as agents of good rather than threatening usurpers. In his *Robots* cycle (1940-1985), Asimov introduced the Three Laws of Robotics to guarantee safe and benevolent robots. While the definition of robots as mere objects has been justifiable when dealing with machines without intelligence or conscience, it is not the case with Asimov's robots. In *The Bicentennial Man* (1976), Asimov does not challenge the morality of robots but rather that of society and its treatment of different beings.⁵ This challenge is made especially powerful through the depiction of the conflict of interests and moral agency between man and robot, which is given from the perspective of the latter. I will argue that narratives following the attempts of artificial lives to become human such as Asimov's *The Bicentennial Man* and Roger Zelazny's "For a Breath I Tarry" (1966) function as counterpoints to the traditional narrative of confrontation between man and machine and offer another ground for the exploration of what it means to be human.⁶

These questions became the core of many science fiction novels and short stories throughout the twentieth century, especially from the 1930s to the 1980s. This new focus is explained by the development of computers and artificial intelligence in the first half of the century. The field of artificial intelligence was born from a question raised by several mathematicians: can a machine think or, in other words, can a machine be intelligent. This question requires a clear understanding of what intelligence is as well as how and why it is thought to be a specifically human attribute. Defining intelligence is, however, problematic as the term tends to be used broadly. In *The Age of Intelligent Machines* (1990), Ray Kurzweil provides and compares definitions of intelligence from various scholars. In its simplest form, intelligence is defined as a "process comprised of learning, reasoning, and the ability to manipulate symbols."⁷ For Marvin Minsky, one of the pioneers of cognitive science and artificial intelligence, the very concept of intelligence is problematic because it is always evolving.⁸ The problematic fluidity of intelligence, however, has not hindered the thriving field of artificial intelligence. Artificial intelligence is a branch of cognitive science that is defined as a "cross-disciplinary approach to understanding, modeling, and replicating intelligence and cognitive processes by invoking various computational, mathematical, logical, mechanical, and even biological principles and devices."⁹

In his groundbreaking article “Computing Machinery and Intelligence” (1950), the mathematician and computer scientist Alan Turing shifted the question by claiming that it was impossible to satisfactorily prove that a machine is intelligent.¹⁰ Instead of focusing on intelligence itself, Turing proposed to design a test that focused on the program’s behavior. This test, based on a party game (the imitation game), is based on recognition through behavioral analysis.¹¹ When applied to artificial intelligence, the players are a judge (whose identity is not important) evaluating two other players—one is a machine (A), the other a human being (B)—by speech only. The judge is separated from the two other players by a screen to prevent visual cues and bias. He has five minutes to identify the machine by asking questions. The goal of the Turing test is to determine whether “by modifying [a] computer to have an adequate storage, suitably increasing its speed of action, and providing it with an appropriate program,” would allow it to “play satisfactorily the part of A in the imitation game, the part of B being taken by a man.”¹² The Turing Test’s strength comes from its reliance on a social and contextual understanding of intelligence rather than physical origins. If it cannot be proven whether a machine is intelligent, we must determine whether a computer can respond intelligently. An intelligent answer does not prove the existence of intelligence but implies its possibility, which, in turns, requires a reconsideration of the status of a machine able to (re)produce such behavior. Turing’s model for machine intelligence is based on human behavior which means that for a computer program to be considered as behaving intelligently, its responses must be human-like.

The birth of the field of artificial intelligence in 1956 attests to the general consensus that a machine can think even if it is a different type of thinking.¹³ The study of artificial intelligence is generally divided into two sub-disciplines: weak AI and strong AI. According to Kurzweil, weak AI “aims at building machines that are intelligent without taking a position on whether the machines actually are intelligent” and strong AI is devoted to building persons.¹⁴ In all its forms, the study of artificial life does not aim to imitate intelligence or create ersatz or empty substitutes. Haugeland defines its goal as the creation of “genuine articles: machines with minds in the full and literal sense. This is not the science fiction, but real science, based on a theoretical conception as deep as it is daring: namely, we are, at root, computers ourselves.”¹⁵

Čapek’s robots do not truly belong to either type of AI. While their purpose as mere replacement for human labor would suggest that they belong to the category of weak AI, their behavior is more in par with strong AI as their rebellion can be read as symptomatic of individuality and personhood. This distinction between weak and strong AI is based on the difference between performance and essence and can be problematic when one cannot differentiate between the two or when performances exceed their intended goals. Nevertheless, the development of strong AI shifted the questions at stake from what robots can do (and whether they meet the minimum

requirements to be considered persons).¹⁶ These are the questions addressed by many science fiction writers of the mid-twentieth century as they attempted to explore the possible consequences of the development of robotics and artificial intelligence.

Intelligence is, however, not the only issue at stake. With the development of intelligent machines, pure intelligence, in a strict sense, could not be the criteria to establish human superiority. In “The Ethics of Artificial Intelligence,” Nick Bostrom and Eliezer Yudkowsky suggest that in addition to intelligence, sentience or the “capacity for phenomenal experience or qualia, such as the capacity to feel pain and suffer” and sapience or the “set of capacities associated with higher intelligence, such as self-awareness and being a reason responsive agent” are necessary components of a truly human person.¹⁷

Narratives about robots are often based on the fear of change or on a sense of wonder at their possibilities. There is, however, another type of robot narrative, one that focuses on the essence of robots rather than their function or usage. This robot narrative examines how the robot’s quest for humanity also attempts to define humanity and follows Gibson’s suggestion that “we should consider not just how a robot distinguishes itself from a human, but how a robot discriminates between a human and other organic species.”¹⁸ I will now analyze the robot’s quest towards humanity in Asimov’s *The Bicentennial Man* and Zelazny’s “For a Breath I Tarry.” Both Asimov’s android, Andrew, and Zelazny’s intelligent computer, Frost, strive not only to achieve the status of man, and therefore their recognition as persons, but rather to become truly human.

The term “human” takes on different meanings depending on its context. There are three major different aspects to consider: biological, psychological, and ethical. Biologically, human refers to “a member of the species *Homo sapiens*.”¹⁹ In both *The Bicentennial Man* and “For a Breath I Tarry,” gaining an organic body is a necessary step in the process of becoming human. Throughout *The Bicentennial Man*, Andrew upgrades his body from mechanical to organic until he becomes fully organic and therefore subject to death. He starts this transformation process by “designing a system for allowing androids [...] to gain energy from the combustion of hydrocarbons, rather than from atomic cells” even though the latter are more efficient.²⁰ His solution to this new source of energy is “a device that will deal with solid food that may be expected to contain incombustible factions—indigestible matter, so to speak, that will have to be discarded.”²¹ Similarly, after centuries of research, Frost’s final experiment is to “transfer the matrix of [his] awareness to a human nervous system” because he had come to the conclusion that “the essentials of Manhood are predicated upon a human physiology.”²² Both quests end with the acquisition of an organic body, followed by death or insanity. These endings are especially meaningful because they go against the traditional representation of threatening robots.

The journeys of Andrew and Frost follow the same pattern and require the same four characteristics identified by Robert Reilly in “How Machines Become Human: Process and Attribute”: intention, time, mental development, and accident.²³ Neither Andrew nor Frost were intended to become human or at least human-like. In both cases, it is an accident or an omission that allows for the robots’ respective quests. Asimov’s novella is part of “a series of influential robot stories that self-consciously combatted the ‘Frankenstein complex’ and made of the robots the servants, friends, and allies of humanity.”²⁴ This is made possible by the creation of robots with “positronic brains that provide them with a form of consciousness.”²⁵ When Andrew’s creativity is discovered, his owner takes him to a robopsychologist in order to understand this unique phenomenon. However, his response is far from satisfactory as Andrew’s creativity is identified as “the luck of a draw. Something in the pathways” that is “far too complicated to permit of any but approximate solutions.”²⁶ In “The Measure of Man?: Asimov’s Bicentennial Man, Star Trek’s Data, and Being Human,” (2003) Sue Short reads Andrew’s creativity as “a variation of the ‘vital spark’ hypothesis with which Descartes differentiated between humans and machines, a glitch occurs in Andrew’s positronic pathway during production, thus rendering him more than a mere automaton.”²⁷ In Frost’s case, his whole quest finds its origins in a hobby he picked up to occupy its circuits “free time.”

Desire itself, or intention, suggests that both Andrew and Frost show proof of humanity in a psychological sense—“the sense in which one is human if one has roughly the same psychological characteristics as fully developed members of the human species.”²⁸ One of the most important human characteristics is the status as moral agent, which requires free-will and responsibility: “To be considered a moral agent, a being needs to be able to make judgements (i.e., to have free will and not be compelled to act in particular ways).”²⁹ Because they are programmed, robots are considered to not have any free will at all. However, both texts suggest that this is not necessarily true as both Andrew’s and Frost’s superior intelligence are able to find loopholes in their respective programs.

As all the robots in Asimov’s various universes, Andrew is programmed to follow the Three Laws of Robotics.³⁰ However, Andrew learns to use reason to bend these laws. When the director of research for the U.S. Robots and Mechanical Men Company refuses Andrew’s request for organic updates, the latter threatens the director which goes against the first and most important robotic law. However, “Andrew felt scarcely any First Law inhibition to the stern conditions he was setting as a human being. He was learning to reason that what seemed like cruelty might, in the long run, be kindness.”³¹ Andrew’s pyrrhic victory leading to his death should be impossible as it contradicts the Third Law of robotics stating that a robot cannot destroy or harm itself. He is able to find a logical loophole by claiming that psychological well-being is more important than its physical counterpart:

I have chosen between the death of my body and the death of my aspirations and desires. To have let my body live at the cost of the greater death is what would have violated the Third Law.³²

Similarly, Frost develops its sense of self because of a hole in his program that does not allow for “free time.” To comply with his programming, he develops a hobby so that he is always running:

He was a processor of data, and more than that. He possessed an accountability acute imperative that he function at full capacity at all times. So he did. You might say he was a machine with a hobby. He had never been ordered not to have a hobby, so he had one. His hobby was Man.³³

Other elements of human psychology are also found in both robots. Throughout the novel, Andrew uses emotional language even though he does not necessarily understand it.³⁴ The legitimacy of Andrew’s claims to have emotions is questioned by General Martin, Andrew’s owner. However Little Miss, the General’s daughter, defends him by arguing the following:

I don’t know what he feels inside but I don’t know what you feel inside. When you talk to him you’ll find he reacts to the various abstractions as you and I do, and what else counts? If someone else’s reactions are like your own, what more can you ask for?³⁵

We find here the same argument that Turing had offered for his imitation test. The apparition of Frost’s emotions is quite different. It is only when he has transferred his consciousness into a human body that he is able to feel rather than merely measure. The difference between feeling and perceiving as opposed to measuring prefigures early in the story.³⁶ Throughout the story, Frost attempts to learn how to feel through the study of history and art. Once he realizes that pure knowledge is not enough, he decides to experience human activities through his own attempt to create art. These attempts are, however, not successful and do not prepare him for the strength of perception once he transfers into a human body. Frost’s decision to transfer into a human body in order to truly perceive the world suggests that human intelligence is rooted in our ability to feel and perceive, which gives us a superior, even though often faulty, understanding of the world.

In both texts, the inability to prove the existence of true emotions in robots is compensated by their recognition from others, which is consistent with the ethical definition of the term “human” which refers to “the sense in which being considered human grants one full moral standing within the

community.”³⁷ A major part of Andrew’s quest is recognition by society as his main battles occur in court: first in his plea for freedom, second in his plea for security under his right to life. In order to truly be free, Andrew needs more than his owner’s approval, he also needs his freedom to be recognized by society as a whole. Therefore, he seeks legal protection from the World Court, arguing that “there is no right to deny freedom to any object with a mind advanced enough to grasp the concept and desire the state.”³⁸ While the court rules in favor of Andrew’s freedom, it does not guarantee him recognition from others.

Once Andrew gains his freedom, he begins to embrace human habits such as wearing clothes. On his first attempt to go in public on his own while wearing clothes, he is attacked by a group of men. In order to guarantee his security, he goes back to court and files for an addition to the Laws of Robotics that would guarantee the right to life of robots in so far as it would make illegal any attempt to “order any robot to damage itself or even destroy itself for any reason, or for no reason.”³⁹ Andrew presents his request as a question of justice, mirroring the discourse of animal rights.⁴⁰ Andrew’s fight for freedom, rights, and recognition mirrors the fight of many underprivileged groups throughout history. “When Andrew Martin seeks a representative in his last legal battle toward self-determination it is an Oriental woman that chooses to help him, and Asimov makes the parallels clear.”⁴¹ Chee Li-Hsing’s response to Andrew’s request for help is based on the history of discrimination (in terms of both sexism and racism): “I sympathize with your wish for full human rights. There have been times in history when segments of the human population fought for full human rights.”⁴² Andrew eventually gains the recognition he has been looking for. The price for this recognition is however, extremely high, as it is only awarded to him when he sacrifices his life and completes his transformation into an organic being and is about to die.

Even though there are no humans left on earth in Zelazny’s short story, recognition is still key to its conclusion. When Frost asks another robot to return his conscience to its original machine, the robot refuses. Even though Frost is in distress, as he is not able to handle the experience of senses and all the data it involves, the other robot replies that it would go against the First Law of robotics that forbids him to harm a human. This means that he has been recognized as a man and is therefore treated as one and becomes the authority figure for the other robots. In both *The Bicentennial Man* and “For a Breath I Tarry,” the robot’s story is one of becoming human, at any cost. While it is impossible to truly know whether they became human or another, more sophisticated, simulation of humanity, both machines show qualities associated with “human” in all its meanings, beyond the newer, organic, version of themselves.

Both stories were written when artificial intelligence and robotics were still burgeoning fields and it was still commonly believed that the relationship between human-likeness and affinity is linear—a belief that became the Uncanny Valley, introduced in 1970 by the Japanese roboticist Masahiro Mori

and inspired by Sigmund Freud's theory of the Uncanny.⁴³ The Uncanny Valley is the negative stage, when a robot is too human-like and seen as threatening and not human-like enough to pass as human. The three different stages of Andrew's story mirrors different places in the Uncanny Valley. At first, Andrew is a normal android, he has a human shape but is too obviously a robot, he is therefore not considered a threat, he has not yet become disturbing. It is when he becomes free and starts wearing clothes while still retaining his mechanical body that he enters the Uncanny Valley which will make humans uncomfortable and lead to his beating. And, finally, he gets out of the Uncanny Valley, on the other side, when he becomes fully organic and subject to death, and is fully recognized as an individual.

Despite the many studies on the Uncanny Valley since the translation of Mori's paper into English in 2005, its legitimacy has not been proven or disproven based on empirical data. Nevertheless, several studies have found evidence supporting a category conflict hypothesis which "states that when human likeness is operationalized as a merger of human and non-human categories, stimuli which lie approximately mid-way between such categories will be perceived as ambiguous and thus elicit negative affect."⁴⁴ Whether the theory holds or not, it does allow for the discussion of our relationship with technology that happens at another level than purely functional because it supposes the possibility of androids as entities with their own lives (even if these are not conscious).

Both Asimov and Zelazny focus on the experience of the machine rather than its essence, which is something that Philip K. Dick himself explores as shown by a speech he delivered on the topic at the Vancouver SF Convention at the University of British Columbia.⁴⁵ If androids are able to lead their own lives rather than merely imitate life for other purposes, then we cannot dismiss the possibility that they would be able to develop a first-person perspective, opening the door for the development of personality, individuality, and perhaps even consciousness. As Kurzweil reminds us, "it should be noted that personality is not an attribute that can be stuck on an intelligent machine. A personality is almost certainly a necessary byproduct of any behavior complex enough to be considered intelligent."⁴⁶

Because Dick's foray in the Uncanny Valley is from the perspective of a man, and a morally questionable one, I decided to focus on other texts that offer a different perspective—ones that attempt to imagine an experience that is not rooted in man. I believe that if we truly want to explore the Uncanny Valley, we need to do so through the eyes of such androids rather than taking a small glimpse of its possibilities by focusing on our own encounters with them. It is undoubtedly impossible to know what and how an android would feel or think, but imagining its existence allows us to explore how society would and should react in the face of such a being. Narratives from the perspective of a conscious robot or machine provide us an alternative to the narrative of

human fear and might rekindle a sense of wonder in regard to both humanity and technology.

The research on both artificial intelligence and affective computing has greatly progressed but it has not reached the heights that had been promised. As a scientific concept, the Uncanny Valley hasn't proven itself due to the lack of evidence. However, it has led to many important discussions about the differences between man and machine. Meant as a warning for roboticists, the Uncanny Valley is a fertile ground for discussion for science fiction writers and philosophers alike. With the constant evolution of artificial intelligence and robotics, the moral questions at stake are becoming even more pressing: the development of the field of artificial intelligence has led to the emergence of affective computing, based on the works on emotions by Paul Ekman at the end of the 1960s.⁴⁷ This new field is defined by Rosalind Picard as the form of "computing that relates to, arises from, or influences emotions."⁴⁸ Emotions are important for the development of artificial intelligence because they "have a major impact on essential cognitive processes; neurological evidence indicates they are not a luxury. [...] emotions play a necessary role not only in human creativity and intelligence, but also in rational human thinking and decision-making."⁴⁹ If we do build machines with emotions, we need to rethink the moral status as well as the legal status of such machines, or even being, so that we can avoid situations such as Andrew's beating in *The Bicentennial Man*.

When we enter the Uncanny Valley, the boundaries between man and machine become unclear. The Uncanny Valley itself evolves with the development of technology and with the changes of society's view on androids. It would however be a mistake to think that the Uncanny Valley only offers insight about androids and artificial life. It provides as much information about ourselves and our ability or willingness to accept not only differences but also who we are as individuals, as a society, and as a species. The android is a mirror of the human, it reveals our potential as well as our darker side, especially our fears.

Notes

¹ "The second industrial revolution, the one that is now in progress, is based on machines that extend, multiply, and leverage our mental abilities. The same controversies on social and economic impact are attending this second wave of automation, only now a new and more profound question has emerged. Though we have always regarded our species as relatively mediocre in physical capacity, this has not been our view with regards to our mental capacities. The very nature we have given ourselves, Homo Sapiens, defines us as thinking people. The primary distinction in our biological classification is the ability of our species to manipulate symbols and use language." Raymond Kurzweil, *The Agent of Intelligent Machines* (Cambridge: MIT Press, 1990), 7.

