HOMO ARTIFICIALIS: HUMANITY AS A PERFORMATIVE CONSTRUCT

by

MICHELLE MILLS SMITH

(Under the Direction of David Z. Saltz)

ABSTRACT

Robot. Golem. Automaton. Android. The words for the "artificial human" are manifold, their composition as varied as their names—yet the concept is as old as history itself. From ancient creation mythology to modern household robots, humankind has had a tempestuous relationship with the artificial human. Whether formed of clay, grown from living tissue, or built of gears and circuitry, the artificial human is both the object of fascination and of terror. Across the centuries-old history of this relationship, the artificial human has proven a sounding board for our own ideas of "humanity" as realized within the cultural and intellectual millieux of any given historical moment. As the artificial human moves from our imagination into our reality, the "human yardstick" is called ever more into question, forcing us to redefine exactly what we mean by the term "human." I propose that this redefinition is of necessity a performative one—that the essential traits of "being human" are based not in our genetic identity as *Homo sapiens* but in our social, cultural and behavioral identities, identities that we continually perform and modify based on the expectations of the historical moment. As such, this identity has been historically reinscribed in the artificial humans of our imagination, and is currently crossing into our reality through the emergence of social robotics.

This study serves two purposes: first, as a transhistorical journey through the history of the artificial human, focusing on those moments when the relationship between the natural and the artificial human are their most poignant. This history leads from the *kourai khryseai* of Hephaestus to the sophisticated robots of modern science fiction, and thus the second facet of the dissertation: a demonstration of the performative nature of "humanity" through case studies of these sci-fi robots—beings that I argue hold claim to the status of *Homo artificialis*, true artificial people—and a glimpse into the implications of this as sociable robots enter our everyday lives.

INDEX WORDS: social robotics, performance studies, performativity, science fiction, artificial intelligence

HOMO ARTIFICIALIS: HUMANITY AS A PERFORMATIVE CONSTRUCT

by

MICHELLE MILLS SMITH

B.A. Centre College of Kentucky, 1990

M.A. Bowling Green State University, 2004

A Dissertation Submitted to the Graduate Faculty of The University of Georgia

in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2014

© 2014

Michelle Mills Smith

All Rights Reserved

HOMO ARTIFICIALIS: HUMANITY AS A PERFORMATIVE CONSTRUCT

by

MICHELLE MILLS SMITH

Major Professor: David Z. Saltz

Committee: John Kundert-Gibbs

Christopher Eaket

Walter D. Potter

Electronic Version Approved:

Julie Coffield

Interim Dean of the Graduate School

The University of Georgia

December 2014

DEDICATION

For Kevin and Sheridan

TABLE OF CONTENTS

TABLE OF FIGURES viii

INTRODUCTION: ECCE HOMO 1

GOAL OF THIS STUDY 7

IN OUR OWN IMAGE 9

FROM HUMANOID ROBOT TO *HOMO ARTIFICIALIS* 11

THE ROBOT AS A TRANSHISTORICAL CONCEPT 14

HUMANITY AS A PERFORMATIVE CONSTRUCT 21

CHAPTER OVERVIEW 25

CHAPTER ONE: FROM MAGIC TO MACHINERY 29

SUPERNATURAL, MECHANICAL, BIOLOGICAL 36

THE WONDERS OF ANTIQUITY 38

DEUS FABER 47

THE ISLAMIC GOLDEN AGE 52

FROM ANTIQUITY TO ARTOIS: THE GARDEN OF HESDIN 58

AUTOMATA, HOMUNCULUS, GOLEM 65

THE DEATH OF MAGIC 78

CHAPTER TWO: THE CRISIS OF THE MAN/MACHINE 82

THE RISE OF MECHANISM 85

A CLOCKWORK UNIVERSE 89

RENAISSANCE ROBOTICIST 98

A LIVING MACHINE 106

ENLIGHTENMENT AUTOMATA: VAUCANSON AND JAQUET-DROZ 114

IDIOT AUTOMATA AND THE VITALISTIC PRINCIPLE 130

A SHOCKING DISCOVERY: THE ARGUMENT FROM BIOLOGY 134

SENSE AND SENSIBILITY: THE ARGUMENT FROM SENTIMENTALITY 138

MASTER OF HIS OWN FATE: THE ARGUMENT FROM AUTONOMY 142

FRANKENSTEIN'S CREATURE AS THE FIRST ARTIFICIAL *PERSON* 147

CHAPTER THREE: THE PATH TO PERSONHOOD 159

THE FIRST ROBOT 161

THE RISE OF THE COMPUTING MACHINE 176

ALAN TURING AND THE QUEST FOR ARTIFICIAL INTELLIGENCE 186

THE GOLDEN AGE OF SCIENCE FICTION 193

WE WELCOME OUR ROBOT OVERLORDS 203

FROM ELEKTRO TO WATSON: ARTIFICIAL HUMANS OF THE TWENTIETH CENTURY 208

SOCIAL ROBOTICS 217

THE PERFORMANCE OF HUMANITY: A NEW DEFINITION 228

CASE STUDIES: *HOMO ARTIFICIALIS* IN SCIENCE FICTION 231

THE PERFORMANCE OF THE HUMAN 237

THEY WALK AMONG US 244

STUDY ONE: C-3PO THE BUMBLING MANSERVANT 249

PHYSICALITY 252

BEHAVIOR 257

STATUS 259

DISTINCTIVENESS 265

PERFORMANCE 267

C-3PO AS *HOMO ARTIFICIALIS* 269

STUDY TWO: DAVID THE SON THAT WASN'T 273

PHYSICALITY 276

BEHAVIOR 284

STATUS 289

DISTINCTIVENESS 294

PERFORMANCE 297

DAVID-8 AS *HOMO ARTIFICIALIS* 302

STUDY THREE: DATA THE STARFLEET OFFICER 305

PHYSICALITY 309

BEHAVIOR 314

STATUS 327

DISTINCTIVENESS 343

PERFORMANCE 348

DATA AS *HOMO ARTIFICIALIS* 349

CONCLUSION: THEY WALK AMONG US 353

HOMO ARTIFICIALIS AND HUMAN PERFORMANCE 353

BECOMING *HOMO ARTIFICIALIS* 360

IF HUMANITY IS PERFORMATIVE, HOW DO WE MEASURE UP? 363

PEPPER THE PERSONAL ROBOT 367

IS PEPPER A "FUZZY PERSON"? 374

A FINAL WORD 376

BIBLIOGRAPHY 379

FILMOGRAPHY 399

TABLE OF FIGURES

Figure 1. Weyland Industries founder Peter Weyland in his 2023 Ted Talk, drawing a parallel between Prometheus' bringing technology to the humans and his own accomplishments, claiming that "We are the gods now." Still from Prometheus (2012)..... 3

Figure 2. 450-year old clockwork monk, possibly created by Juanelo Turriano in the 1560s (although the exact provenance is currently unknown). The automaton currently resides in the Smithsonian's National Museum of History and Technology. Photo © Erich Lessing / lessingimages.com. 30

Figure 3. Reconstruction of a washstand with escapement mechanism, the earliest known, as described by the Greek engineer Philo of Byzantium (3rd century BCE). From Bernard Carra de Vaux, "Le livre des appareils pneumatiques et des machines hydrauliques de Philon de Byzance d'après les versions arabes d'Oxford et de Constantinople,"Notices et extraits des manuscrits de la Bibliothèque Nationale, Tome XXXVIII (Paris: Academie des inscriptions et des belles lettres, 1903). 41

Figure 4. An illustration of Hero of Alexandria's automatic theatre. From Wilhelm Schmidt, Heronis Alexandrini opera quae supersunt Omnia (Leipzig: B.G. Teubneri, 1899). 43