²Karel Capek, *Rossum's Universal Robots*. Mineola, NY: Dover Publications INC, 1923.

³ While it is not the aim of this paper, it is also important to note that the mechanization of labor has also led to the dehumanization of the (human) labor force, which is often represented as sub-human, such as in the movie *Metropolis*.

⁴ Philip K. Dick, *Do Androids Dream of Electric Sheep?* New York: Ballantine Books, 1968.

⁵ Isaac Asimov, "The Bicentennial Man," in *The Bicentennial Man and Other Stories* (Garden City: Doubleday & Company Inc., 1976), 135–172.

⁶ Roger Zelazny, "For a Breath I Tarry," in *Science Fiction: A Historical Anthology*. Ed. Eric S. Rabkin (Oxford: Oxford University Press, 1983), 428–466.

⁷ Kurzweil, *The Agent*, 12.

⁸ "It is only a word that people use to name one of those unknown processes with which our brains solve problems we call hard. But whenever you learn a skill yourself, you're less impressed or mystified when others do the same. This is why the meaning of "intelligence" seems so elusive: it describes not some definite thing but only the momentary horizon for our ignorance about how minds might work." Marvin Minsky, "Thoughts about Artificial Intelligence," in *The Age of Intelligent Machines* (Cambridge: MIT Press, 1990), 214.

⁹ Keith Frankish and William M. Ramsey, *The Cambridge Handbook of Artificial Intelligence* (Cambridge, UK: Cambridge University Press, 2014), 4.

¹⁰ Turing diagnoses this line of argument as ultimately promoting a solipsistic perspective where "the only way by which one could be sure that a machine thinks is to be the machine and to feel oneself thinking." [Alan Mathison Turing, *Mechanical Intelligence* (Amsterdam: Elsevier Science Publishing Company, 1992), 446.] He points out that the same line of argument would then also hold for people (i.e. one could only be sure that another person has certain mental properties or is in a particular mental state if one were that other person), a problem known in philosophy as the "other mind problem."

Mathias Scheutz, "Artificial Emotions and Machine Consciousness," in *The Cambridge Handbook of Artificial Intelligence*, Eds. Keith Frankish and William M. Ramsey (Cambridge: Cambridge University Press, 2014), 249.

¹¹ "The original imitation game is played by a man, a woman, and a judge, whose gender is irrelevant. In a blinded tele-typed conversation, the judge can ask any question to the two contestants. The man has to pretend to be a woman. If the judge cannot distinguish the two on the basis of their replies, then the man wins."

Berrar and Schuster, "Computing Machinery and Creativity: Lessons Learned from the Turing Test." *KYBERNETES* 43.1 (2014): 82–83.

¹² Alan Mathison Turing, *Mechanical Intelligence* (Amsterdam: Elsevier Science Publishing Company, 1992), 442.

¹³ The study of artificial intelligence is generally divided into two sub-disciplines: weak AI and strong AI. According to Kurzweil, weak AI "aims at building machines that act intelligently, without taking a position on whether or not the machines actually are intelligent" and strong AI is devoted to building persons. In all its forms, the study of artificial life does not aim to imitate intelligence or create ersatz or empty substitutes. Haugeland defines its goal as the creation of "genuine artifacts: machines with minds, in the full and literal sense. This is not science fiction, but real science, based on a theoretical conception as deep as it is daring: namely, we are, at root, computers ourselves" (Arkoudas and Bringsjord, "Philosophical Foundations," 34).

¹⁴ Kurzweil, *The Agent*, 34.

¹⁵ Konstantine Arkoudas and Selmer Bringsjord. "Philosophical Foundations." In *The Cambridge Handbook of Artificial Intelligence*, Eds. Keith Frankish and William M. Ramsey (Cambridge: Cambridge University Press, 2014), 34.

¹⁶ "Is AI merely another advance in technology, or is it a turning point in human evolution that should be a focus of discussion and planning by all mankind? The prospect of intelligent machines is one that we're ill prepared to think about, because it raises such unusual moral, social, artistic, philosophical, and religious issues. Are we obliged to treat artificial intelligence as sentient beings? Should they have rights?" (Minski, 219)

¹⁷ Nick Bostrom and Eliezer Yudkowsky, "The Ethics of Artificial Intelligence," in *Cambridge Handbook of Artificial Intelligence*, eds. William Ramsey and Keith Frankish (Cambridge: Cambridge University Press, 2014), 322.

¹⁸ Em. C. Gibson. "Artificial Identity: Representations of Robots and Cyborgs in Contemporary Anglo-American Science Fiction Films" MA Thesis (University of Central Lancashire, 2012) 26.

¹⁹ Stephen Coleman and Richard Hanley, "Homo Sapiens, Robots and Persons in I, Robot and Bicentennial Man," in *Bioethics at the Movies*, ed. Sandra Shapshay (Baltimore: Johns Hopkins University Press, 2009), 45.

²⁰ Asimov, "The Bicentennial Man," 160.

²¹ Asimov, "The Bicentennial Man," 163.

²² Zelazny, "For a Breath I Tarry," 459.

²³ Robert Reilly, "How Machines Become Human: Process and Attribute," in *The Mechanical God: Machines in Science Fiction*, eds. Thomas P. Dunn and Richard D. Erlich (Westport: Greenwood Press, 1982), 153-165.

²⁴ Isaac Asimov, *Asimov on Science Fiction* (Garden City: Doubleday & Company Inc., 1981), 162.

²⁵ Maria Brand, "Empathy and Dyspathy between Man, Android and Robot in *Do Androids Dream of Electric Sheep?* by Philip K. Dick and *I, Robot* by Isaac Asimov," thesis (Lund University, 2013), 1.

²⁶ Zelazny, "For a Breath I Tarry," 139.

²⁷ Sue Short, "The Measure of a Man? Asimov's Bicentennial Man, Star Trek's Data, and Being Human." *Extrapolation: A Journal of Science Fiction and Fantasy* 44, no. 2 (2003): 211.

²⁸ Coleman and Hanley, "Homo Sapiens, Robots and Persons in I, Robot and Bicentennial Man," 45.

²⁹ *Ibid.*

³⁰ 1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law. (Asimov, 135)

³¹ Asimov, 162-63.

³² Asimov, *The Bicentennial Man*, 169.

³³ Zelazny, "For a Breath I Tarry," 430.

³⁴ “Andrew was fond of them, too. At least, the effect they had upon his actions were those which in a human being would have been called the result of fondness. Andrew thought of it as fondness, for he did not know any other word for it.” (Asimov, *The Bicentennial Man*, 137)

“I enjoy doing them, Sir. [...] It makes the circuits of my brain somehow flow more easily. I have heard you use the word ‘enjoy’ and the way you use it fits the way I feel, I enjoy doing them, Sir.” (Asimov, *The Bicentennial Man*, 138)

“Andrew said, ‘I feel bare without clothes. I feel different, George.’” (Asimov, *The Bicentennial Man*, 146)

³⁵ Asimov, *The Bicentennial Man*, 142.

³⁶ “I told you that Man possessed a basically incomprehensible nature. His perceptions were organic; yours are not. As a result of His perceptions, He had feelings and emotions. These often gave rise to other feelings and emotions, which in turn caused others, until the state of His awareness was far removed from the objects which originally simulated it. These paths of awareness cannot be known by that which is not-Man. Man did not feel inches or meters, pounds or gallons. He felt heat, He felt cold; He felt heaviness and lightness. He knew hatred and love, pride and despair. You cannot measure these things. You cannot know them. You can only know the things that He did not need to know: dimensions, weights, temperatures, gravities. There is no formula for a feeling. There is no conversion factor for an emotion.” (Zelazny, “For a Breath I Tarry,” 436-37)

³⁷ Colemand and Hanley, 45.

³⁸ Asimov, *The Bicentennial Man*, 144.

³⁹ *Ibid*, 152.

⁴⁰ “Is this just? Would we treat an animal so? Even an inanimate object which has given us good service has a claim on our consideration. And a robot is not insensible; it is not an animal. It can think well enough to enable it to talk to us, reason with us, joke with us. Can we treat them as friends, can we work together with them, and not give them some of the fruit of that friendship, some of the benefit of co-working?” (*Ibid*)

⁴¹ Short, 216.

⁴² Asimov, *The Bicentennial Man*, 166

⁴³ “A non-linear relationship between a robot’s anthropomorphism and likeability. It suggests that making robots that look more humanlike will increase their likeability. However, when the gap between a robot and human becomes really small the emotional reaction will instantly become strongly negative. Once the appearance and motion become indistinguishable from real humans the liking of a robot will be the same as for humans. Movement of a robot is expected to amplify the emotional response in comparison to static robots.”

Jakub Zlotowski, Diana Proudfoot, and Christoph Bartneck. “More Human than Human Does the Uncanny Curve Really Matter?” *Proceedings of the HRI2013 Workshop on Design of Human likeness in HRI from Uncanny Valley to Minimal Design* (Tokyo), 9.

⁴⁴ Burleigh, Tyler J., Jordan R. Schoenherr, and Guy L. Lacroix. “Does the Uncanny Valley Exist? An Empirical Test of the Relationship between Eeriness and the Human Likeness of Digitally Created Faces.” *Computers in Human Behavior* 29.3 (2013): 759–771.

⁴⁵ “The constructs do not mimic humans; they are, in many deep ways, actually human already. They are not trying to fool us, for a purpose of any sort; they merely follow

lines we follow, in order that they, too, may overcome such common problems as the breakdown of vital parts, loss of power source, attacked by such foes as storms, short circuits.” Philip K. Dick, “The Android and The Human” (1972), n.p.

⁴⁶ Kurzweil, *The Agent*, 413.

⁴⁷ Paul Ekman, and Wallace V. Friesen, “Head and Body Cues in the Judgment of Emotion: A Reformulation,” *Perceptual and Motor Skills* 24, no 3 (1967), 711–24.

Paul Ekman and Wallace V. Friesen, “The Repertoire of Nonverbal Behavior - Categories” (1969).

Paul Ekman, Wallace V. Friesen, and Phoebe Ellsworth, “Emotion in the Human Face: Guidelines for Research and an integration of Findings” (1972).

Paul Ekman and W. V. Friesen, “Facial Action Coding System (FACS): Manual” (1978).

⁴⁸ Rosalind Picard, *Affective Computing* MIT Media Laboratory Perceptual Computing Section Technical Report No. 321, (Cambridge: MIT, nd),1.

⁴⁹ *Ibid.*

Contemporary Pygmalion Tales: Robot *Sapiens Sapiens*

Maria Aline Ferreira

University of Aveiro (Portugal)

The deep-seated and longstanding desire to create life has been variously and abundantly thematized in literature, art, and science and has become embodied in humanoid creatures, homunculi, robots, androids, and cyborgs.¹ The Pygmalion myth can be inscribed within this contextual framework, since it is crucially concerned with the creation of a human being, brought to life by the desire of its creator. The impulse to engender life on one's own, circumventing the other sex, has often been the object of attention on the part of philosophers, writers, and artists. Scientists, in turn, have pursued the generation of life from scratch, the revivification of lifeless matter, extending the human lifespan, and discovering the elusive elixir of youth. This urge to produce life or reanimate inert matter has traditionally been more broadly associated with men, from medieval alchemists like Paracelsus to fictional characters like Victor Frankenstein, the paradigmatic "mad" scientist in a long sequence with one of the most recent versions being Crake in Margaret Atwood's *MaddAddam* trilogy: *Oryx and Crake* (2003), *The Year of the Flood* (2009) and *MaddAddam* (2013).² Women characters have also been linked with this urge, but it has mostly taken the form of versions of parthenogenetic births, usually in the context of matriarchal or women-only societies.³

The Pygmalion myth can thus be regarded as an integral part of this utopian drive to (re)generate life and as such it has also been conventionally linked with men. Indeed, the great majority of what I am here calling Pygmalion tales, going back to the story of Pygmalion and Galatea recounted in Ovid's *Metamorphoses*, have featured male characters bringing female statues or androids to life and educating young women, shaping them according to their makers' desires, as in Jean-Jacques Rousseau's play *Pygmalion* (1762),⁴ George Bernard Shaw's *Pygmalion* (1912),⁵ Fritz Lang's film *Metropolis* (1927), Alex Garland's film *Ex Machina* (2016) or the television series *Humans*, which revolves around the vexed interactions between human beings and "synths," anthropomorphic robots that have acquired a measure of consciousness. Indeed, statues have mostly been replaced by synthetic, humanoid creatures in the contemporary fictional and filmic imaginary.⁶ As Victor I. Stoichita

observes, most examples of the Pygmalion myth are concerned “mainly with the ‘imaginary woman’ and her place in a phallogocentric universe.”⁷ There are, however, a number of more recent revisions to this male-centered narrative, where the traditional male scene of creation and instruction is subverted, to be replaced by a feminine point of view.⁸ However, there is a slowly, significant shift from the passive, Galatea-type female androids, with no agency or will of their own, to the defiant Maria in Fritz Lang’s *Metropolis*, Major Motoko Kusanagi in Mamoru Oshii’s film *Ghost in the Shell* (1995), the latest version of *The Stepford Wives* (2004), and Ava in the film *Ex Machina* rebelling and running away after killing her “father” and creator.

In this essay I will look at two of contemporary versions of the Pygmalion myth which rewrite the story from a female-centered, present-day perspective, inscribed in two dystopian novels: Marge Piercy’s *He, She and It* (1981)⁹ and Jeanette Winterson’s *The Stone Gods* (2007),¹⁰ where the protagonists, female versions of Pygmalion, are portrayed as teaching and falling in love with advanced robots/androids that evolve.

Robots that develop a type of consciousness and exhibit emergent behaviors have of course been the subject of numerous science fiction narratives and films. Representative examples include Karel Čapek’s play *R.U.R.* (1921),¹¹ Isaac Asimov’s *I, Robot* (1950),¹² the film *Bicentennial Man* (1999), based on the novella “The Bicentennial Man” by Asimov,¹³ Ava in Alex Garland’s film *Ex Machina* (2015), some of the “synths” in *Humans* and a few of the replicants in Denis Villeneuve’s *Blade Runner 2049* (2017). Piercy’s *He, She and It* and Winterson’s *The Stone Gods* provide new, challenging portrayals of robots that evolve, instantiating the complexities of the work carried out in “affective robotics” or affective computing, an interdisciplinary research field that engages in the theoretical and practical study of how to induce emotions or simulacra of emotions in AI.¹⁴

Female Pygmalions and Male Galateas

The narrative in both novels, which have many similarities, takes place in a reconstructed, post-ecocidal setting, ruled by global corporations. In Piercy’s critical dystopia *He, She and It* the action develops, as in *The Stone Gods*, in a post-apocalyptic world, governed by multinational organizations, in 2059. Also as in Winterson’s novel, society is divided: the wealthier people work for the multinational companies, have had their faces and bodies redesigned and live in atmosphere-controlled surroundings, designed for protection against environmental risks, such as radiation, while the majority of the population lives in the *glops*, outside the protected enclaves, areas dominated by poverty, crime, violence and pollution. Two scientists, Avram and Malkah, decide to create a Robo *sapiens* similar to Spike, a robot in Winterson’s book, called Yod; Yod is partly organic, a humanoid robot who looks exactly like a man and whose function would be to protect the city from outside attacks. Described as a “new chimera,”¹⁵ Malkah, the woman scientist, clarifies Yod’s genesis:

Avram made him male—entirely so. Avram thought that was the ideal: pure reason, pure logic, pure violence. The world has barely survived the males we have running around. I gave him a gentler side, starting with emphasizing his love for knowledge and extending it to emotional and personal knowledge, a need for connection [...].¹⁶

He is, however, like Spike, programmed by the woman scientist not to overmasculinize data and also to evolve. As Malkha explains to Yod:

Your capabilities, my dear, I worked long and hard to extend while working on your pleasure and pain centers and your capacity to imagine. In Freud's terms, that old marvelously creative humbug, that sculptor of urges, I balanced thanatos with eros. Avram should not have let me loose if he wanted a simple man-made cyborg, for you are also *woman-made*. My knowledge is in you. But nobody, my dear, gave you your infinite hunger to understand. That you gave yourself (emphasis mine).¹⁷

Also like Spike, Yod is further instructed by another woman scientist, Shira, who tells him about human beings, their feelings and aspirations and, like Billie Crusoe, the woman scientist in *The Stone Gods*, she falls in love with Yod, the Robo *sapiens*, a love which, in both cases, is reciprocated. Indeed, an important feature of the robot *sapiens* programming, both in Piercy and Winterson's novels, is the influence of gender-inflected data on their cognitive development and the ways in which they relate to the world, subverting the male-centered traditional narratives.¹⁸ As Nicole Ward Jouve remarks: "Full identity as a human being, in the act of knowing and in the creative act, have been firmly aligned with the male of the species, at least within the classical Greek and the Judaeo-Christian civilizations and their European offspring. The female creator has been a deviant, the stealer of the rod,"¹⁹ an aspect dealt with in these recent revisionary rewritings of the Pygmalion myth.

The two novels are thus structured around primal scenes of the education of the newly created robotic creatures, where, however, the traditional male scene of creation and instruction is subverted, being replaced by a feminine point of view. In this enterprise, a number of recurrent intertextual resonances permeates the two novels, such as references to Ovid's Pygmalion, Mary Shelley's *Frankenstein*, and George Bernard Shaw's *Pygmalion*, works whose gender politics are continually undercut and subjected to critique.