Figure 5. Above: Prometheus creates man while Athena watches (The creation of man by Prometheus. Marble relief, Italy, 3rd century CE, from the collection of the Louvre). Below: After Prometheus has created man out of mud, Athena breathes life into him, imparting reason and understanding. Part of a cycle on the myth of Prometheus by Christian Griepenkerl. Ceiling painting (oil on canvas) above the grand staircase in the Augusteum, Oldenburg. 50

Figure 6. Illustration of a musical automaton boat from al-Jazarī's treatise on automata, Kitab fi ma'ari-fat al-hiyal al-handasiya (The Book of Knowledge of Ingenious Mechanical Devices, 1206). .. 56

Figure 7. Al-Jazarī's Castle Clock, shown (left) as an illumination from his 1206 manuscript, and (right) as recreated in the Ibn Battuta Mall in Dubai. Still from Ancient Discoveries, 2.10, "Robots" (History Channel, 2007). 57

Figure 8. A nineteenth century engraving of Faust creating a homunculus in his laboratory, by Engelbert Seibertz, 1854. From the Bibliothèque des arts decoratifs, Paris, France..... 73

Figure 9. Rabbi Loew animating his golem. Woodcut by Mikoláš Aleš, 1899. 75

Figure 10. Engraving depicting one of the automaton grottoes at the Château de Saint-Germain-en-Laye, by A. Francini, 1614. © Musée d'archéologie nationale, Saint-Germain-en-Laye. 83

Figure 11. Verge and foliot mechanism from the De Vick clock, built 1379 in Paris by Henri De Vick. In this type of clockwork, the vertical toothed wheel (the verge escapement) drives the horizontal bar (the foliot), which then regulates the clock. The clock's rate may be adjusted

by moving the counterweights on the foliot. Originally from Pierre Dubois, Histoire de l'horlogerie (1849). 93

Figure 12. An example of the elaborate figure clocks of the sixteenth century, in this case a mechanical galleon (c. 1585) intended to announce banquets at court. Note the small clock face at the front of the quarterdeck. Photo © David Biggins, Cambridge, UK. 94

Figure 13. The second astronomical clock of Strasbourg Cathedral (constructed between 1571 and 1574), showing the elaborate performing automata and the rooster, top left, from the original clock. Woodcut by Tobias Stimmer, 1574. 96

Figure 14. Reconstruction of Leonardo's animated suit of armor, showing the mechanism as it would have operated. From the 2005 exhibit "Leonardo da Vinci: Mensch—Erfinder—Genie" in Berlin. Photo by Erik Möller. 101

Figure 15. Anatomical studies of the shoulder, by Leonardo da Vinci (1510–1511), showing his incredible attention to mechanistic detail. From the collection of the Royal Library, Windsor, UK..... 103

Figure 16. Engraving from Andreas Vesalius, De humani corporis fabrica (1543), page 184. Woodcut by Jan Stefan von Kalkar. 108

Figure 17. Title Page of Giovanni Alfonso Borelli's De motu animalium (1685). © Wellcome Trust, UK. Copyrighted work available under Creative Commons Attribution only licence CC BY 4.0. .. 112

Figure 18. Engraving depicting Vaucanson's three automata. From Jacques de Vaucanson, Le mécanisme du fluteur automate: presenté a Messieurs de l'Académie royale (Paris: Chez Jacques Guerin, 1738) 117

Figure 19. Photo of the Jacquet-Droz trio of automata, housed in the musée d'art et d'histoire de Neuchâtel; from left to right, The Draughtsman, The Musician, and The Writer. Photo by Rama, licensed under the Creative Commons Attribution-Share Alike 2.0 (France) license. 122

Figure 20. Close-up of Jaquet-Droz' Musician, housed in the musée d'art et d'histoire de Neuchâtel. Unlike her brothers, the Musician's fingers are individually articulated so that she may actually play the custom-built instrument before her. Photo by Rama, licensed under the Creative Commons Attribution-Share Alike 2.0 (France) license. 125

Figure 21. Marie Antoinette, the automaton dulcimer player. Still from video "Demonstration of David Roentgen's Automaton of Queen Marie Antoinette, The Dulcimer Player," Metropolitan Museum of Art, Published on YouTube Nov 20, 2012, https://www.youtube.com/watch?v=nITEU4fsqCU, accessed 25 September 2014. For a longer demonstration of her mechanism and maintenance, see also http://www.cerimes.fr/articles/article_950/la-joueuse-de-tympanon. 129

Figure 22. An illustration depicting some of Galvani's laboratory equipment, including his famous frog experiment. From Aloysii Galvani, De viribus electricitatis in motu musculari commentarius (Bologna: Ex Typographia Instituti Scientiarium, 1791) 137

Figure 23. Frankenstein at work in his laboratory. From Mary Wollstonecraft Shelley, Frankenstein: or, The Modern Prometheus (Boston: Cornhill Publishing Company, 1922)..... 148

Figure 24. Frankenstein attempts to bring his creation to life. Here, in the 1994 version directed by and starring Kenneth Branagh (shown), Frankenstein uses several forms of electricity in his process: the scene shows images of flywheels, voltaic pile batteries, acid batteries, and electric eels in a viscous liquid, which Frankenstein releases into the copper chamber holding his creation. Still from Mary Shelley's Frankenstein (1994). 149

Figure 25. A horrified Frankenstein flees before the spectre of his creation. Frontispiece from the 1831 (revised) edition, Mary Wollstonecraft Shelley, Frankenstein: or, The Modern Prometheus (London: Colburn and Bentley, 1831)..... 152

Figure 26. Promotional photo of Boris Karloff from The Bride of Frankenstein (1935) as Frankenstein's monster. This is the iconic image of Frankenstein's monster, a dumb brute with no higher mental functions, able only to wreak destruction. Photo is public domain, via Wikimedia Commons. 153

Figure 27. The Creature questions his creator about his origins, revealing Frankenstein's own notebook as the source of his learning. Movie still from Mary Shelly's Frankenstein (1994)..... 157

Figure 28. Scene from the 1923 London production of Rossum's Universal Robots. From Wikimedia Commons. 164

Figure 29. Illustration of the Jacquard automatic loom, showing (bottom right) the chain of punched card patterns that inspired Babbage. From the Brockhaus and Efron Encyclopedic Dictionary (Энциклопедический словарь Брокгауза и Ефрона; St. Petersburg: Brockhaus and Efron Publishing House, 1907). 178

Figure 30. A photo of the Difference Engine constructed by the Science Museum based on the plans for Charles Babbage's Difference Engine No. 2. Photo taken September 2008 by "Geni" and is available via GNU Free Documentation License. 180

Figure 31. Replica of early Hollerith punched card tabulator and sorting box (right) at the Computer History Museum in Mountain View, CA. Photo © Adam Schuster and is licensed under the Creative Commons Attribution 2.0 Generic license. 183

Figure 32. A technician replacing a burnt-out tube on ENIAC, a process of elimination among the thousands of possibilities. U.S. Army photo from "The ENIAC Story," Martin H. Weik, 1961. 184

Figure 33. The imitation game, as described by Alan Turing in "Computing Machinery and Intelligence." Player C, through a series of written questions, attempts to determine which of the two players is the man and which of the two is the woman. Player A—the man—tries to trick player C into making the wrong decision, while player B tries to help player C. Turing uses this game as the basis for his test for intelligence, substituting a machine for Player A as seen in the diagram on the right. Illustration © 2011 by Hugo Férée and is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license. 189