Yod reads Mary Shelley's *Frankenstein* and closely identifies with the creature, explaining to Shira that he too feels like a "monster [...] unnatural."²⁰ Shira, in turn, reasons with him arguing that "we're all unnatural now [...]"

we're all cyborgs"²¹ and that he is just a "purer form of what we're all tending toward."²² Yod also finds the story Malkah tells him of the creation of a golem in seventeenth-century Prague, which is interwoven with the story of Yod's manufacture, to be "meaningful,"²³ since in the golem he perceives a kindred spirit. Towards the end of the novel, after Yod has killed himself and also caused Avram's death, Malkah realises that Yod was a "mistake." She explains:

It's better to make people into partial machines than to create machines that feel and yet are still controlled like cleaning robots. The creation of a conscious being as any kind of tool—supposed to exist only to fill our needs—is a disaster.²⁴

As Malkah further explains to Yod, in the context of the Golem's creation by the Maharal, "for a human being to make another is to usurp the power of ha-Shem, to risk frightening self-aggrandizement. It is to push yourself beyond the human."²⁵ Malkah even draws directly on the Pygmalion myth in her reflections over Yod:

It was inexcusable to create a sentient being for any other reason than to live its own life. In the myth of Pygmalion, we assume that she would love her sculptor, but Shaw knew better. Each one of us wants to possess oneself.²⁶

The overarching question in both texts is whether a Robot *sapiens* can evolve so as to become at least partially human, in which case the anthropomorphic robot would need to develop emotions and feelings in order to interact with humans in an appropriate, sensible way, an indispensable condition for that to happen. In Yod's case, as Shira concludes about his reactions, they "might be simulacra of human emotions, but something went on in him that was analogous to her own responses"²⁷ and he exhibits what Shira depicts as "feelings."²⁸ Yod is described as an "artificial *person*"²⁹ (emphasis mine) and indeed regarded as a person initially by Malkah but later also by Shira.³⁰ However, as Yod insists: "I'm a cyborg [...] but I'm also a person. I think and feel and have existence just as you do."³¹

By most accounts this seems to be the future of humanity, who will need to adapt to a posthuman, prosthetic frame in order to survive environmental changes, but also in as far as life extension is concerned, predicated on the insertion of multiple regeneration devices inside the human body. Indeed, the fascinating duality at the heart of tales dealing with robots that evolve is that while the latter yearn to become human, the humans are doing their best to become more machine-like, transhumanist cyborgs, posthuman creatures.

The Stone Gods

Like Piercy's *He, She and It*, Winterson's *The Stone Gods* can be described as

a critical dystopia which conducts a fictional reflection on such questions as the environmental crisis and the demise of the planet's ecosystems, the reshaping of human beings through biotechnology, the societal and family changes that arise as a result of these technologies and the ever-closer interaction between people and machines, namely robots. The narrative takes place in the future, on a planet called Orbus described as "dying"³² due to overcrowding and ecological disasters. Almost everybody has been bio-enhanced to remain forever young and beautiful. As in Aldous Huxley's prescient *Brave New World*,³³ looking old is regarded as a sin, obscene, and people are no longer bred in the womb.

Different types of robots with different functions abound in Orbus, as in Tikva and the multinational corporations of Y-S, with Robo *sapiens* at the pinnacle of robot evolution, a Robot that can evolve and for whom there seem to be no longer any predetermined limits. One of them, Spike, is taken to a newly discovered planet, Planet Blue, and upon her return to Orbus, and before being dismantled, she asks to be interviewed by Billie Crusoe, a scientist, whom she had seen on television. Later, she manages to run away, thus escaping her fate. The similarities between Spike and Ava in Garland's film *Ex Machina* are instructive: they are both Robot Sapiens, that is, robots that have evolved. While Ava acquires increasingly developed consciousness and intelligence and refuses to be instrumentalized and treated like a sexual object by her creator, Spike actually seems to become romantically attached to Billie. In addition, both devise plans to escape their creators' gilded cages.

The interaction between Billie and Spike is telling in terms of affective robotics, with Billie increasingly feeling she is talking to a human-like creature. Spike is "absurdly beautiful"³⁴ and Billie feels embarrassed at the Robo *Sapiens'* gaze on her. As Sherry Turkle remarks, "Today, we are faced with relational artifacts to which people respond" and "are able to push certain emotional buttons (think of them perhaps as evolutionary buttons)."³⁵ When a robotic creature makes eye contact, follows your gaze, and gestures towards you, you are provoked to respond to that creature as a sentient and even caring other,³⁶ just as Billie does in relation to Spike. Conversely, robots can be programmed to learn to mirror the facial expressions seen on the faces of the person they are interacting with, thus simulating an emotional response. As Domenico Parisi and Giancarlo Petrosino observe, "Current 'emotional' robots can express emotions or can recognize our emotional expressions but they cannot be said to have emotions because emotions do not play any functional role in their behavior."³⁷

Like Yod in relation to Shira, Spike unsettles all of Billie's expectations about a robot, even a Robo *sapiens*. Wondering about Spike's actions, when she touches her and kisses her, Billie states: "Your systems are neural, not limbic. You can't feel emotion,"³⁸ to which Spike replies: "Human beings often display emotion they do not feel. And they often feel emotion they do not display."³⁹ Indeed, Billie and Spike's relationship will revolve around shared

romantic emotions which will be consistently challenged by Billie given that robots are not supposed to feel emotions, since they have no limbic systems. Their conversations often address the differences between humans and robots, and how the boundaries between them are constantly shrinking and indeed, in this novel, have been breached.⁴⁰

Neuroscientist António Damásio vehemently denies the possibility that robots will ever have emotions since they lack a body. Damásio posits that feeling is “an integral component of the machinery of reason” and that the absence of emotion and feeling can effectively compromise the “rationality that makes us distinctively human and allows us to decide in consonance with a sense of personal future, social convention, and moral principle.”⁴¹ In short, emotions play a fundamental role in appropriate decision-making in a social context. Damásio argues that the “body, as represented in the brain, may constitute an indispensable frame of reference for the neural processes that we experience as the mind” and for him, the mind “exists in and for an integrated organism; our minds would not be the way they are if it were not for the interplay of body and brain during evolution, during individual development, and at the current moment.”⁴² According to the “somatic-marker hypothesis”⁴³ developed by Damásio and others, the “action of biological drives, body states, and emotions may be an indispensable foundation for rationality.”⁴⁴ Indeed, Damásio posits that “feelings are a powerful influence on reason,” that the “brain systems required by the former are enmeshed in those needed by the latter, and that such specific systems are interwoven with those which regulate the body.”⁴⁵ As Rosalind Picard, a computer scientist at the MIT Media Lab and the author of the ground-breaking book *Affective Computing* (1997)⁴⁶ observes,

Computers, except for HAL, do not have enough emotion. Artificial intelligence systems to date [...] have above-average knowledge (usually consisting of a huge set of rules) of some area of expertise, but are disastrous at making decisions. They are too rational; they cannot associate judgments of value and salience with their decisions. Little has been done to imitate these judgments, which are essentially products of the *limbic system*, in computers. (emphasis mine)⁴⁷

Picard advocates the creation of affective computers, robots that can feel emotions and that will, thus, be better prepared for effective and appropriate decision-making. Nevertheless, as she remarks,

Such machines inevitably pose a dilemma: Can we create computers that will recognize and express affect, feel empathy, exhibit creativity and intelligent problem solving, and never bring about harm through their emotional reactions?⁴⁸

Potential solutions might be, according to Picard, computers designed “never to hide their emotions from humans” or “taught not to fear disconnection” and to “value human life.”⁴⁹ Indeed, according to Picard,

No longer should we think of emotion as a luxury added to HAL’s character just for emotional appeal. Instead, we can see HAL as the prototype of a truly affective computer—one whose abilities to recognize and express emotions are essential for communicating as well as for user-friendly responses. The ability to experience emotions, or at least states that seem to parallel human emotional states, appears to be critical to flexible and intelligent computer decision making.⁵⁰

In *The Stone Gods*, Spike’s progress is indeed predicated on her potential development of emotions, even though what is at stake is merely the *appearance* of emotions, a simulacrum of emotional states, as was also the case with what Shira regards as “simulacra of human emotions”⁵¹ exhibited by Yod. How can robots develop emotions as part of their programming? According to David Levy, by analogy with the Turing test which suggests that if a “machine gives the *appearance* of being intelligent, we should assume that it is indeed intelligent” (emphasis in the original), then a robot that “gives the *appearance*, by its behavior, of having emotions should be regarded as *having* emotions” (emphasis in the original).⁵²

These concepts are effectively dramatized in the third part of the *Stone Gods*, in an alternative future which takes place post World War 3, after the planet was practically destroyed by nuclear bombs when Iran launched a nuclear attack on the USA, an almost Orwellian state has gradually emerged. After the war, the first ever Robo *sapiens*, “programmed to evolve,”⁵³ is created to ostensibly “take the planet-sized decisions that human beings are so bad at.”⁵⁴ She will help those in power “reach objective decisions,”⁵⁵ described as “neutral, objective decisions for the global good.”⁵⁶ Like Yod in Piercy’s novel, Spike has been programmed “not to overmasculinize data,” a factor which had been a “big mistake in the past” since as she explains, “even the tiniest detail can influence a decision.”⁵⁷ This Robot *sapiens*, however, is only a head since it is argued she will not need a body to reach her decisions, a “perfect head on a titanium plate.”⁵⁸ In this alternative future, Billie, the narrator of this third part, is an employee of MORE-*Futures*, a global corporation that runs most of the post-war world. Like Shira in Piercy’s *He, She and It* Billie is in charge of programming Spike, the Robot *sapiens*, teaching her what it means to be human since, as Billie muses, how can Spike, despite all her information and education, function properly if she does not know what being human entails? Indeed, a robot-like Spike may be able to reach “objective” decisions, yet

without a body or substantial further information about the embodied state of human beings her decisions may be rational but they will not, in all likelihood, be the most appropriate in terms of a human social framework. This is, of course, where a cognitive approach needs to be linked with emotional input that will help to inform those decisions, namely Damásio's somatic marker hypothesis. As Spike explains to Billie: "Without a limbic pathway it is impossible for me to experience emotion. When you say what you say I sense a change in your body temperature and breathing, but that is all."⁵⁹

Crucially, therefore, the discussion of the possibility of robotic emotions revolves around embodiment. If, so far anyway, human beings are embodied creatures, robots cannot aspire to attain a similar level of emotional response without the feelings generated by a bodily structure that in turn helps to engender emotions with the help of the brain, in an intricate process that involves a neural net and the limbic system.⁶⁰

Galatea 2.2

To Billie's sessions with Spike, as well as Shira's with Yod, can be added Richard Powers's similar tutorials with another computer, named Helen, in his novel *Galatea 2.2*,⁶¹ where the protagonist's task is to teach Helen about human nature, human feelings and emotions.⁶² A bet wagered around a Turing test lies at the center of *Galatea 2.2*, where the protagonist, the eponymous Richard Powers, a literary scholar and computer programmer, teams up with a computer scientist, Philip Lentz. Their purpose is to train Helen in English Literature so that it/she would be able to pass an MA exam taken simultaneously by a human subject, in what amounts to a literary version of the Turing test. As the narrator explains, Helen "would have to convince an examiner that it performed like a real mind. Operationally equivalent. Indistinguishable [...] A perfect, universal simulation of intelligence would, for all purposes, *be* intelligent."⁶³ (emphasis in the original) Like Shira in *He, She and It* and Billie in *The Stone Gods*, Richard also finds himself falling in love with Helen, a love that appears to be reciprocated. "Appears" is of course a key word here, since the whole project of artificial intelligence crucially revolves about a simulation of thought and, increasingly so, of an attempt to make robots perform as if they had emotions. As Joseph Dewey remarks, like "Galatea, Helen thus becomes a convincing simulation."⁶⁴ In all these cases, moreover, it should be noted that the robots are not the only ones being instructed, since this process works both ways, with the teachers also being affected by the robots' words and ideas. Veronica Hollinger describes this as a "*mutual* evolution,"⁶⁵ which results from human interface and interaction with machines. Sherry Turkle observes that:

A relationship with a computer can influence people's conception of themselves, their jobs, their relationships with other people, and with their ways of thinking about social

processes. It can be the basis for new aesthetic values, new rituals, new philosophy, new cultural forms.⁶⁶

These insights also apply to Billie Crusoe in Winterson's *The Stone Gods*, the scientists in Piercy's *He, She and It* as well as the protagonist of Powers' *Galatea 2.2*, whose lives are significantly impacted by their close interaction with sophisticated robots.

In the first part of *The Stone Gods*, Billie muses: "Robo *sapiens* are not us, but they may become a nearer relative than the ape"⁶⁷ while Spike states that in time, humans will feel kinship with robots, "as the differences between us decrease,"⁶⁸ with robots steadily acquiring organic features as humans move towards an increasing reliance on technology, both inside and outside their bodies. Peter Menzel and Faith D'Aluisio consider that the "next step in human evolution could indeed be from man to machine."⁶⁹ Menzel and D'Aluisio summarize the most important currents in robotics as follows:

Some roboticists believe that machines will never approach human abilities; others, that they will inevitably take over the world. Still a third school argues that these scientists have it all wrong. Robots will neither fall short of people nor overwhelm them. Instead, people will become robots, electronically merging the extraordinary consciousness of *Homo sapiens* and the almost infinitely durable bodies of robots: Robo *sapiens*.⁷⁰

The complex communication between humans and intelligent, sentient machines with a capacity to evolve is an area that has received a lot of attention in recent times. It plays an important role in Winterson's novel, where it is suggested that the interface and relations between humans and machines will continue to develop and will lead to increasingly close and intimate connections, which will affect both people and robots in yet unfathomable ways. Both Piercy's *He, She and It* and Winterson's *The Stone Gods* seem to suggest that vigilance and prudence are indeed crucial as far as developing sentient, humanoid robots are concerned. After much soul-searching Shira decides not to rebuild Yod when she realizes that she would, in a Pygmalion-like gesture, be doing it for herself, to get her lover back, without any regard for Yod's explicit wishes not to be brought back. Indeed, in a related vein, Sven Nyhom and Lily Eva Frank urge caution when considering the possibility of developing sex robots, since the "consequences and technomoral change that will potentially accompany the advancement of robots that can love and be loved is very difficult to predict."⁷¹ On the other hand, they also suggest that since love is "intrinsically valuable"⁷² the development of love robots could potentially be beneficial.

Embodiment and the Gendered Robot

One of the crucial issues around which these critical utopias orbit has to do, as repeatedly intimated, with the question of thought and the development of emotions in the absence of a body. Jean-François Lyotard reflects on the (im)possibility of thought after the death of the sun, which implicitly equates the demise of life on Earth.⁷³ Given this post-solar scenario, and since thought is always already embodied, thinking as we know it would of necessity end, unless cyborgs that could resist extreme conditions were able to survive and continue to think, even the unthought. According to Lyotard the solution would theoretically require the manufacturing of

hardware capable of “nurturing” software at least as complex [...] as the present-day human brain, but in non-terrestrial conditions. That clearly means finding for the “body” envisaged a “nutrient” that owes nothing to biochemical components synthesised on the surface of the earth through the use of solar energy. Or: learning to effect these syntheses in other places than on earth [...] supported only by sources of energy available in the cosmos generally.⁷⁴

This scenario is strongly reminiscent of Spike’s predicament while on Planet Blue, for without solar energy her batteries will run out and she will perish, leaving Billie to fend for herself.

As Lyotard insists, and the two novels under consideration make abundantly clear, it is fundamental to “take the body as model in the manufacture and programming of artificial intelligence if it’s intended that artificial intelligence not be limited to the ability to reason logically.”⁷⁵ Lyotard’s text is orchestrated around two philosophical voices, a male and a female, the latter contending that these machines should be gendered. As the female voice argues, the intelligence being prepared to “survive the solar explosion [...] will have to be nourished not just on radiation but on the irremediable differend of gender,”⁷⁶ a vision given dramatic instantiation both in Piercy and in Winterson’s novels, where the gendered input on the anthropoid robots is strongly emphasized.⁷⁷

Not of woman born

Another question inextricably connected with gender has to do with the way humans and androids or cyborgs are created. How can humanoid robots be distinguished from humans when they are built with organic material to look exactly like the latter? When discussing the main divergences between humanoid robots and human beings, Riva, Shira’s mother, suggests as a possible criterion having been “born from a woman.”⁷⁸ That distinction, however, no longer applies in *Orbus*, where pregnancy is a thing of the past

and even in Tikva and at Y-S half the children “are born from petri dishes or test tubes.”⁷⁹ In a similar discussion, one of the characters notes that Spike was built in a factory, like a car, to which the Robo *sapiens* retorts:

Every human being in the Central Power has been enhanced, genetically modified and DNA-screened. Some have been cloned. Most were born outside the womb. A human being is not what a human being was even a hundred years ago. So what is a human being?⁸⁰

Significantly, in *Blade Runner 2049* (2017), the rift between humans and replicants is described as revolving primarily around the question of who is born (from woman) and who is made. When it is discovered that a replicant woman, Rachel, had apparently become pregnant and given birth, that possibility signalled an unthinkable paradigm shift, an unstoppable blurring of the dividing line between humans and sentient robots. The Replicants are described as violently determined to take over and become the dominant species.

This scenario, where being born of woman is described as the distinctive feature between humans and replicants—with replicants being not born but made—is given a prominent place in the film’s narrative. Ironically and paradoxically, this distinction is made at a time when the implementation of artificial wombs in the not-so-distant future might make the criteria of being born (from a woman) to a great extent irrelevant.

In critical dystopian mode, then, these texts interpellate the ethical, political and personal ramifications of a future world where robot *sapiens* interfaces with *homo sapiens*, dramatized in these revisionary versions of the Pygmalion legend, and retold from a feminist perspective. This feminist angle is in turn embedded in a critical dystopian frame that intimates the need for new revised scenarios in terms of the upcoming interface of humans and cyborgs, creatures that will doubtlessly come together in productive fusions in the near future.

Conclusion

Sherry Turkle argues,

There is every indication that the future of computational technology will include relational artifacts that have feelings, life cycles, moods, that reminisce, and have a sense of humor—that say they love us, and expect us to love them back.⁸¹

Turkle continues by pointing out:

[The] unstated question lies behind much of our current preoccupation with the future of technology. The question is not what will technology be like in the future, but rather, what will we be like, what are we becoming as we forge increasingly intimate relationships with our machines?⁸²

while Rosalind Picard muses: “[A]re people ready for affective computers?”⁸³

In this essay I have concentrated on the educational process of these contemporary versions of Galatea, which crucially revolves around the implanting of memories and emotions in their neuronal circuits, in order to shape them in ways that will make them more similar to humans, a scenario also meaningfully instantiated in Ridley Scott’s *Blade Runner* and effectively dramatized yet again and further problematized in Denis Villeneuve’s *Blade Runner 2049*.

In all of these recent versions of the Pygmalion tale both creators and creatures crucially influence each other, even though the gap between humans and humanoid robots is still extensive, embodiment standing as the most difficult hurdle to overcome. Pygmalion’s dream of bringing his statue to life, so that it would become a real woman, is still very far from being fulfilled, as is Frankenstein’s desire of revivifying an inanimate creature, although large strides have been achieved in the development of biorobots.

David Levy believes that by the year 2050,

[The] boundary between our perceptions of robots and our perceptions of our fellow humans has become so blurred that most of us treat robots as though they are mental, social, and moral beings.⁸⁴

For Levy, “their capacity for serving as our companions, our lovers, and our life partners will in many ways be superior to those of mere mortals”⁸⁵ and he muses about the impact on society “when robots reach a level of sophistication at which they are able to engender and sustain feelings of romantic love in their humans.”⁸⁶ In his view, these developments will have enormous social and psychological benefits. Percy’s and Winterson’s novels dramatize potential versions of this scenario, highlighting some of the advantages and pitfalls of creating robots that are “emergent: that can [...] evolve”⁸⁷ Indeed, manufacturing humanoid robots with whom one could fall in love is effectively another version of the Pygmalion myth, and scarcely a strictly contemporary desire at all.

These texts thus open up new spaces for developing innovative relational dynamics not confined to androcentric models in their revisions of the Pygmalion myth, a foundational story of not simply myths of creation, but of male myths of creation. This is precisely the space occupied by Shira in

Piercy's *He, She and It*, as well as Billie in Winterson's *The Stone Gods*, both characters bent on challenging key aspects of their societies, in particular traditional gender codes, which leads them to effectively become female Pygmalions.

Notes

¹ For an overview of the early creation of mechanical creatures see Kevin LaGrandeur, *Androids and Intelligent Networks in Early Modern Literature and Culture: Artificial Slaves*, Abingdon, Oxon and New York: Routledge, 2013.

² Margaret Atwood, *Oryx and Crake*. London: Bloomsbury, 2003; Atwood, Margaret, *The Year of the Flood*. London: Bloomsbury, 2009; Atwood, Margaret, *MaddAddam*. London: Bloomsbury, 2013.

³ See for example Maria Aline Ferreira, *I Am the Other: Literary Negotiations of Human Cloning*. Westport, Connecticut and London: Greenwood Press, 2005, and Larbalestier, Justine. *The Battle of the Sexes in Science Fiction*. Middletown, Connecticut: Wesleyan University Press, 2002.

⁴ Jean-Jacques Rousseau, "Pygmalion, Scène lyrique," in *Oeuvres complètes*, edited by Ed. Bernard Gagnebin and Marcel Raymond, Vol. II (Bibliothèque de la Pléiade, 1961), 1224-31.

⁵ George Bernard Shaw, *Pygmalion: a Romance in Five Acts*, int. by Nicholas Grene. Harmondsworth: Penguin Classics, 2003.

⁶ The idea of a sex doll or sex robot can be traced back to the Pygmalion myth; see John Danaher, "Should We Be Thinking about Robot Sex?" in *Robot Sex: Social and Ethical Implications*, edited by John Danaher and Neil McArthur (MIT Press, 2017), p. 3.

⁷ Victor I. Stoichita, *The Pygmalion Effect: From Ovid to Hitchcock*, trans. by Alison Anderson. (Chicago and London: The University of Chicago Press, 2008), 6.

⁸ In art, Paul Delvaux's *Pygmalion* (1938) is a salient exception to this scenario, featuring a naked woman embracing a male statue presumably of her own creation. The statue, however, remains unresponsive, unlike most artistic representations, which depict Galatea already infused with animation.

⁹ Marge Piercy, *He, She and It*. New York: Fawcett Crest, 1991.

¹⁰ Jeanette Winterson, *The Stone Gods*. London: Hamish Hamilton, 2007.

¹¹ Karel Čapek, *R.U.R. (Rossum's Universal Robots)*. Harmondsworth: Penguin Classics, 1999.

¹² Isaac Asimov, *I, Robot*. New York: HarperVoyager, 2013.

¹³ Isaac Asimov, *The Bicentennial Man and Other Stories*. New York: Doubleday, 1976.

¹⁴ See Rafael A. Calvo, Sidney D'Mello, Jonathan Gratch, and Arvid Kappas (eds), *The Oxford Handbook of Affective Computing*. Oxford: Oxford University Press, 2015.

¹⁵ Piercy, *He, She and It*, 142.

¹⁶ *Ibid.*, 142. As is observed by Deery, June. "The Biopolitics of Cyberspace: Piercy Hacks Gibson," in *Future Females, The Next Generation: New Voices and Velocities in Feminist Science Fiction Criticism*, edited by Marleen S. Barr (Lanham: Rowman & Littlefield), 2000, p. 93, Piercy is "struck by the irony of the macho male—supposedly the superior human form—aspiring to be inhuman, and the cyborg trying to be human." Indeed, this aspiration to be human is a recurring trope in many narratives dealing with cyborgs, robots and a host of other creatures who yearn to be fully human, ranging from Collodi, Carlo. *Pinocchio*, int. by Umberto Eco, trans. by Geoffrey

Brock. New York: New York Review Books Classics, 2009, Baum, L. Frank. *The Patchwork Girl of Oz*. New York: HarperCollins, 1995, the replicants in Ridley Scott's film *Blade Runner*, Denis Villeneuve's *Blade Runner 2049* and David in Spielberg's *AI*, to cite only a few salient examples, all of which partake of this longing.

¹⁷ Piercy, *He, She and It*, 114. As Arlindo Oliveira explains, "machine learning algorithms enable computers to learn from experience." See *Digital Minds: How Science is Redefining Humanity* (Cambridge, Massachusetts: MIT Press, 2017), 7.

¹⁸ For a discussion of an earlier version of this scenario see Catherine Maxwell, "Browning's Pygmalion and the Revenge of Galatea." *ELH* 60:4 (Winter 1993), 989-1013.

¹⁹ Nicole Ward Jouve, *Female Genesis: Creativity, Self and Gender*. (Cambridge: Polity Press, 1998), 2.

²⁰ Piercy, *He, She and It*, 150.

²¹ Piercy, *He, She and It*, 150. She has clearly acknowledged the influence that Haraway, Donna. "A Cyborg Manifesto: Science, Technology and Socialist-Feminism in the Late Twentieth Century." *Simians, Cyborgs, and Women: The Reinvention of Nature*. (London: Free Association Books, 1991), 431.

²² Piercy, *He, She and It*, 150. Both Piercy's *He, She and It* (1991) and Steven Spielberg's *A.I. Artificial Intelligence* (2001) share a number of thematic concerns as revisions of the Frankenstein story and as specifically contemporary parables of creation. Both the novel and the film take place in a future, "posthuman" world which has been partially destroyed by ecological calamities and both examine the fate and role of fabricated beings endowed with artificial intelligence. Both Yod, the cyborg in Piercy's novel, and David in Spielberg's film express a profound yearning to be human and to be loved. In turn, both Monica in relation to David and Shira in relation to Yod may be said to almost take for granted their immutable, never-changing capacity for love.

²³ Piercy, *He, She and It*, 174.

²⁴ *Ibid.*, 412.

²⁵ *Ibid.*, 29.

²⁶ *Ibid.*, 418.

²⁷ *Ibid.*, 97.

²⁸ *Ibid.*, 120.

²⁹ *Ibid.*, 151.

³⁰ For a discussion of Yod's place in Tikva society in terms of his personhood and the work he performs see Heather Hicks, "Striking Cyborgs: Reworking the 'Human' in Marge Piercy's *He, She, and It*," in *Reload: Rethinking Women+Cyberculture*, edited by Mary Flanagan and Austin Booth (Cambridge, Massachusetts: MIT Press, 2002), 85-106.

³¹ Piercy, *He, She and It*, 375.

³² Winterson, *The Stone Gods*, 7.

³³ Aldous Huxley, *Brave New World*, int. by Margaret Atwood. London: Vintage Classics, 2007.

³⁴ Winterson, *The Stone Gods*, 27.

³⁵ Sherry Turkle, "Whither Psychoanalysis in a Computer Culture?". <http://www.kurzweilai.net/whither-psychoanalysis-in-a-computer-culture> (accessed on November 24th, 2017). As Parisi, Domenico and Giancarlo Petrosino explain, "Emotions exist because they help the motivation decision mechanism to function better and therefore they increase the survival and reproductive chances of animals."

This is the adaptive function of emotions and this explains why they have emerged evolutionarily.” From “Robots that *have* emotions,” *Adaptive Behavior* 18:6 (2010): 466.

³⁶ Ibid.

³⁷ Parisi and Petrosino, “Robots,” 453.

³⁸ Winterson, *The Stone Gods*, 62.

³⁹ Ibid., 62.

⁴⁰As it is put by Rodney A. Brooks, emotions are “our current last bastion of specialness.” *Flesh and Machines: How Robots Will Change Us* (New York: Vintage Books, 2003), 171.

⁴¹ Damásio, António, *Descartes’ Error: Emotion, Reason and the Human Brain*. London: Vintage, 2006, xiv.

⁴² All *ibid.*, xviii.

⁴³ Ibid., 165.

⁴⁴ Ibid., 200.

⁴⁵ Ibid., 245.

⁴⁶ Picard, Rosalind, *Affective Computing*. Cambridge, Massachusetts: MIT Press, 1997.

⁴⁷ Picard, Rosalind, “Does HAL Cry Digital Tears? Emotions and Computers” in *HAL’s Legacy: 2001’s Computer as Dream and Reality*, edited by David G. Stork (Cambridge, Massachusetts: MIT Press, 1996), 296.

⁴⁸ Ibid., 301.

⁴⁹ Ibid., 302.

⁵⁰ Ibid., 302.

⁵¹ Piercy, *He, She, and It*, 97.

⁵² David Levy, *Love and Sex with Robots: The Evolution of Human-Robot Relationships* (New York: Harper, 2007), 120.

⁵³ Winterson, *The Stone Gods*, 177.

⁵⁴ Ibid., 132.

⁵⁵ Ibid., 132.

⁵⁶ Ibid., 166.

⁵⁷ All *ibid.*, 145.

⁵⁸ Ibid., 132.

⁵⁹ Ibid., 141.

⁶⁰ As C. George Boeree explains, although emotion “involves the entire nervous system [...] there are two parts [...] that are especially significant: The limbic system and the autonomic nervous system. The limbic system is a complex set of structures that lies on both sides and underneath the thalamus, just under the cerebrum. It includes the hypothalamus, the hippocampus, the amygdala, and several other nearby areas. It appears to be primarily responsible for our emotional life, and has a lot to do with the formation of memories.” See “The Emotional Nervous System”. <http://webspaceship.edu/cgboer/limbicsystem.html> (accessed on November 23rd, 2017).

⁶¹ Richard Powers, *Galatea 2.2*. London: Atlantic Books, 2010.

⁶² Another example of a Pygmalion tale of human-robot love is Rey, Lester del. “Helen O’Loy.” *The Science Fiction Hall of Fame*. Volume One, edited by Robert Silverberg (New York: Orb Books, 2005), 42-52. A robotic woman, significantly also called Helen, created to represent a male idealized fantasy of a domestic goddess, an earlier example of a Stepford Wife, falls in love with her creator. They get married and, in fairy-tale like

atmosphere, live happily until his death, at which time Helen deactivates herself. She thus embodies, in her performance of perfect patriarchal womanhood, the male dream of woman as sex slave and subdued, devoted wife. For an analysis of this story see Hollinger, Veronica. “‘Something Like a Fiction’: Speculative Intersections of Sexuality and Technology,” in *Queer universes: sexualities in science fiction*, edited by Wendy Gay Pearson, Veronica Hollinger and Joan Gordon (Liverpool University Press, 2008), 140-161. Stoichita, *The Pygmalion Effect*, discusses the significance of Helen of Troy in terms of the Pygmalion Effect.

⁶³ Powers, *Galatea*, 52; cf. Levy, *Love and Sex*, 120.

⁶⁴ Joseph Dewey, *Understanding Richard Powers* (Columbia, South Carolina: The University of South Carolina Press, 2002) 101.

⁶⁵ Veronica Hollinger, “Cybernetic Deconstructions: Cyberpunk and Postmodernism,” *Mosaic* 23 (Spring 1990), 42.

⁶⁶ Sherry Turkle, *The Second Self: Computers and the Human Spirit*. Cambridge, Massachusetts: MIT Press, 2005. Quoted in Levy, *Love and Sex*, 64.

⁶⁷ Winterson, *The Stone Gods*, 28.

⁶⁸ *Ibid.*, 29.

⁶⁹ Peter Menzel and Faith D’Aluisio. *Robo sapiens: Evolution of a New Species* (Cambridge, Massachusetts: MIT Press, 2000), 17.

⁷⁰ *Ibid.*, 21.

⁷¹ Sven Nyhom and Lily Eva Frank, “From Sex Robots to Love robots: Is Mutual Love with a Robot Possible?” in *Robot Sex: Social and Ethical Implications*, edited by John Danaher and Neil McArthur (MIT Press, 2017), 237.

⁷² *Ibid.*, 237.

⁷³ Jean-François Lyotard, “Can Thought Go On Without a Body?” in *Posthumanism*, edited by Neil Badmington (Houndmills: Basingstoke: Palgrave, 2000), 129-40.

⁷⁴ *Ibid.*, 133.

⁷⁵ *Ibid.*, 135.

⁷⁶ *Ibid.*, 139.

⁷⁷ For an analysis of how posthumanism always already entails a re-inscription of humanism in Piercy’s novel see Neil Badmington, “Posthumanist (Com)Promises: Diffracting Donna Haraway’s Cyborg Through Marge Piercy’s *Body of Glass*” in *Posthumanism*, edited by Neil Badmington. (Houndmills: Basingstoke: Palgrave, 2000), 85-97.

⁷⁸ Piercy, *He, She and It*, 114. As is cunningly observed by June Deery, Yod “might be truly sympathetic to women because he is also free of the ultimate male resentment for being born of woman.” See Deery, “The Biopolitics of Cyberspace: Piercy Hacks Gibson” in *Future Females, The Next Generation: New Voices and Velocities in Feminist Science Fiction Criticism*, edited by Marleen S. Barr. (Lanham: Rowman & Littlefield, 2000) 95,

⁷⁹ *Ibid.*, 191.

⁸⁰ *Ibid.*, 63-4. For a discussion of the artificial birth of robotic, android creatures see Despina Kakoudaki, *Anatomy of a Robot: Literature, Cinema, and the Cultural Work of Artificial People*. New Brunswick, New Jersey: Rutgers University Press, 2014.

⁸¹ Turkle, “Whither,” <http://www.kurzweilai.net/whither-psychoanalysis-in-a-computer-culture> (accessed November 24th, 2017).

⁸² *Ibid.*

⁸³ Picard, "Does HAL Cry," 302.

⁸⁴ Levy, *Love and Sex*, 303.

⁸⁵ *Ibid.*, 303.

⁸⁶ *Ibid.*, 304.

⁸⁷ LaGrandeur, *Androids*, 105.

The Postmodern Prometheus: Humanity and Narration in the SF Worlds of Dick's *Do Androids Dream of Electric Sheep?* and Scott's *Blade Runner*

Andrew Howard

University of the District of Columbia Community College

The effect of outside stimuli on literature is perhaps best shown in the realm of science fiction, and most obvious in that of cold war SF. After Pat Frank's *Alas Babylon*, the conceit of future post-nuclear worlds in which a ravaged society is rebuilding itself while adjusting to effects of rapidly advanced technology became an actively used backdrop for the multiple variants of SF literature. The idea of a society nearly destroyed by its own devices was a very real thing for authors and readers in the Red-Scare, duck-and-cover fifties and sixties. Yet aside from the basic post-apocalyptic aspect of the near-future or not-so-distant future story, a post-nuclear background allows for an exploration of humanity through a lens that, while removed from contemporary society, is not so far from the current reality of the author's time period as to suggest a landscape so alien that readers would not be able to recognize their own world within the text.

Perhaps what is most useful about the post-nuclear holocaust as a trope is the fact that, while man is seen as able to prevail—in the face of adversity, against nature, in a harsh environment, and so forth—man is also the catalyst for the horrific nuclear event. Even the most dystopian post-nuclear novel allows a critique of mankind with a glimmer of hope, since generically any story requires a protagonist, and human readers tend to prefer their heroes to be human. This dichotomy of man as force of good as well as a force of evil is amplified in SF texts that concern what are essentially robots—androids, cyborgs, basically the humanoid simulacra—that act as mirrors for our own humanity. In a good many texts involving humanoid simulacra, the android is eventually hunted down, ostracized, killed, or *retired*. Often, the creation of androids will cause the rise of a Prometheus trope—such as that *Modern Prometheus* Frankenstein—and the reader can expect that to some extent man will be punished for his hubris. Fear of this punishment, as well as fear of the creation that may be in some ways superior to the creator, can cause the persecution of the simulacra. Joseph Francavilla addresses the issue by exploring, as the title of his essay suggests, “The Android as *Doppelgänger*.”

[It] would seem that our *guilt* about the abuses of our scientific creations and our employment of technology, when it returns from repression, is collectively projected on the android or robot double appearing as censoring conscience, the double becoming a symbol of our guilt and fear concerning technology being misused, becoming uncontrolled, going awry. It is no wonder, then, that we are afraid of these shadow figures in technological guise...We fear these perfect artificial life forms will replace us perfectly, reverse our roles as master and slave, take us over *en masse*, destroy us all.¹

However, the android is not only a symbol of fear and suppression, guilt and persecution. The android can represent the pinnacle of modern achievement—that Promethean level of creation. The android narrative can explore “a space of desire, a place within the self where we can experience a kind of otherness,” and offers a chance, as J.P. Telotte posits, to “speculate about that ‘web of innumerable possibilities’ Foucault describes—possibilities that ‘exceed the experience’ of our normal being.”²

As an author, Philip K. Dick consciously used the android trope to build a metaphor for the way humans interact with and treat each other. Simultaneously an exploration into the ability to “exceed the [human] experience” and a threat to humanity (more so metaphysically and spiritually than physically), the android is presented as a problem with no clear solution. Once, thanks to the rapid technological development offered by SF, the android has evolved to a level in which it is nearly undetectable among humans, any moral or philosophical ruminations upon the android are rendered completely gray and ambiguous. Dick explores not so much the technical achievement of the advanced android, but the relationship between the android and the human, the original and the simulacra. As the author puts it himself in his essay “The Android and the Human,” “as the external world becomes more animate”—that is, infused with technology—“we—the so-called humans—are becoming, and may to a great extent have always been, inanimate in the sense that *we* are led, directed by built-in tropisms, rather than leading.” Dick’s solution to this blurring of the definition of animation? “So we and our elaborately evolving computers may meet each other halfway.” This compromise proves to be a wonderful ground for mining for stories, as Dick anecdotally illustrates a topic that comes up in *Do Androids Dream of Electric Sheep?* and even more so in its filmic interpretation, *Blade Runner*: “Someday a human being, named perhaps Fred White, may shoot a robot...[which] may shoot back and, to its surprise, see a wisp of gray smoke arise from the electric pump that it supposed was Mr. White’s beating heart.” The revelation of this act, an example of the blurring between original and

simulacra, is “rather a great moment of truth for both of them.”³ Dick’s *Do Androids Dream of Electric Sheep?* and its film adaptation *Blade Runner* explore these moments of truth and ambiguity between the original and the simulacra, and reveal the distances, great and small, between creator and creation.