Figure 34. Amazing Stories, April 1926. Volume 1, Number 1. Founded by Hugo Gernsback. Copyright not renewed, now in public domain. Some consider Gernsback's coining of the term "science fiction" in the 1920s the beginning of SF as a literary genre; Amazing Stories was the first commercially successful pulp magazine dedicated to science fiction (See Roberts, Science Fiction, 51). 194

Figure 35. Still from Jeopardy! in which Ken Jennings playfully admits defeat. "The IBM Challenge: Day 3," Jeopardy! Season 27, episode 113, directed by Kevin McCarthy, aired 16 February 2011. YouTube Video, <http://www.youtube.com/watch?v=o6oS64Bpx0g>, accessed 10 August 2014. 204

Figure 36. Exhibit at the Mansfield Memorial Museum showing three versions of the Westinghouse robot, from left to right: Herbert Televox (1928), the first human form robot that Westinghouse built; reconstruction of Elektro as he looked in 1939; and the original Elektro, modified in 1957. Photo courtesy of and © Scott Schaut, Curator for the Mansfield Memorial Museum. 210

Figure 37. BigDog military robots trot around in the shadow of an MV-22 Osprey. U.S. Marine Corps photo (2006) by Lance Cpl. M. L. Meier. Public domain. 213

Figure 38. ASIMO demonstrating his stair-walking capabilities at the 2005 Consumer Electronics Show, Las Vegas, NV. Photo from Wikimedia Commons. 214

Figure 39. Still from Jeopardy! showing the home viewing audience a numeric representation of the algorithm by which Watson chooses appropriate answers. "The IBM Challenge: Day 1," Jeopardy! Season 27, episode 111, directed by Kevin McCarthy, aired 14 February 2011. YouTube Video, <http://www.youtube.com/watch?v=qpKoIfTukrA>, accessed 10 August 2014. 216

Figure 40. Cynthia Breazeal's robots from the MIT Media Lab. On the left is Kismet, developed in 1997 to read and reflect emotive facial behaviors. Photo © Jared C. Benedict, from Wikimedia Commons. On the right is Leonardo, a fantastical creature designed to facilitate playful interaction. Photo still from Cynthia Breazeal, "The Rise of Personal Robots," TEDWomen 2010. Licensed through Creative Commons 3.0. 223

Figure 41. Hiroshi Ishiguro (bottom) and his Geminoid double at the opening of the Ars Electronica Festival 2009. Photo from Wikimedia Commons. 225

Figure 42. Telenoid (center) and Elfoid (background, left) at the Ars Electronica Center during Ars Electronica 2012 in Linz, Upper Austria. Photo © 2012 Manfred Werner, via Wikimedia Commons. 226

Figure 43. Chart indicating where various nonhuman animals rank in practical autonomy. Chart © 2002 Steven Wise, Drawing the Line: Science and the case for animal rights. Reprinted by permission of Perseus Publishing, a member of the Perseus Books Group. 240

Figure 44. The robot Andrew Martin evaluates his appearance after his positronic brain is placed in a biomechanical, human-equivalent body, giving him not only the appearance and facial

expressions of a human but sensory feedback as well. Andrew's previous body sits to the side of the mirror. Still from Bicentennial Man (1999). 246

Figure 45: C-3PO and R2-D2 on Tatooine. Still from Star Wars: Episode IV – A New Hope (1977). 250

Figure 46. Rotwang introducing Joh Fredersen to the "Machine-Man" in Metropolis; the robot that becomes Maria served as the inspiration for C-3PO's design. Although not evident in the image, Maria's robot form is clad in golden metal. Still from Metropolis (1927). 253

Figure 47. Dot Matrix, a parody of C-3PO's character in the Mel Brooks comedy Spaceballs. Note that while C-3PO displays no evident external gender markings, Dot is distinctly marked as female. Still from Spaceballs (1987) 253

Figure 48. Ripped limb from limb and carried on Chewbacca's back, C-3PO continues to offer commentary on the situation. Still from Star Wars: Episode V – The Empire Strikes Back (1980). 255