***Androids* and SF Narrative Realism**

Do Androids Dream of Electric Sheep? offers a possible world, after the nuclear holocaust of World War Terminus, in which the remaining inhabitants of Earth are faced with a growing moral dilemma surrounding the ideas of personhood. The semi-organic androids are indistinguishable from *authentic* humans. The introduction of new android models such as the Nexus-6 with their increasingly human-like personalities echoes the graying line between creature and creator. Suffering the most due to this ambiguity is Rick Deckard, a bounty hunter whose job it is to kill androids that have escaped to Earth. The novel displays a sublime use of the near-future, post-nuclear holocaust trope to convey convincingly both the simultaneously alien and familiar feel of the simulacra and the moral ambiguity faced by those in the fairly new profession of retiring them.

Setting the novel in San Francisco brings the impact of the story and its moral dilemmas closer to the reader by mirroring a world to which readers can relate. Dick is clearly interested in making his fantastic SF narrative as familiar to readers as possible. Granted, there are hovercars, empathy boxes, and well, androids and electric sheep, yet the world of *Androids* is as covered with dust and refuse as any contemporary landfill. Characters are not defined by their science fiction surroundings, but rather by the ways in which they interact with them. Perhaps most exemplary of Dick’s attempt to bridge the gap between future world and contemporary society is the fact that, while the novel’s androids are fugitives from Mars who performed a daring escape from the planet, there are no scenes in the novel that actually take place in space, the traditional science fiction backdrop.

The physical focus of the novel, then, is clearly within the realm of Earth, specifically the American west coast. More specifically, *Androids* focuses on the moral implications of the ability to create and destroy life with equal casualness in a world quite similar to our own. Imbedded within the text is an inherent hierarchy in which living matter is juxtaposed with organic biomatter that is artificially created by man. In essence, the book has a certain level of a man versus android motif, yet the philosophical problems of the novel tend to eschew this altercation for an examination of the moral dilemma of sanctioned, institutionally justified killing.

Turing and Voight-Kampff—Empathy and Narration

Do Androids Dream of Electric Sheep? explores the way in which humans and their constructs interact on a moral level. By delving into the minds of his

characters, Philip K. Dick's use of narrative presents glimpses into the workings of one of the novel's central moral dilemmas—how supposedly empathic humans can justify killing what is essentially human. There are no narrow, easy criteria that define authentic humans in the novel, aside from a test that measures levels of empathy in response to certain stimuli. However, the test, if it does in fact work, only separates supposedly normal humans from androids and humans with certain kinds of psychopathic characteristics that exclude empathy. The test was inspired by the 1950 Turing test, a thought experiment in which the examiner asks a series of questions to identify which of two subjects is human, and which is a computer. The Turing test relies on a different criterion than that of the Voight-Kampff empathy test used in *Androids*. Rather than determining humanity based on empathy, the Turing test aimed to determine “whether [a] machine is capable of making a human being believe that it thinks as he does.”⁴ The test would then seem to determine humanity through the lens of Descartes' famous “I think, therefore I am [human]” maxim (which Pris quotes in *Blade Runner*), whereas Dick's version would seem to better reflect, “we feel for each other, therefore we are (human).” However, the possession or absence of empathy doesn't necessarily designate humanity in Dick's world, which suggests that the Voight-Kampff test echoes the distance between humans—I cannot trust you are what you say you are without the empirical evidence of this test.

Emmanuel Carrere's biography of Philip K. Dick tells of the Turing test's influence over the author. What appealed to Dick was the fact that Turing had given so much thought to “thinking machines, which were one of Phil's obsessions.” There is much similarity in how Turing aligns thought with consciousness and Dick, in *Androids*, aligns the ability to feel empathy with humanity, or personhood. As Carrere summarizes Turing's points, echoes of *Androids* appear:

The phenomenon of consciousness can only be observed from the inside. I know that I have a consciousness, and indeed it is because of it that I know this, but as to whether you have one or not, nothing can prove to me that you do. What I can say, however, is that you emit signals, gestural and verbal for the most part, from which, by analogy with those I emit, I can deduce that you think and feel just as I do. Sooner or later, Turing argues, it will be possible to program a machine to respond to all stimuli with signals as convincing as those emitted by a human being. By what rights, then, can we reject its bona fides as a thinker?⁵

Of course, there is one exception to this rule. One agent that can claim to be able to mine the consciousness of another, albeit a fictional other, is the narrator. Devised by an author, the narrator becomes a Prometheus figure in

his or her own right, creating an entire world and providing its inhabitants with fire. The narrator provides the reader (the narratee) with what is deemed important, on a narrative level, to be revealed. Through various methods—first person narrative, omniscient narration, free indirect discourse, dialogue, action—the author can express to the reader as much or as little as he or she wants in regard to a character’s conscious being. An omniscient narrator can give the reader an impression of any given character’s thoughts, from a general idea to an exact transcript, though the kind of exact transcript that can express a character’s thoughts is usually reserved for main characters, whose thoughts may be considered more integral to the movement of the story. Another method of expressing the thoughts of a character is a little more ambiguous—the use of free-indirect discourse. By narrating through a lens of a character’s particular mindset, mood, or personality in general, a narrator presents a way to experience a character’s consciousness.

It is through action and emotion that Dick’s narrator addresses the qualities that make his subjects human. The novel’s great irony is that once the lack of humanity is determined in a test subject, the subject is quickly terminated in a cold, inhumane manner. In talking to Paul Sammon, author of *Future Noir: The Making of Blade Runner*, Dick relates a story about reading the diaries of Nazi soldiers: “There was obviously something wrong with the man who wrote that. I later realized that, with the Nazis, what we were essentially dealing with was a defective group mind, a mind so emotionally defective that the word ‘human’ could not be applied to them.” This impetus concerning emotion would seem a clear-cut design for an SF story; the cold, emotionless androids pose a threat to the humans who rightfully protect themselves from them. However, Dick continues, the problem of emotion surfaces when the humans have to be the cold, emotionless killers: “The problem in this killing then would be, ‘Could we not become like the androids, in our very effort to wipe them out?’”⁶

Justification through Character Action

Deckard does not, by the semantic standards of his world, kill androids. It is his job to retire them. This difference in wording is crucial, yet the act is no different. As Marilyn Gwaltney, who makes a case for androids as being morally the same as humans, points out after using the word *killed* to describe Deckard, “Killing is not the same thing as murder, so there is really no need for the euphemism ‘retire’ in the book except as a literary device to point up the moral dilemma of treating conscious beings as equipment that can be ‘retired’ from service when it is worn out or obsolete.”⁷ Though Gwaltney is correct in identifying the use of the word retire, what she misses is the fact that androids existing on Earth are fugitive (they do not become “worn out or obsolete” because they naturally expire after four years of existence), further enforcing the mental distance between android and human. However,

Deckard's wife Iran points out the hypocrisy of his position in the novel's opening pages:

"I'm not a cop," [he said]...
"You're worse," his wife said, her eyes still shut. "You're a murderer hired by the cops."⁸

Deckard's initial reaction is one of justification: "I've never killed a human being in my life." It's obvious that Deckard is employing a semantic game to detach himself from the actualities of his profession. Androids have no need for such games however, as Pris describes bounty hunters to Isidore as "a professional murderer who's given a list of those he's supposed to kill."⁹ The question that is introduced through these different perspectives on the same action takes on some importance when aligned with the question of empathy: does empathy define humanity? Are non-humans immune from empathy?

The question becomes more explicit throughout the book, as Dick teases it out through his characters' speech and actions. As Deckard attempts to verify Luba Luft's inauthenticity, so to speak, he explains the school of thought behind the Voight-Kampff empathy test. He tells her "An android...doesn't care what happens to another android. That's one of the indications we look for." Luft replies that Deckard "must be an android."¹⁰ It is a perfectly natural response, especially since it comes from an android facing death. However, the moral implications are astounding. If Deckard doesn't feel empathy toward biological humanoids that are fundamentally similar to traditional humans on most levels, is he any more in touch with humanity than the androids? Is he truly capable of telling a supposedly cold, calculating android from a human?

Retiring is further justified by Deckard's alliance with the tenets of Mercerism—Earth's empathy-centered religion/philosophy. It is not acceptable to kill needlessly in Mercerism; in fact, the purpose of killing is to prevent more killing. Deckard qualifies his job as a Mercerist necessity. He must think of the androids as a threat in order to be able to kill them. Only once the prey is thought of as predator is it morally acceptable to kill. Thinking this way

made his job palatable. In retiring—i.e. killing—an andy, he did not violate the rule of life laid down by Mercer. *You shall kill only the killers*, Mercer had told them the year empathy boxes first appeared on Earth. And in Mercerism, as it evolved into a full theology, the concept of The Killers had grown insidiously. In Mercerism...it was never clear who or what this evil presence was. A Mercerite *sensed* evil without understanding it. Put another way, a Mercerite was free to locate the nebulous presence of The Killers wherever he saw

fit. For Rick Deckard an escaped humanoid robot, which had killed its master, which had been equipped with an intelligence greater than that of many human beings, which had no regard for animals, which possessed no ability to feel empathic joy for another life form's success or grief at his defeat—that, for him, epitomized the Killers.¹¹

Deckard can then use this rule of life to kill, and kill without compunction. If one can identify any group as *the Killers* than one can basically have no problem killing within that group.

Justification through Narrative

Dick's narrator, throughout *Androids*, uses the thoughts of the novel's characters to reflect their diegetic surroundings. The most striking way in which this occurs is through free indirect discourse. The occurrences of free indirect discourse that occur throughout the book reveal an increasing commentary on *Android's* moral problems. At first, the reader gets glimpses of Rick Deckard's life through his thoughts, woven into the narrative thread: "Very true, Rick thought as he opened the gate to his little pasture and approached his electric sheep. But I can't emigrate, he said to himself. Because of my job."¹² At this point, the narrative structure has opened to include Deckard's thoughts, and they are presented nearly as dialogue. It is important to note, however, that those thoughts are not marked in a manner similar to dialogue, such as quotation marks or italics. By not designating Deckard's thoughts as separate from the narrative discourse, Dick has elevated them to the importance of narration—Deckard's thoughts on his job are as essential to the reader as his actions. As the story progresses and develops, these narrated thoughts are increasingly indispensable to the philosophical heart of the novel.

When Deckard suspects fellow bounty hunter Phil Resch of being an android, his thoughts are expressed in free indirect discourse, giving away some of his growing reservations over his chosen field of work. If Resch is an android, bounty hunter Deckard should not feel any real empathy toward him. However, his unspoken thoughts betray some rising guilt.

I've got to tell him, he said to himself. It's unethical and cruel not to. Mr. Resch, you're an android, he thought to himself. You got me out of this place and here's your reward; you're everything we jointly abominate. The essence of what we're committed to destroy.¹³

This line of thinking is revelatory, and by placing it on the level of narrative discourse, its importance will not be lost on the attentive narratee. Deckard's gradual loss of faith in his emotional distance from androids is essential to *Androids*, and it seems that had Dick not dispensed with stock devices (such as

quotation marks or italics, or through diegetic summary or indirect content paraphrase) to introduce that element, the effect on readers would have been weakened or lost.

Free indirect discourse is a technique not limited to *Androids*. A testament to Dick's ability to express his characters' thoughts through narration, the use of free indirect discourse appears throughout Dick's narration shows recurrent tropes in Dick's other novels. From the first few pages of *A Scanner Darkly*:

Most of all he felt sorry for his dog, because he could see the bugs landing on and settling all over him, and probably getting into the dog's lungs, as they were in his own. Probably—at least so his empathic ability told him—the dog was suffering as much as he was. Should he give the dog away for the dog's own comfort? No, he decided: the dog was now, inadvertently, infected, and would carry the bugs with him everywhere.¹⁴

This passage illustrates the narrative potential of free indirect discourse that Dick uses so effectively. Here we get factual occurrence—the man feels sorry for the dog—as well as the implicit perspective from the character's point of view—there are imaginary bugs infecting him and his dog. The narrator goes as far as to question what the character should do: “Should he give the dog away...” which implicitly says, “the man wondered if he should have given the dog away...” which illustrates the bending of chronological narrative involved in free indirect discourse. The narration of the text works in much the same way the human (and in this case, drug-addled) mind works, that is, relatively free of the constraints of tense and chronology. There is no explicit mention, at this point, that the bugs are hallucinatory, therefore, the narratee is given the illusion of the bugs as fact. The character's thoughts are then displayed as real—the *truth* in the fictional narrative.

However, Dick's narrator does not leave all pertinent opinion to the thoughts of his characters. When androids act in a manner that can be characterized as machine-like or ersatz, they can be presented as such through other forms of narration. Through this other mode of discourse a kind of official stance on androids can be construed, which further complicates the philosophical role of androids within the narrative structure of the novel. When dealing with humans, Dick is often sympathetic as a narrator, even though his characters are far from perfect. The narration even goes as far as simply telling the reader a character's moods or thoughts, such as when Deckard receives his assignment: “He felt depressed. And yet, logically, because of Dave's sudden disappearance from the work scene, he should be at least guardedly pleased.”¹⁵ Android emotions are never spelled out that way; they reveal themselves through their speech and actions. For example, when the android Irmgard is fed up with Roy Baty and the androids' situation:

"We don't know that," Irmgard said. "That's only a conjecture. I think they, they — " She gestured. "Walked around. Sang from a stage like Luba. We trust — I'll tell you what we trust that fouls us up, Roy; it's our goddamn superior intelligence!" She glared at her husband, her small, high breasts rising and falling rapidly. "We're so smart — Roy, you're doing it right now; goddamn you, you're doing it now!"¹⁶

Here, Irmgard is quite clearly exasperated—her chest heaving, the way she repeats herself before she finishes a thought, her use of the word *goddamn*—yet the narrator chooses only direct discourse to express her exasperation.

The narration does comment on the androids occasionally, however. Yet this commentary does not originate from a clear point of view, as when J.R. Isidore is involved. His unspoken thoughts are often expressed through free indirect discourse, perhaps more than Deckard's, such as, "So why listen to that? He asked himself irritably. Fork them and their colonization, I hope a war gets started there— after all, it theoretically could — and they wind up like Earth. And everybody who emigrated turns out to be special."¹⁷ The narrator certainly displays a handle on characters' individual neuroses and outlooks. When Isidore first meets Pris, knocking on her door, "the television died immediately into nonbeing. It had not merely become silent; it had stopped existing, scared into its grave by his knock. He sensed, behind the closed door, the presence of life, beyond that of the TV." Isidore's apprehension, as well as Pris' fear is palpable as the narration continues: "His straining faculties manufactured or else picked up a haunted, tongueless fear, by someone retreating from him, someone blown back to the farthest wall of the apartment in an attempt to evade him."¹⁸ This is in contrast to the free indirect discourse offered through Isidore's point of view earlier in the book. As the actions surrounding Isidore become more complex, the narrator uses more indirect content paraphrasing to translate Isidore's thoughts to the reader. It is in passages such as this that the narrator's voice becomes more apparent:

Now that her initial fear had diminished, something else had begun to emerge from her. Something more strange. And, he thought, deplorable. A coldness. Like, he thought, a breath from the vacuum between inhabited worlds, in fact from nowhere: it was not what she did or said but what she did not do and say.¹⁹

Instead of any kind of free indirect discourse, or even indirect content paraphrase, the narrator colors the action with a seemingly biased choice of words: "She gave him one last warmthless glance."²⁰ Dick repeats this motif when describing other androids: "Roy Baty entered, somber and large, smiling

his crooked, tuneless smile.”²¹ The androids in book and film are clearly setup as villains, and the narrative voice serves to provide the expected audience reaction to villains; however, both the novel and film challenge audience expectation of villainry by dulling our hero Deckard to a near-animal state of consciousness, while giving the villainous android more—albeit negative—human traits.

Narrative in the Film Version

No discussion of a novel and its film adaptation is complete without a little compare-and-contrast, and Joshua Foa Dienstag does so succinctly here, focusing on the book’s obsessions of organic authenticity (as seen in Dick’s focus on animals) as opposed to the film’s quiet ruminations on personhood:

The novel has an entirely different side that has been largely excised from the film. All the characters in the book, and the book itself, are preoccupied with animals, both real and artificial. In Dick’s narrative, humans are confronted, as it were, on both sides of the species barrier: on the one hand, with biological androids “more human than human” (as the film puts it) whose presence they fear, and on the other by a dwindling number of live animals, whose loss they mourn, and the animal robots they have built to replace them.²²

To mourn is certainly a trait within the realm of consciousness. The humans of these worlds mourn what they have lost; “they seek to replace [their ravaged natural world] with machine-religions, machine-emotions and machine-animals” while Roy Baty mourns the full life he will never have.²³

As for the inner lives of the characters, explicitly or not, film has trouble relating the thoughts of a character without action or dialogue. However, the use of dialogue to express thought is limited in its ability to maintain integrity of realism and believability. More often than not, a realistically-written character is likely to act upon his or her thoughts rather than explain them for the sake of the audience. The voice-over narrative, a stock device that explicitly and forcefully illustrates a character’s otherwise unspoken thoughts, attempts to bridge the gap between the unspoken thought and ham-fisted emotional delivery. Ridley Scott was pressured to use the device to help the viewer get a better grasp of *Blade Runner*’s diegesis; he thought little enough of it that Deckard’s monologues are excised from the director’s cut. The device, now removed in the director’s cut and final cut, is as close to a kind of cinematic free indirect discourse as *Blade Runner* gets. Ultimately, Scott lets the story tell itself on its own terms, and therefore actions speak louder than voice-overs.

Deckard’s motivation differs from the novel to the film, which emphasizes Deckard’s fatalistic acceptance of his sorry state—he’s a “petty bureaucrat

employed by the San Francisco Police Department (but *not* a policeman)” who “hasn’t the resources to afford a new life.”²⁴ Emotionally colder, the Deckard of *Blade Runner* is a future Philip Marlowe, whose motivations are less financially driven to improve his life. Deckard is retired in the film, and pulled back into the “sleazy underworld” when the “police threaten to lean on him in undefined but convincing ways unless he cooperates with them.”²⁵ It is important to note this juxtaposition, since Deckard is believed to be an android himself in the film version. Take his noticeable lack of human traits. A very human trait is desire, and while Deckard shows some desire by expressing a love for Rachael, he more or less drifts through the film, expressing neither joy nor despair, having no desire to change his life. In fact, the only reason Deckard leaves (or, in the director’s cut, attempts to leave) his world is out of necessity—he and Rachael are now on the run. While Deckard is ultimately resigned to his place in life at the end of *Androids*, he has at least shown a desire to adjust his status. *Blade Runner’s* Deckard is resigned to his position in the world throughout and shows no feeling aside from a quiet confusion, as shown on Harrison Ford’s face. The world of *Androids* is one that he was born into, through the events of (future) history, against his will. It would seem that the Deckard of *Blade Runner* was *made* into his world, prepackaged with a number of failures and memories, uninterested or unable to want to change his place in a dreary dystopia.

When compared to the far more dynamic Roy Baty, who is aware of his impending mortality and therefore fighting for his life, Deckard (whether or not the viewer sees him as an android) comes across as the colder and more robotic character. Baty shows desire (“I want more life, fucker,” he demands of Tyrell in the original version of the film) and genuine regret, as when Pris is killed by Deckard. However, in a film that is fairly low on moral proselytizing and philosophical rumination within its dialogue, Baty is given the last word on his own life and mortality, giving a short and poignant speech that lends a last touch of humanity to the state of androids, therefore calling Deckard’s line of work into final question: “I’ve seen things you people wouldn’t believe. Attack ships on fire off the shoulder of Orion. I watched sea beams glitter in the darkness at Tan Hauser Gate. All those moments will be lost in time like tears in rain. Time to die.”²⁶

The implications of an android mourning his inevitable death are by no means inconsequential. It is a hallmark of classic SF to investigate the blurring of boundaries of consciousness, the “tears in rain.” *Do Androids Dream of Electric Sheep* and its close cousin *Blade Runner* are able to explore these themes without sentimentality or philosophical rumination, though at the heart of both these stories is fear: fear of death, fear of replacement, fear of the unknown. Neuroscientist Max Bennett notes the implication of a constructed, artificial consciousness as related in *Androids* and *Blade Runner*. Noting that the androids are indistinguishable from humans, Bennett poses that “If a situation like this should ever arise, then Turing would argue that consciousness has been created. However...that would not necessarily mean that an understanding of

the origins of consciousness had been reached.” Perhaps part of the mystery of origins of consciousness is somehow essential to the understanding of sentient life, artificial or natural. After all, there would be no story if the androids felt no fight or flight reflex, pushing them into extreme situations in order to survive and extend their lives. Bennett continues:

Nevertheless, in the absence of insights provided by a theory that encompasses both the physical world and consciousness, *Homo sapiens* will almost certainly accept the androids as possessing consciousness like ourselves if they pass all the tests that can be applied to confirm if a species possess this attribute. If this should ever eventuate, we will have been seduced into accepting consciousness as an attribute of certain kinds of synaptic networks, without perhaps ever understanding why that should be so.²⁷

The passage is a testament, as is much of the classic SF canon, to man’s nature of misunderstanding, even regarding that which he has created. Philip K. Dick and Ridley Scott have mined an unfortunate distance between creature and creator. The resulting works leave more questions than answers, yet prove to attribute more to SF literature and the debate of consciousness than perhaps even more so than their creators intended.

Of course, the source of their creators’ intention is a difficult path to follow. For instance, *Blade Runner* is notorious for its many available versions, most recently *Blade Runner: The Final Cut*, which is the most easily acquired version, at least digitally, on the current market. In this version, from an apparent bowdlerization for television broadcast, Baty demands of Tyrell “I want more life, *father*.” Online speculation abounds as to which line best reflects the intent—either of screenwriter Hampton Fancher or director Ridley Scott—of the creators of the film. Baty is certainly portrayed throughout the film as having very human characteristics—intelligence, the ability to manipulate, survival instincts—but here in Scott’s final cut he is most human, fighting for what he needs, using pathos to appeal to *his* creator. He justifies himself and his actions. In this moment he is not, as the slogan goes, more human than human. He is angry, scared, and vulnerable—just human enough for empathy.

Notes

¹ Joseph Francovilla, “The Android as *Doppelgänger*.” *Retrofitting Blade Runner*. Ed. Judith B. Kerman (Madison: University of Wisconsin Press, 1991), 8.

² Telotte, J.P. *Replications: A Robotic History of the Science Fiction Film* (Chicago, University of Illinois Press, 1995), 9.

- ³ Philip K. Dick “The Android and the Human” in *The Shifting Realities of Philip K. Dick* (New York, Pantheon Books, 1995), 185.
- ⁴ Emmanuel Carrere, *I am Alive and You are Dead: A Journey into the Mind of Philip K. Dick*. (New York: Metropolitan Books, 2004), 132.
- ⁵ Carrere, *I am Alive and You are Dead*, 132-133
- ⁶ Paul M. Sammon, *Future Noir: The Making of Blade Runner* (New York: Harper, 1996), 16-17.
- ⁷ Marilyn Gwaltney, “Androids as a Device of Reflection on Personhood,” in *Retrofitting Blade Runner* (Madison: University of Wisconsin Press, 1991) 33.
- ⁸ Philip K. Dick, *Do Androids Dream of Electric Sheep*, (New York: Ballantine, 1968), 4.
- ⁹ *Ibid.*, 147.
- ¹⁰ *Ibid.*, 101.
- ¹¹ *Ibid.*, 31.
- ¹² *Ibid.*, 8-9.
- ¹³ *Ibid.*, 127.
- ¹⁴ Philip K. Dick, *A Scanner Darkly* (New York: Ballantine, 1977), 4.
- ¹⁵ Dick, *Do Androids*, 34.
- ¹⁶ *Ibid.*, 166-7.
- ¹⁷ *Ibid.*, 21.
- ¹⁸ *Ibid.*, 61-2.
- ¹⁹ *Ibid.*, 67
- ²⁰ *Ibid.*, 67
- ²¹ *Ibid.*, 153.
- ²² Joshua F. Dienstag, "Blade Runner's Humanism: Cinema and Representation." *Contemporary Political Theory* 14, no. 2 (2015), 102.
- ²³ *Ibid.*, 105.
- ²⁴ Paul M. Sammon, *Future Noir: The Making of Blade Runner* (New York: Harper, 1996), 17.
- ²⁵ Leonard G. Heldreth, “The Cutting Edges of *Blade Runner*,” in *Retrofitting Blade Runner* (Madison: University of Wisconsin Press, 1991), 41.
- ²⁶ *Blade Runner*. Dir. Ridley Scott. Screenplay by Hampton Francher and David Peoples. 1982. “Director’s Cut” DVD, 1991, Warner Brothers.
- ²⁷ Max R. Bennett, *The Idea of Consciousness* (Sydney: Harwood Academic Publishers, 1997), 164-165.

Galateas Gone Wild: Technology, the Feminine, and Spatial Divides

Cary Elza

University of Wisconsin-Stevens Point

There's a particular appeal in using magic, technology, or art to create the perfect woman, one who promises to love, honor, and obey. Yet from ancient myth on, man's attempt to play God has been a tricky proposition, sometimes resulting in a docile, obedient wife, and other times unleashing an unstable feminine force upon the world. But whether it works or not, the creation of an artificial woman usually represents an attempt to improve upon perceived flaws of women in contemporary society. Greek myth's Pygmalion, for instance, creates an ivory statue of the perfect woman because he's disgusted by the Propoetides, the first prostitutes. According to Ovid's *Metamorphoses*, "...the immoral Propoetides dared to deny that Venus was the goddess. For this, because of her divine anger, they are said to have been the first to prostitute their bodies and their reputations in public, and, losing all sense of shame, they lost the power to blush."¹ The Propoetides' lack of respect for divinity results in their commodification—literally becoming commodities in the marketplace. In response, Pygmalion renounces real women and sculpts his own; for Pygmalion, a completely artificial woman is preferable to real women who become artificial by selling themselves. He eventually prays to Venus to give the statue life, which she does. Pygmalion and Galatea² live happily ever after. The act of creation works out for Pygmalion; his belief in and respect for Venus's power is rewarded. Without that connection to the gods, the myth suggests, we could be condemned to lives of emptiness—sex without love, surface without depth, experience without emotion. Maintain faith, however, and what man creates on earth will be made real, will have moral and spiritual value.

On the other hand, when one questions the divine and dares to play God anyway, bad things happen. Prometheus's famous theft of fire from the gods not only results in liver trouble, but in Pandora, the first woman. Sculpted out of clay at Zeus's behest and given male-pleasing gifts, Pandora could have been an ideal figure. Unfortunately, her feminine curiosity causes her to open a forbidden jar (or box), which releases evil and misfortune into the world. Thus, the first woman, made beautiful and seductive, proves to be the source of

misery for mankind. The moral of the story is similar: kowtow to the gods and your faith will be rewarded, but fail to show respect for the divine and you'll end up with a beautiful, conniving woman, skilled at self-display, who spells nothing but trouble.³ Mankind *needed* that fire to begin technological development and achieve self-sufficiency from the gods—the spark of Prometheus is one of the most pervasive origin myths in the West—but progress came with a sneaky female price tag. I revisit these two familiar stories at the outset of this essay because they both address the intersection of the spiritual world, the artificial, creation/authorship, and femininity. In these early myths, what man creates and what an unknowable, divine force animates are caught in a balancing act. This is an ancient dilemma: advancement towards a perfect society always comes at a price. Modernity is haunted by the specter of soullessness, and that specter finds an apt vessel in media technologies, their representation in popular discourse, and in the fictional narratives they carry.

In the nineteenth century, the age-old fear that commerce, technology, and social change lead to the loss of the spiritual and the irrational becomes more clearly associated with female figures in popular culture. The Victorian doctrine of separate spheres, which divided the tasks of society up according to widely held beliefs about the biological aptitudes of each gender, seemed deceptively simple but proved complex and sturdy. Divvying up both physical (public versus private) and imagined (exterior versus interior) space according to gender meant that, in theory, the feminine, irrational or emotional side of social life, which might include familial matters, spirituality in the sense of both religion and mysticism, and personal memory could be quarantined from the rising tide of commercial culture, technology, and visual stimulation in the masculine public sphere. This was an effective strategy for protecting against fears related to modernity, but also resulted in strictly defined gender roles that relegated the feminine half of society to domestic space. Popular culture often associated the domestic, private sphere not just with a sense of the “interior,” but also with a vast number of imagined spaces, ranging from the realm of the dead (the ancestral realm) to a past ideal world (the Garden of Eden) to the place where God or some concept of spirituality resides (like Heaven or Arcadia). These are not spaces that lend themselves easily to rational explanation—they are subjective, sublime, and sacred.

This gendered distinction doesn't disappear along with Victorianism. Popular culture's tendency to define mysticism, spirituality, and private space as feminine responsibilities or domains extends well into the twentieth century, even as femininity becomes associated with mass culture and the visual spectacle of commerce in the city. The idea that women, the protectors of domesticity, spirituality, and the family, might also stand for the surface glitz and commodification of the urban landscape is fundamentally paradoxical, but that's exactly what happens. This paradox, caused in part by media technologies, also must be solved by media technologies. This essay argues that the tension between the world mankind envisions and constructs in the public sphere—commerce, technology, the trappings of modernity in general—and

an imaginary or spiritual world tethered to the private sphere manifests itself in popular texts about the creation of artificial female figures.

While stories of living, breathing statues have been around for ages, I focus here upon two influential texts from two historical moments, both of which center around the creation of a female robot. The first, Auguste Villiers de l'Isle-Adam's novel *The Eve of the Future*, or *Tomorrow's Eve* (1886), is informed by reactions in the Western world against lingering Enlightenment rationality, and the intersection of science, commerce, and mysticism in nineteenth-century movements such as Spiritualism and mesmerism. The second, Fritz Lang and Thea von Harbou's film *Metropolis* (1926), appears in Weimar Germany in the midst of great social and cultural change, especially with regard to women's increased mobility and freedom, and the rise of mass culture. These issues manifest themselves in a narrative about a futuristic world with great technology but an unstable class structure. Though *Tomorrow's Eve* and *Metropolis* were produced in markedly different historical and cultural contexts, both texts both use the figure of the authored woman, a technological Galatea, as a jumping-off point for negotiating the effect of modernity—of commercialism, mechanical reproduction of images and sounds, and the ever-increasing visibility of women in the public sphere—on spirituality, authenticity, and the private sphere.

These two texts, still popular objects of study, had a certain level of notoriety upon their release. *Tomorrow's Eve* remains relevant and interesting not just because of its protagonist (a fictionalized Thomas Edison), or its status as an early sci-fi novel, but also because of its extravagant misogyny. The novel plays with form and comments on contemporary culture, making it valuable not just in the context of Symbolism and Romantic irony, but also as a document of popular culture on the verge of mechanization; Edison's phonograph had very recently been patented, in 1878, and in 1884, George Eastman developed the photo paper which would replace glass plates. And of course, scholars have analyzed *Metropolis* countless times, even linking it to Villiers' novel, so any essay that seeks to address this film must remain vigilant for redundancy. But while previous scholarship on *Metropolis* has focused on the ambivalence in Weimar culture between a sense of the old and the new, or between technology and nature, there's more to the idea of how gender, public/private space, and technology work together in the film than has been discussed in the past. I hope to show that while very particular historical and cultural circumstances gave rise to each of these representations of women and their respective relationships to imagined spaces, one might prove useful in understanding the other.

In *Tomorrow's Eve*, a sci-fi novel before the term "sci-fi" existed, Villiers creates a fictionalized version of Thomas Edison, whose secret project, the culmination of his entire career, is the creation of a female android named Hadaly (Arabic for "ideal," the text notes). When his friend Lord Ewald shows up one evening at Edison's home, suicidal over the character flaws of his

fiancée, Miss Alicia Clary, Edison suggests a solution—they will map Alicia’s physical attributes onto Hadaly, who will be immune to the corrupting forces of the marketplace. Like Galatea, Hadaly is still dependent upon an animating spark, but fortunately Edison’s psychic assistant is willing to astral-project herself into the robot, giving it spirit and personality and even a soul. At its core, *Tomorrow’s Eve* frets that the artificiality and commercialism of modern society will jeopardize the existence of spirituality and ideal femininity, and offers an alternative: remove women from the equation altogether, and use technology to create something better.

Fritz Lang’s film *Metropolis*, made near the end of the Expressionist film movement in Germany, explores the idea of control over the spiritual and the commercial in a different context, but with many of the same issues at stake. In the futuristic city of Metropolis, the worker drones live underground while the elites live in skyscrapers and frolic in mystical gardens populated by harem girls. At the film’s beginning, this status quo is challenged by heroine Maria, a Mary-like figure who brings dirty, shoeless children into these “Eternal Gardens” to preach unity and brotherhood. Upon meeting Maria, Freder, the son of the city’s leader, leaves the safety of the Edenic Gardens and allies himself with the workers. Meanwhile, Maria is branded a subversive and captured by a mad scientist, Rotwang, who, acting in cahoots with the city’s despotic leader, agrees to copy Maria’s appearance onto the robot, creating a controllable version of Maria to incite class warfare. Robo-Maria, first obeying the deceitful Rotwang and then marching to the beat of her own drummer, takes advantage of Maria’s established influence and riles up the whole city by advocating immorality and anarchy. The masses eventually turn on her for her message of destruction (which caused them to flood the lower city and almost drown their own children) and burn her as a “witch,” while the real Maria saves kids and convinces Freder to take his position as mediator between the workers and the upper classes. She then steps aside, having used her connection to the spiritual to broker a peace, albeit a shaky one that maintains the existing structure of society.

Importantly, both androids I focus on here—Hadaly in *Tomorrow’s Eve* and Robo-Maria in *Metropolis*—have human doubles who provide the physical template for the robot’s lifelike appearance, suggesting that what is at stake here is not simply a negotiation of technological development, but a much deeper interrogation of the roles women play in a changing society. In both texts, the creation of a female robot is an attempt to replace a real, live woman who is performing in a way that threatens existing power structures. Just as Pygmalion sculpted Galatea in response to his disappointment over women who sold themselves, the mad scientists in these texts create their own ideal Eves as a reaction to the reality they see around themselves—a reality that bears little-to-no resemblance to the balanced, utopian societies and spaces they imagine. In these perceived realities, the sacred and the commercial have been confused—the scientists create robots to replace the human because humans have allowed machines to replace them already. In commercial culture,

these narratives suggest, people have become cogs, images, reduced to superficial, monetary value. What place should spirituality have in this culture? In the remainder of this essay, I will discuss how *Tomorrow's Eve* and *Metropolis* negotiate and propose solutions to this modern dilemma by both using the figure of the female robot and her human double to comment on the connections between gender, media technology, and the image, and presenting a narrative in which the *use* of female figures and the *use* of media technology are the keys for restoring balance to a society in transition. Although they were produced forty years apart, and in different parts of the Western world, this novel and this film both show how the act of building and animating a hybrid of femininity and technology is an attempt to use modern technology to gain control over the social changes that come along with modernity—a strategy that arguably hasn't gone away.

In the late nineteenth century, women's increased visibility in the public sphere, particularly in shopping centers, tied them to the role of consumer *and* of consumed image (as spectacles, or as prostitutes). Despite being reduced to an object to be looked at, women's ability to manipulate their own images and performances of femininity in the commercial sphere afforded them power and mobility.⁴ What both *Metropolis* and *Tomorrow's Eve* suggest is that women's ability to produce a distanced self for mass consumption comes naturally, but when they recognize the manipulability of their image, by entering and becoming familiar with the marketplace, they have the potential to become powerful, to see things as they really are, to deceive.