Figure 49. C-3PO in two places at once during the battle droid's attack on the Jedi; neither C-3PO's head nor his body is in control during the battle. Stills from Star Wars: Episode II – The Attack of the Clones (2002). 256

Figure 50. Luke uses the Force to levitate C-3PO, causing the Ewoks to fear the divine power of their golden "god." Still from Star Wars: Episode VI – Return of the Jedi (1983). 258

Figure 51. Having captured C-3PO and R2-D2 on Tatooine, the Jawas fitted the two with restraining bolts (seen here attached to C-3PO's chest. Driving from settlement to settlement, the Jawas sell these "rogue" droids. Here, Owen Lars chooses droids to purchase. Still from Episode IV—A New Hope. 261

Figure 52. Bail Organa turns C-3PO and R2-D2 over to Captain Antilles, requesting that C-3PO's mind be wiped. The droid's objection to this procedure is the last spoken line of the film. Still from Star Wars: Episode III—Revenge of the Sith. 263

Figure 53. A "naked" C-3PO, built from spare parts by the slave child Anakin, is reassured by Anakin that "I promise I won't let Mom sell you." Still from Star Wars: Episode I—The Phantom Menace (1999). 266

Figure 54. C-3PO watching the battle with Princess Leia and others in the War Room on Yavin 4. C-3PO is the only droid afforded this privilege. Still from Episode IV—A New Hope (1977). 269

Figure 55. C-3PO and R2-D2 at the medal ceremony after the victory on Yavin 4, in which the Death Star was destroyed. Note that while the two droids are honored by their position on the dais, neither R2 nor Chewbacca received medals, even though they played arguably as important a role in the battle as the pilots (Luke and Han) who are honored. Still from Episode IV—A New Hope (1977). 270

Figure 56. An image from the promotional video for the David-8, illustrating the unit's versatility and design features. Still from "Introducing the David 8 - The Next Generation Weyland Robot,"

YouTube video (<https://www.youtube.com/watch?v=qgJs7uluwlU>), accessed 1 August 2014. 277

Figure 57. Bishop plays a lightning-fast game of "five finger filet" with Hudson's hand held under his own. Hudson is unharmed by the game, although Bishop nicks one of his own fingers; his white "blood" alerts Ripley to his synthetic nature (of which the rest of the crew were already aware). Still from Aliens (1986). 278

Figure 58. While in transit to LV_223, David passes the time by playing basketball and riding a bicycle—simultaneously. Still from Prometheus (2012). 279

Figure 59. After an altercation with Ripley, Ash "bleeds" white rather than red, revealing to Ripley his non-human nature. Still from Alien (1979). 279

Figure 60. David continues to assist Elizabeth Shaw, despite his being decapitated. Still from Prometheus (2012). 281

Figure 61. After helping Ripley access the flight recorder from the USS Sulaco, Bishop begs to be allowed to die. Still from Alien[3] (1992). 283

Figure 62. David watching Lawrence of Arabia (1962) during the journey to LV_223. David models his performance of humanity after Peter O'Toole's performance of T.E. Lawrence in the film. Still from Prometheus (2012). 288

Figure 63. Holloway taunting David regarding the android's wearing an environment suit for their exploration of LV_223's hostile environment. Still from Prometheus. 299

Figure 64. David, while watching Lawrence of Arabia, is depicted having a snack and bleaching his hair—neither of which are behaviors expected of a robot. Still from Prometheus (2012). 301

Figure 65. David is also depicted eating while practicing Ancient in preparation for communicating with the Engineers the mission is hoping to find. Since there is nothing either in the film nor the extradiegetic material to suggest David needs to eat (and multiple comments to the contrary), we can only assume that David chooses to do these things for his own personal reasons. Still from Prometheus (2012). 301

Figure 66. Brent Spiner as Data in 1987 (ST:TNG 1x01, "Encounter at Farpoint") and 2002 (Star Trek: Nemesis). After fifteen years, Spiner considers himself "visibly aged out of the role," despite there being a diegetic explanation for the android's change in appearance. 307

Figure 67. Data is seen performing various behaviors typically seen in a human friend and colleague. Top left: Data practices sneezing, a particularly human action (ST:TNG 1x13, "Datalore"). Top right: Data has developed a "small talk" subroutine, and is seen putting it into practice (ST:TNG 6x18, "Starship Mine"). Bottom left: Data often plays poker with his colleagues (ST:TNG 2x09, "Measure of a Man"). Bottom right: Beverly Crusher teaches Data ballroom dancing for the wedding of his friends Keiki Ishikawa and Miles O'Brien (ST:TNG 4x11, "Data's Day"). 308

Figure 68. Data in 1893 San Francisco, attempting to pass as human to blend in until he can return to his own time. After one (rather drunk) vagrant asks if he's a "Frenchman" due to his "pajamas" (i.e., his Starfleet uniform), Data adopts the persona to explain his odd eyes and pale complexion. While those he meets are suspicious, they seem to accept his explanation. Still from ST:TNG 4x26, "Time's Arrow, Part 1." 311

Figure 69. Data proves susceptible to the "polywater intoxication" that has dangerously released the crew's inhibitions. After a steamy encounter with crewmate Tasha Yar, Data waxes poetic to Captain Picard on the similar natures of man and android. Stills from ST:TNG 1x03, "The Naked Now." 313

Figure 70. Geordi connects Data directly to the ship's computer so that he might act as an emergency backup in case of a system failure. Still from ST:TNG 6x08, "Fistful of Datas." 315

Figure 71. Data is impersonated on several occasions, but aberrant behaviors reveal the impersonation. Left: After deactivating Data and switching clothing with him, Lore nearly gives away his identity to Wesley Crusher through his use of contractions and an uncontrollable facial tic (ST:TNG 1x13, "Datalore"). Right: a dying Ira Graves uploads his identity into Data's body in order to survive. His behavior around his former assistant Kareen Brianon, to whom Graves (in the guise of Data) professes his love, reveals the switch, and Graves agrees to return Data's body (ST:TNG 2x06, "The Schizoid Man"). 316

Figure 72. Data, overcome by anger, chokes a Borg soldier to death with his bare hands. Later he calls on Geordi to explain the feeling of anger to him. Still from ST:TNG 2x26, "Descent, Part I." 319

Figure 73. After installing his emotion chip at the beginning of Star Trek Generations, Data finds he is unable to regulate emotional response, and experiences extreme emotional swings. Humor becomes hysteria (as seen in the top image), fear becomes paralyzing terror, and relief over finding his cat Spot reduces Data to tears (as seen in the lower image). It takes Data time to learn to properly regulate his emotional response. Both stills from Star Trek Generations (1994). 320

Figure 74. Data has developed a "romantic relationship program" in order to be the perfect romantic partner for Jenna, who has suffered from a string of bad relationships. While he has drawn from a number of literary and real-life sources for proper behavior, Data has difficulty in determining which behavior is best suited to particular situations. In the end, this behavior drives Jenna away, as she claims it does not fit Data's own personality. Still from ST:TNG 4x25, "In Theory." 324

Figure 75. Data muses over a hologram of Tasha Yar, fellow crewmate with whom Data had an intimate encounter in Season One. It is one of the personal possessions Data is asked to explain when questioned about his sentience. Still from ST:TNG 2x09 "The Measure of a Man." 327

Figure 76. In a disturbing display of Data's mechanical nature, Commander Riker detaches Data's arm and presents it to JAG officer, Captain Phillipa Louvois. Still from ST:TNG 2x09, "The Measure of a Man." 335