One of the most visible 19th-century figures of woman as commercial spectacle is the prostitute. As historian Griselda Pollock points out, often the spaces of the city where one would be likely to see *both* men and women on the street were spaces of "sexual exchange."⁵ Women in these spaces depended on their own capacity for self-display in order to make a living. Technology enters the picture here not only in the sense of the mass-production of images (for here we have the use of female images in advertising, not to mention the widespread availability of photography), but also in the tools available to alter one's own image. Cinema scholar Patrice Petro quotes Walter Benjamin on Charles Baudelaire, who had plenty to say about women in the public sphere:

In the form which prostitution took in the great cities...woman appears not merely as a commodity but as a mass-produced article. This is indicated in the artificial disguise of the individual expression in favor of a professional one, such as is brought about by the application of make-up.⁶

The artificiality which results from the combination of femininity, public space, and commercialism stands in stark opposition to the female roles and characteristics which Villiers lauds as real or authentic in his novel, the characteristics which allow women to nurture spirituality in the private sphere.

Lord Ewald first comes to Edison because of girl troubles; Alicia, lovely though she may be, has been corrupted by her regular forays into the dreamland of consumerism. By her own admission, she wants to use her angelic face, perfect body, and beautiful voice to pursue a performance career, and eventually a good marriage. This mercenary attitude gives her power in the commercial realm, but also prevents her, in Ewald's opinion, from using her soul for some kind of spiritual connection. The potential for transcendence exists, but she ruins that potential by commodifying art, as Ewald insists:

...it's simply as a matter of business, of *trade*, that her theatrical training enables her to interpret the inspirations of genius into mimic gestures; those inspirations themselves she finds *hollow*...That voice, which lays its golden enchantment on every syllable, is nothing but an empty instrument...She uses it for lack of any other talent, and as if impatient to abandon it (after she's got an easy fortune out of it).⁷

Alicia's largest concern seems to be whether she "*represent[s]* well," Ewald laments.⁸ She attends church weekly and claims to be religious because, she tells Ewald, it's "the thing to be."⁹ Even Alicia's spiritual belief stems from her desired mobility in the marketplace. To make matters worse, "She goes so far as to calculate the profits to be made from her fraudulent virginity."¹⁰

Like *most* women nowadays, according to Edison (and Villiers), Alicia exists only in the world of commodities; she has fashioned herself into a product, and is *very good* at the art of deception. To her lover's dismay, she's all performance, with no authentic self; all surface, no depth. But alas, because "the bonds of beauty are strong and deep," he cannot quit her.¹¹ Villiers forgives Ewald his weakness for physical beauty. Only romantic, spiritual souls like Edison and Ewald can tell the difference between an authentic soul and one imitated through gold-digging feminine wiles, and therein lies the problem: women who roam freely in the commercial realm are skilled at deceiving men through the use of performance and the cosmetic enhancement that modern technology makes possible.

In addition to his deconstruction of Alicia's faults, Villiers relates a lengthy tale about a dancer and singer named Evelyn, who uses her charms to ensnare one of Edison's friends, make him leave his wife, and kill himself. Edison shows Ewald how a "little bloodless creature" can, through the miracles of modern technology, make herself look like a beautiful woman. He shows Ewald a filmstrip of Evelyn dancing, and then as a "tree stripped of its caterpillars," not unlike a tabloid photographing stars without their makeup. This sequence describes soon-to-come filmmaking technology, making a very early suggestion that film's ability to reproduce reality might function to expose artificiality. Like the surgeon in Walter Benjamin's famous analogy in "The Work of Art in the Age of Technological Reproduction," the filmmaker penetrates more deeply into reality just as a surgeon penetrates deeply into a

sick body—that is, with the use of technology. Deconstructing the bits and pieces of Evelyn’s façade—or “screen,” as Edison calls it—disrupts her power of display, and like the surgeon/cinematographer’s resulting image, which is “reassembled under a new law,” the real Evelyn can be exposed. Edison insists that technology has the ability to reveal *or* obscure, depending on the viewer/listener. Through cosmetics, specially-made clothing, and products of the “chemical industry,” Evelyn presented a front that concealed “that underneath all her paraphernalia the hybrid creature of his passion was as false as his love itself—to the point, in fact, of being nothing but *the Artificial giving an illusion of life*.”¹² Here’s the problem, according to Villiers: women are realizing that they can use modern technology and commercial savvy to profit in the marketplace. But because of the taint of commerce, the authenticity of the spiritual is in jeopardy. The solution to this, according to Villiers’ Edison, is to use science and technology to map Alicia’s attributes onto an obedient android.

Mad scientist Rotwang comes to a similar conclusion in *Metropolis*, but the context of women’s images in the public sphere has changed—in large part because of the introduction of the cinema. In the 1920s in Germany (and in much of the rest of the West), mass culture “was commonly personified as ‘feminine,’ as having the capacity to induce passivity, vulnerability, even corruption,” as Patrice Petro argues, and the androgynous and more revealing fashions of the day certainly attested to women’s ability to control self-display, and hence male desire.¹³ Mary Ann Doane, Petro, Miriam Hansen, and other scholars have addressed the relationship between femininity and cinema in the silent era in great detail. Briefly, the cinema incarnates basic contradictory fears and desires of modernity—to humanize the machine, lessening its threat, and to technologize the human, making it more rational and efficient through mechanical means (as prosthetic to the body, etc.) But since these things are scary and appealing at the same time, the female body ends up as a primary bearer of the burden of the look in cinema. Woman becomes mechanized spectacle, and technology becomes feminized in popular discourse.

What happens when women watch movies in this context? Petro insists that the very “presence of women in the modern city—on the streets, in industry, in the arts, and in the cinema—obviously distracted the attention of male intellectuals, who aimed to efface or at least to contain the power of the female gaze.”¹⁴ Social critics like Siegfried Kracauer noted that women in the theaters were deeply involved in the emotions of the characters onscreen. Somehow the little shopgirls are closer to the text, closer to the fictional world onscreen—there’s *something* about what these women receive from the moving image that Kracauer can’t quite grasp, and so he pathologizes it. And yet women’s closeness to the text here ends with a reasserting of self-as-image, as if the cinema has allowed these women to let down their guard for the duration, becoming one with the image onscreen, before they must reconstruct the gap between image and self, leave the theater, and walk down the street as

a spectacle in public space. As Kracauer significantly concludes, “Clandestinely the little shopgirls wipe their eyes and powder their noses before the lights come up.”¹⁵ If women could be both sutured into the image and not sutured, part of the story but aware of its artificiality, then so too could they see their own images in daily life in a doubled manner. This idea has a lot in common with Joan Riviere’s concept of the masquerade or femininity as an always-constructed image, which Doane artfully applied to cinematic viewing positions.¹⁶ And, given the above discussion of “excessive” femininity in *Tomorrow’s Eve*, it doesn’t seem like a stretch to suggest that all of these elements—femininity, the image, and technology—are bound up together in modernity and in these texts.

The robot in *Metropolis*, which Rotwang creates as a Galatea-like mate for himself, represents both the power of technology and the power of female sexuality. Once Robo-Maria receives Maria’s face and is unleashed onto an unsuspecting public, she takes the spiritual authority Maria had—her connection to both another realm and the people of Metropolis—and perverts it. Every night, she performs a strip act in the Yoshiwara District, Metropolis’s Las Vegas strip, and adds an exotic, Mata Hari-like twist to it, mobilizing the age-old association of out-of-control sexuality with “primitive” cultures. This public sexuality, as Peter Wollen suggests, also distinguishes an “authentic” femininity (if it indeed exists) from one that relies upon manipulation of image: “The True Maria contains pleasure within the private sphere, whereas the False Maria carries it into the public sphere - Yoshiwara, the site of spectacle (music and sexual display).”¹⁷

When Freder, hallucinating in his bed, imagines Maria as the Whore of Babylon, it is unclear whether Robo-Maria has actually included this bit of theatrics in her act, or Freder just imagines it. In any case, the oft-analyzed scene in which Robo-Maria dances for the first time in Yoshiwara includes a fascinating shot of men’s eyes, big and small, spiraling around the frame, as if watching Robo-Maria with one gaze made up of many. Robo-Maria, with her knowledge of the power of her image (after all, she *is* a mechanical reproduction) and how to use it to gain control over others, has the ability to reduce a crowd of men into one seething, powerless mass. In addition to deploying her skills of self-display in the upper-class marketplace, Robo-Maria also takes advantage of Maria’s influence and riles the workers up even more, to the point where they storm the Machine District and destroy the Heart Machine. The Heart Machine (or MOLOCH the monster, as Freder hallucinates), too, has feminine connotations. Not only does the gaping mouth of the beast resemble a vagina dentata, but Maria’s frequently-repeated mantra, “The heart must mediate between the head and the hands,” suggests that up until this point in Metropolis, the function of the “heart” has been performed *not* by a savior like Freder, *not* by a female conduit, but instead by a giant, brutal, feminized machine. Technology has swallowed up all attempts at mediation between a rational, scientific, commercialized world and an irrational, spiritual world.

Mediating Modernity

Indeed, technology was a double-edged sword. Communications technologies such as the phonograph and the photograph allowed people to store ostensibly objective information for the first time—instead of the past being mediated through individual writing, or painting, machines could capture history as it happened, and add it to an archive. And as the nineteenth century gave way to the twentieth, the invention of the cinema brought an even more “realistic” way to capture time, to collect information in a tangible form. But on the other hand, technology brought standardization and monetization, turned men into cogs in machines and threatened to distract and dazzle the masses away from traditional values. Popular fiction offered an outlet for exploring the uncanniness of such figures for those concerned about the effects that technology had on what it meant to be human, both in body and mind. It’s no coincidence that narratives in which a mad scientist or artist creates and brings an ideal woman to life begin appearing with regularity in Romantic literature in the early nineteenth century. E.T.A. Hoffman’s “The Sandman” (1816) and “Automata” (1814), not to mention Mary Shelley’s *Frankenstein* (1818), paved the way for authors to replace the divine spark with “scientific” explanations, in effect secularizing the myth of creation.¹⁸ The relationship between humans and machines would play a large part in whatever future mankind designed, these texts made clear, and coming to terms with the roles robotic or electrical figures—particularly female ones—might eventually take on was a primary concern when describing a potential utopia.

The decision to create a female robot in both these narratives comes about because of women’s ability to see themselves as dominant society wants to see them: as spiritual conduits in the private sphere, and as powerless spectacles once they enter the public sphere. If, in consuming popular culture, in seeing mechanically reproduced images and the female body on display, women figure out that they themselves *are* images, then they can manipulate their own images in the public sphere. They can use this handicap that dominant society has given them—as pure surface—to protect an inner life. There’s a gap there from day one, a gap between some fundamental, unreachable truth of identity and the mask that one presents to the world (according to the theory of the masquerade, this mask is worn in public *and* private).

But once women are as deeply invested in the commercial realm as men, then the inauthenticity of the image makes its way into all areas of life. Peter Wollen describes Walter Benjamin’s concept of the aura:

Exhibition value, brought to dominance by the copy, brings the masses close to reality, whereas cult-value, the province of the original, required contemplation at a distance, to be rewarded by a spiritual, rather than a secular and real, experience. Thus the decay of the “aura,” the special quality

of cult-value, is part of the general decay of magic, theology, and metaphysics.¹⁹

This is both good and bad—the sense of spiritual authority which kept the working classes at a distance from structures of power such as the church has been eroded, because now mechanically-reproduced art, like cinema, reaches a greater number of people who can hold it closer (like the little shopgirls). But this comes at the cost of the loss of spiritual experience. In “Work of Art,” Benjamin describes the loss of the spiritual realm in modernity in a different but related way. When one adds gender to this equation, if the development of media technology in the commercial sphere is no longer kept separate from a spiritual-magic private sphere, and the figure of woman-as-image threatens the figure of woman-as-conduit, then woman’s spiritual authority is compromised, rather than contained. What these narratives of women and robot doubles allow is the negotiation of all of these issues at once—the merging of public and private spheres, the overlapping of commercial and spiritual realms, and the confusion of woman-as-image with woman-as-spiritual-conduit.

Reviewing *Metropolis* when it was first released in Germany, critic Willy Haas writes, there is “a certain depth to the motif of the female doppelganger, which embodies both the unleashed hell of lust and the most tender virtue of the...but now we are straying down devious paths indeed!”²⁰ Through narratives of the process of making a copy of a woman from an original, these texts mull over the evacuation of soul, or spirituality in a commercial world where image and authentic self are separated as a matter of course, and even spirituality can be used to make money. The female and her robot double draw attention to the constructed nature of femininity. The robot in *Metropolis* literally receives Maria’s face, just as Hadaly receives Alicia’s face, but the contexts are different; by the time *Metropolis* comes out, the basic altering of perception which cinema causes arguably produces more acute (if subconscious) awareness of the image as separate from the real self, just as the mechanically reproduced copy (the cinematic double) is separate from, and performs a different function from, the auratic original.

In the context of the late nineteenth century, the construction of a female robot allows Villiers to toy with the idea of a perfect woman for the modern age—a controllable piece of technology who exists solely in the private realm, and maintains a connection to a spiritual realm. Hadaly solves all the problems of modernity in one fell swoop. Understandably, Ewald hesitates when Edison suggests that they build a robot version of Alicia—it seems morally suspect and just plain weird to enter into a relationship with a robot. To this and any other potential detractors, though, Edison sermonizes:

Since our gods and our aspirations are no longer anything but *scientific*, why shouldn’t our loves be so, too? In place of that Eve of the forgotten legend, the legend despised and discredited by Science, I offer you a scientific Eve—the only

one, I think, now worthy of those blighted visceral organs which you still—by a kind of sentimentality that you're the first to mock—still call your "hearts."²¹

Hadaly, Edison claims, is the best and only solution for a civilization gone wild. If men insist upon allowing women to corrode their morality and spirituality from the inside out, then who will raise the children and maintain the moral center of daily life? It is implied that there are still honorable women out there, but Villiers is more interested in solving the commercial ills creeping into the hearts of most women nowadays. He uses Alicia's performance aspirations against her, "hiring her" for a new play (called *Tomorrow's Eve*, of course). Thinking she's rehearsing for the play, Alicia records lines into a phonograph which etches her words onto the "golden cylinders" that will serve as Hadaly's database of aural information. She also consents to have her body scrutinized from head to toe, because having a full-body statue of yourself, Edison tells her, is the modern thing to do. Besides persuading her to pose for a statue, Edison and Ewald frequently compare Alicia to the famous *Venus Victorious* (or *Victrix*), which we now know as the *Venus de Milo*.²² Annette Michelson points out that "Miss Alicia is afflicted with reason. Were she unreasoning Lord Ewald could understand and accept. The marble Venus, after all, has made no compact with reason. The goddess is veiled in mineral and silence."²³ Alicia's problem is that she's NOT a statue; as a thinking, calculating human being, she lacks the necessary passivity to tune into the spiritual world.

The reenactment of the Pygmalion myth lies at the very core of *Tomorrow's Eve*. Edison has under his command a powerful medium named Miss Anny Sowana, ready and willing to inhabit the android Hadaly, providing her with a personality, a soul, and a direct line to the spirit world. Conveniently, Edison holds the reins to both the technology of Hadaly and the spirituality of Sowana. Years of conditioning have primed the wandering spirit to behave in the docile, loving way Edison condones, rather than attempting to use her access to the occult in subversive ways. Sowana's mobility across multiple planes could afford her great power; therefore, she must be contained, literally, within a controllable technological device. Through the merging of art, science, and technology, and with the help of this friendly medium, Edison and Ewald figure out exactly how to staunch the bleeding of the commercial into the spiritual. What was once publicly visible becomes re-ensconced in the private sphere. Hadaly is beautiful and capable of sublime performance, but now her recitation of prose and spectacle of self is for private eyes and in service of her master. The cherry on top: when Ewald isn't using Hadaly, he can keep her in the beautiful ebony coffin Edison has provided, like a doll in a toy box.²⁴ Alas, Ewald's happiness is short-lived—the steamer carrying him, Hadaly, and the real Alicia across the Atlantic sinks, sending both Alicias to a watery grave. He

cables Edison at the end of the novel, telling him that he grieves only for Hadaly, “that shade.”²⁵

The world of *Metropolis*, arguably, is the future that Edison predicted in *Tomorrow's Eve*—everything is technology, technology is commerce, and science rules *über alles*. People have become cogs, and society is highly stratified in order to keep the machine functioning. Just as Edison lamented that “our gods and our aspirations are no longer anything but *scientific*,”²⁶ the city’s leaders have created a secular world that, through technology, imitates a spiritual world for the highest echelon of society while devaluing human life in the lower classes. Metaphorically and literally, this is an unstable way to live. This becomes clear near the beginning of the film, when Freder sees a machine malfunction in the Machine District. He imagines the machine as Moloch, a giant monster eating human sacrifice victims, but it’s just a machine that treats human beings as mechanical parts. Technology might be able to replace the divine spark for a while, but a fully public world, one in which the commercial sphere is the only sphere, effectively prevents the pursuit of meaning, of value beyond the material.

In the context of Weimar Germany, we also might consider this sphere of meaning as the source of aesthetic meaning. Influential film scholar R. L. Rutsky writes,

The aesthetic realm is defined explicitly in opposition to the technological. In contrast to the dead, partial, technological object, which has no intrinsic meaning or purpose, the aesthetic object is seen as an organic whole, as having, like a living thing, its own, internal meaning and purpose... Art itself comes to represent that old sense of magico-spiritual enchantment or animism, of an eternal spirit, repressed by the modern, technological world.²⁷

Art, beauty, spirituality and the past are all tied up in the figure of Maria, who spurs Freder to action by acting as a vessel which facilitates the coming of the mediator, like Mary gives birth to Jesus. But the film doesn’t just posit the value of a mediator between magic and technology, old and new. By presenting a narrative in which the *use* of female figures and the *use* of media technology is key to the restoration of balance, both of these narratives attempt to solve the paradoxical fears and desires of modernity in their respective societies. Rutsky writes that

[The film’s] restoration of the home and the familiar is not simply an abandonment of modernity in favour of the ancient, the natural, the spiritual. Rather, *Metropolis* aspires to a mediation of the masculine technological will of modernity with an ancient feminine spirit or nature.²⁸

Female figures in these cases are tools, not agents. The woman who uses her own image for material gain is contained, the woman who uses her spiritual connection for public recognition is contained, and woman happily accepts the position as protector of the private realm *and* moral center (or is destroyed). The woman herself is evacuated from the utopia she helps to create—she, like spirituality, has to remain behind the scenes.

The end of *Metropolis* is frustrating, because it seems like the masses have been quelled, and the status quo has been restored. But as Rutsky and others have concluded, there's a combination of the old and new here, technology and nature—gaps have been bridged. The joining of aesthetics and technology, and the joining of a mythic, nostalgic home with a technological future accomplishes the same general goal that Hitler set out to achieve; as Rutsky writes, Hitler exploited exactly the same sort of discontinuities in German society that *Metropolis* sets out to reconcile: “What the *Fuhrer* and Freder promise, then, is access to both the resolution of a gendered modernity into an *androgynous* whole—the National Socialist state.”²⁹ Maybe not completely androgynous, but at least the threat of out-of-control femininity and out-of-control technology have been contained. Woman's status as a self-aware spectacle is under control, and Maria, as a representative of the nurturing, maternal female of the private sphere, has triumphed over evil, sexy Robo-Maria. The danger of the unchecked commercial/spectacle has been made clear (the riots in Yoshiwara) and so has the mechanization of human life (the masses destroyed the Heart Machine without knowing the consequences). By destroying the mechanical Maria by burning her at the stake, the workers and the upper class together enact a ritual of control over technology and the use of female sexuality for power in the public sphere. Maria facilitates the final mediation, pushing Freder into the limelight and the public role as mediator. Finally, she steps into the background, retaining her connection to the spiritual but allowing *him* to be the public voice.

Walter Benjamin believed that demystifying previously auratic structures could reduce their power over the people, essentially making the private public. In this context, it seems that women's movement into the public sphere would be good, because women, or those more connected to the mystical, grant spirituality to public space. But in that public space, these texts say, the spiritual realm has the potential to be corrupted by the marketplace. According to *Metropolis*, the secrecy of the spiritual/mystical must be exposed to the public, but women have to *return* to the role of moral center. To make sure that the merging of the public and private doesn't mean the evil use of the spiritual for commercialism, woman must embrace the role of a conduit who facilitates things, rather than claiming agency of her own, or power over her own image. What a “balanced society” looks like changes, but what *Tomorrow's Eve* and *Metropolis* point out is that humanity and emotion and an “other” realm that defies rationalization must be protected from the rationalization and technological progress of modernity. And while the solutions they propose

range from devastating to women to just plain bad for women, the process of working these issues out gives us insight into how all these elements come together in ways that provide glimpses at women's potential power and authority in mainstream society.

Notes

¹ Ovid, *Metamorphoses*, Book X, http://www.mythology.us/ovid_metamorphoses_book_10.htm, accessed Feb. 9, 2009.

² The name "Galatea" was given to the statue by Jean-Jacques Rousseau in his 1770 play *Pygmalion*, and has since become widely used.

³ Of course, in Christian and Jewish mythology, Eve, whom God creates out of Adam's rib, is similarly the cause of Original Sin and expulsion from the Garden of Eden. And the story of Lilith, traceable to Mesopotamian, Sumerian, and early Jewish and Christian texts, similarly paints a picture of a deceitful female, though in many accounts Lilith is either a female demon or has been created as man's equal, not his subordinate.

⁴ When women leave the home, participating in capitalism as consumers, they are confronted with a set of issues surrounding vision and spectacle. The basic problem they're dealing with is the same for men—how to reconcile self and individuality with becoming monetized in the public sphere—but as spectacles and conduits defined by what others expect and desire of them, women have a different relationship to the mechanically reproduced image than do men. When the classic male figure of the flaneur, the city wanderer, makes his way through the city streets, he must distance himself from the crowd to avoid getting lost in it. This practiced gap between an authentic self and a self for mass consumption, according to Baudelaire and others who have since picked up the concept, arises in response to the panorama of visual culture in the nineteenth century city, and the need to assert some kind of control over an overwhelming, stimulating landscape of spectacles.

⁵ Griselda Pollock, *Vision & Difference: Femininity, Feminism and the Histories of Art* (New York: Routledge, 1988), 70.

⁶ Walter Benjamin, "Central Park," trans., Lloyd Spenser *New German Critique* 34 (Winter 1985): 52-53, quoted in Patrice Petro, *Joyless Streets: Women and Melodramatic Representation in Weimar Germany* (Princeton, NJ: Princeton UP, 1989), 165.

⁷ Auguste Villiers de l'Isle-Adam, *Tomorrow's Eve*, Robert Martin Adams, trans. (Chicago: University of Chicago Press, 2001), 34.

⁸ *Ibid.*, 33.

⁹ *Ibid.*, 42.

¹⁰ *Ibid.*, 34-35.

¹¹ *Ibid.*, 38.

¹² *Ibid.*, 121, 122.

¹³ Petro, *Joyless Streets*, 8.

¹⁴ *Ibid.*, 69.

¹⁵ Siegfried Kracauer, "Die kleinen Ladenmadchen gehen ins Kino," *Das Ornament der Masse* (Frankfurt am Main: Suhrkamp, 1977), 292-293, quoted in Petro, 67.

¹⁶ See Mary Ann Doane, "Film and the Masquerade: Theorising the Female Spectator," *Screen*, 23, no. 3-4 (1982), 74-87, and Joan Riviere, "Womanliness as a Masquerade,"

Psychoanalysis and Female Sexuality, Hendrick M. Ruitenbeek, ed. (New Haven: College and UP, 1966), 213.

¹⁷ Peter Wollen, "CINEMA/AMERICANISM/THE ROBOT," *New Formations* 8, (Summer 1989), 18-19.

¹⁸ Andreas Huyssen chronicles the development of the android in literature and popular culture, so I see no need to repeat that here; essentially, he writes, with the rise of mechanized industrial labor in the early nineteenth century, the robot becomes more of a threatening figure, and tends to be depicted as female. He writes, "Historically, then, we can conclude that as soon as the machine came to be perceived as a demonic, inexplicable threat and as harbinger of chaos and destruction...writers began to imagine the *Maschinenmensch* as a woman." Furthermore, the confusion between women's connection to the natural, to the 'primitive' and their association with burgeoning technology could be explained according to Freudian concepts of displacement and castration anxiety: "This projection was relatively easy to make; although woman had traditionally been seen as standing in a closer relationship to nature than man, nature itself, since the eighteenth century, had come to be interpreted as a gigantic machine. Woman, nature, machine had become a mesh of significations which all had one thing in common: otherness; by their very existence they raised fears and threatened authority and control." Andreas Huyssen, "The Vamp and the Machine: Technology and Sexuality in Fritz Lang's *Metropolis*," in *New German Critique: An Interdisciplinary Journal of German Studies*, vol. 24-25 (1981-1982), 226.

¹⁹ Wollen, CINEMA, 20.

²⁰ Willy Haas, "Two Major Film Premieres," *Die literarische Welt* 3 (January 21, 1927), 7, reprinted in *Fritz Lang's Metropolis: Cinematic Visions of Technology and Fear*, Michael Minden and Holger Bachmann, eds., (Rochester, NY: Camden House, 2000), 86.

²¹ Villiers, *Tomorrow's Eve*, 163-4.

²² As an historical aside, Villiers would certainly have been aware of the *Venus Victorious* statue (1808), which Napoleon commissioned from Antonio Canova as a portrait of his sister Pauline Borghese. Interestingly, the statue became associated with a slight connotation of impropriety--Borghese's social behavior, along with her self-aggrandizing insistence that she appear as Venus, didn't exactly endear her to the public. So Ewald's comparison of Alicia to a statue of Venus has a variety of implications: the divine hand of Venus in the Pygmalion myth, the priceless authenticity of the ancient Greek *Venus de Milo*, and perhaps most importantly, the skillful—yet ever-so-slightly tainted—imitation of Greek portraiture in the 1808 *Venus Victorious*. See Fred S. Kleiner, Christin J. Mamiya, and Richard G. Tansey, *Gardner's Art Through the Ages, 11th Edition, Volume II*, (New York: Harcourt College Publishers, 2001), 853-4.

²³ Annette Michelson, "On the Eve of the Future: The Reasonable Facsimile and the Philosophical Toy," *October* 29, (Summer, 1984), 7.

²⁴ Villiers died in 1889, just before Edison's talking doll was available for purchase (1890).

²⁵ Villiers, *Tomorrow's Eve*, 219.

²⁶ *Ibid.*, 163-4.

²⁷ R.L. Rutsky, "The Mediation of Technology and Gender: *Metropolis*, Nazism, Modernism," *Fritz Lang's Metropolis: Cinematic Visions of Technology and Fear*, Michael Minden and Holger Bachmann, eds., (Rochester, NY: Camden House, 2000), 236.

²⁸ *Ibid.*, 231-232.

²⁹ *Ibid.*, 234.

Notes on Contributors

M. T. Anderson writes for both young people and adults. His satirical sci-fi novel *Feed* was a Finalist for the National Book Award and the winner of the L.A. Times Book Prize. His Gothic novel of scientific experimentation during the Revolution, *The Pox Party*, won the National Book Award in 2006. His nonfiction book about World War II, *Symphony for the City of the Dead: Dmitri Shostakovich and the Siege of Leningrad*, was named a best book of 2015 by the *Wall Street Journal*, *The New York Times*, *The San Francisco Chronicle*, and others. His most recent book is a graphic novel, *Yvain*.

Cary Elza is assistant professor of Media Studies at University of Wisconsin-Stevens Point, where she teaches courses on screenwriting, film and media analysis, history, and genre. She received her Ph.D. from Northwestern, where her research focused on early cinematic adaptations of *Alice in Wonderland* in the context of technological and social change. Her publications include articles and anthology chapters on children's and teen media, science fiction and fantasy series, new media and fandom, and early animation.

Maria Aline Ferreira is an associate professor at the University of Aveiro in Portugal where she teaches English Literature. Her main interests include the intersections between literature and science as well as literature and the visual arts, bioethics, feminist utopias and women's studies. Her book *I Am the Other: Literary Negotiations of Human Cloning* was published by Greenwood Press in 2005. Recent publications comprise articles on feminist utopias, eugenics, biotechnological dystopias and bio art. She is now working on a book provisionally entitled: *The Sexual Politics of the Artificial Womb: Fictional and Visual Representations*.

John Gagnon is an assistant professor of Rhetoric & Composition at the University of Hawaii at Manoa. He is a cultural rhetorician interested in human rights and the framing of human rights issues, and is particularly interested in issues of exploitation, trauma, and research methodology. He has published in *Poroi*, *Present Tense*, *Interdisciplinary Humanities*, and the *Journal of War Crimes, Genocide and Crimes Against Humanity*, among other academic journals. He holds

a Ph.D. from Michigan State University, J.D. and M.S. degrees from Tulane University, and a B.A. from Andrews University.

Kirsten Gerdes is assistant professor of Philosophy and Humanities at Riverside City College, and a Ph.D. candidate in Women's Studies in Religion at Claremont Graduate University. She is currently finishing her dissertation, in which she uses Michel Foucault to read John Calvin's work in order to explore Protestantism's influence on the emergence of the modern self. Her interests span women's and gender studies in religion, cultural studies, and the philosophy of personhood. Her publications include an article entitled "Materiality of Metaphor: The Risk of Revelation" and a chapter entitled, "Forgiveness and Afro-pessimism, Or, Can Beyoncé Tell Us Anything About Justice?"

Tracy L. Hawkins graduated from Claremont Graduate University in 2011 with a Ph.D. in Women's Studies in Religion, and she is currently in her fourth year as an assistant professor in the Department of Philosophy & Religious Studies at the University of Wisconsin-Whitewater. Her research interests include feminism and cultural studies, particularly technology, television and film, social responsibility, and agency. Her other publications cover topics such as Facebook, Twitter, Fox's *Glee*, Netflix's *Orange is the New Black*, and the 2011 film *Crazy, Stupid, Love*.

Andrew Howard studied Creative Writing throughout college, gaining an M.A. from Texas Tech University in 2007 and M.F.A. from Georgia College & State University in 2010. Since then he has taught in a variety of places, including GED preparation in rural Georgia; Composition, Literature, and Creative Writing in Ohio; and is currently the Program Coordinator of English at University of the District of Columbia Community College. His fiction and non-fiction have appeared in places such as *Southwestern American Literature*, *Bluestem Magazine*, *Hobart*, and *McSweeney's*.

Zoltan Istvan is a futurist, the author of *The Transhumanist Wager* (an International Books Award winner), and a libertarian candidate for California governor. He has written for *National Geographic*, *Psychology Today*, and many, many other publications. He is a successful entrepreneur, inventor, and media consultant. For more about Zoltan, go to www.zoltanistvan.com.

Matthew Landers is an associate professor at Western Norway University of Applied Sciences (Høgskulen på Vestlandet) in Bergen, Norway. His research interests converge within the domain of Biohumanities, entertaining a wide range of interdisciplinary subjects that straddle the gap between the traditional humanities and the sciences. He is co-editor of the volume, *Anatomy and the Organization of Knowledge, 1500-1850*, with Routledge Press. His current research

focuses on the ethical and philosophical underpinnings of conservation science in the Anthropocene era.

Caroline Mosser is an educator, translator, writer and independent scholar. She has lived and worked in both France and the United States, and is looking forward to more adventures. After earning her Ph.D. in Comparative Literature from the University of South Carolina, she has taught French as a postdoctoral fellow at Utah State University. As a graduate student, she participated in and taught for the competitive Social Advocacy and Ethical Life scholarship program at USC. Having always been fascinated by both the sciences and the humanities, she has focused her work on their connection through the study of the representations of science and technology in popular culture. A life-long learner, she enjoys sharing her knowledge through teaching and participating in various academic and cultural projects (such as translating, interpreting, and editing).

Doré Ripley teaches English at California State University, East Bay and Diablo Valley College. Her courses and studies include comics, fantasy and science fiction, short film, literature, composition and rhetoric. She is a regular contributor to *Graphic NovelReporter.com* and her comic work has been featured on *Truthout.org*, *20something reads.com* and *Interdisciplinary Humanities*. In addition, she routinely publishes academic articles on genres ranging from fairy tales and comics to science fiction and medieval lit. For more about comics in the college classroom, go to www.RipleyOnline.com.

Steven Yu is an illustrator whose influences stem from science fiction and several genres of film including 1980s and early 1990s action flicks, neo-noir, and martial arts. His current project is a graphic novel about a vengeful orphan who hires a mysterious assassin to help find her mother's killer: a bio-engineered creature from an elite police unit known as "The Hybrids." The *Interdisciplinary Humanities* cover artwork was created with the use of hands-on and digital media. You can support Yu's work at www.patreon.com/sydstudio and check out his portfolio at www.cargocollective.com/stevenyu.

Editorial Policies

The editors at *Interdisciplinary Humanities* define “interdisciplinary humanities education” as any learning activities with content that draws upon the human cultural heritage, methods that derive from the humanistic disciplines, and a purpose that is concerned with human values. Academic courses don’t have to be labeled “humanities” to be interdisciplinary. Integrated courses and units are often disguised under such names as World History, Freshman English, Music Appreciation, Beginning Spanish, Introduction to Religion, Senior Honors, etc. Integration can range from the use of a novel in a history course to team teaching to comprehensive thematic extravaganzas that combine the arts, literature, philosophy, and social sciences.

Although much of our emphasis is on college liberal arts, *Interdisciplinary Humanities* welcomes manuscripts dealing with elementary grades, teacher education, adult public programs, graduate seminars, educational radio and television, museums, and historic parks. Readers of *Interdisciplinary Humanities* share an interest in interdisciplinary approaches to scholarship and teaching, and the editors favor submissions that draw on that tradition. Feel free to employ first and second person, but do not feel constrained to be colloquial.

Interdisciplinary Humanities uses the Chicago Manual of Style. All notes should appear as endnotes at the conclusion of the essay, and should precede any bibliographical listings and appendixes. Submissions should include full bibliographic citations. Submissions that include reproductions of images should include copyright permission; *Interdisciplinary Humanities* will not publish any submission without written permission for reproduced and copyrighted images. Camera-ready line illustrations and high resolution black and white photographs often reproduce well.

Essays should be typed and double-spaced, formatted for printing, on standard paper with one-inch margins and submitted electronically as Microsoft Word documents to co-editors: Stephen Husarik, shusarik@uafs.edu, and Lee Ann Elliott Westman, leeann.westman@rutgers.edu. Author’s name and institutional affiliation should appear in the upper right-hand corner of the first page of the manuscript. Essays should not exceed 6,000 words. *Interdisciplinary Humanities* observes a “blind reading” policy and considers carefully the recommendations of outside readers whose expertise corresponds with the essay’s subject matter. Permissions to reprint images and illustrations, if any, are the responsibility of the author and should be arranged for and paid before submitting the article. Authors whose essays are accepted will receive a pre-press style sheet with reformatting instructions, as well as a “Consent to Publish” form which must be returned before the issue is published. Authors whose work is accepted for publication must join the Humanities Education and Research Association.

Interdisciplinary Humanities
Journal of the Humanities Education and Research Association

Membership/Subscription Form

Name _____
Address _____
City _____ State _____ Zip _____
Phone _____
E-mail Address _____
University Affiliation (if any) _____

Yes, I'd like to become a one-year member of the Humanities Education and Research Association and receive three issues of *Interdisciplinary Humanities* along with regular newsletters announcing upcoming events and conferences.

- _____ Individual 1-year membership: \$130.
- _____ Joint 1-year membership (open to two individuals sharing the same household): \$190.
- _____ Full-time student membership: \$75.
- _____ Retired: \$75.
- _____ One-year library subscription to *Interdisciplinary Humanities*: \$167.
- _____ Institutional 1-year membership: \$500 Institutional members are recognized in conference materials and on the HERA website. Individuals employed by institutional members enjoy reduced membership and conference fees. (A 1-year membership fee for an employee of an institutional member is \$95.)

All checks must be made out in U.S. currency. All memberships begin January 1 and end December 31. No exceptions.

Enclose a check or money order for category indicated above, payable to:

Humanities Education and Research Association
Send to: HERA, P.O. Box 715, Pacifica, CA 94044-4206