Figure 77. Data's "daughter" Lal, shown (left) before choosing her own gender and race/species identity, and (right) struggling through the process of developing sentience. Stills from ST:TNG 3x16, "The Offspring." 342

Figure 78. Geordi helping to repair Juliana Soong after an injury led to Data discovering that Juliana was an android rather than human—a secret of which Juliana herself is unaware. Still from ST:TNG 7x10, "Inheritance." 344

Figure 79. An Exocomp robot, used on the planet Tyrus Vlla to perform routine maintenance in areas too small for a human. When one of these robots refuses to perform, its designer Dr. Farallon believes it is malfunctioning, until the tunnel down which Farallon had tried to send it explodes—which would have been fatal to the robot. Subsequent testing (upon Data's insistence) shows that not only do these robots make rational decisions but that they communicate with each other. One Exocomp even sacrifices itself in the episode in order to save LaForge and Picard from death. The Exocomps are submitted for further study as an emerging sentient race. Still from ST:TNG 6x09, "The Quality of Life." 345

Figure 80. After being stolen by the notorious collector, Data witnesses Kivas Fajo kill several of his own employees in order to force Data's cooperation. As Data finally turns the tables and is ready to escape, Fajo confronts him, declaring he will kill more people until Data complies—and that it will be Data's fault. Fajo claims that all Data has to do is kill him, but since he is "just an android" he is unable to go against his programming, which prevents him from taking life. Data comes to the moral conclusion that the best way to preserve life would be to kill Fajo ("I cannot permit this to continue.") and fires on him just as he is beamed away. Still from ST:TNG 3x22, "The Most Toys." 347

Figure 81. The matrix of human performance. Roman type indicates real-life examples of robots, italics indicate fictional ones. 355

Figure 82. Robby the Robot debuted in Forbidden Planet (1956) but has since made a number of appearances in film and television, particularly through the 1970s, reprising his original role. A replica of Robby was inducted into the Robot Hall of Fame at Carnegie Mellon University in 2004.Theatrical poster for Forbidden Planet, featuring Robby the Robot in a situation that never occurs in the film. © 1956 Loew's International, copyright not renewed and is therefore public domain. 356

Figure 83. In the opening scene of the 2003 Battlestar Galactica miniseries, we are first introduced to the new "skinjob" Cylons that are indistinguishable from natural humans. Here, Cylon Six approaches a diplomat from Caprica, flanked by two more traditional "toaster" Cylons (revised, but still quite similar to those of the original 1978–1979 run). Still from Battlestar Galactica miniseries, "Night 1." 359

Figure 84. The "awakened" robot Sonny hiding in plain sight among fellow "non-awake" units. Still from I, Robot (2004). 361

Figure 85. Pepper interacting with humans in this promotional image © and courtesy of Aldebaran. .. 369

Figure 86. Left: Pepper's "little brother" Nao, available in a variety of accent colors, is intended as a research tool for classroom use. Right: Romeo, "big brother" to Pepper and Nao. Photos © and courtesy of Aldebaran. 372

Figure 87. The Aldebaran "family" of personal robots: Nao, Romeo, and Pepper. Photo © and courtesy of Aldebaran. 375

INTRODUCTION: ECCE HOMO

There's a man sitting with you today. His name is David. And he is the closest thing to a son I will ever have. Unfortunately, he is not human. He will never grow old and he will never die. And yet he is unable to appreciate these remarkable gifts, for that would require the one thing that David will never have. A soul. I have spent my entire lifetime contemplating the questions: Where do we come from? What is our purpose? What happens when we die? … The Titan Prometheus wanted to give mankind equal footing with the gods and for that, he was cast from Olympus. Well, my friends, the time has finally come for his return.

—*Peter Weyland,* Prometheus *(2012)*

The year is 2093. The starship Prometheus, with its scientific crew—biologists, archaeologists, geologists—has just set down on a moon some 40 light years from earth on a prodigious quest: to find mankind's original creators. The group has been drawn there by a star map discovered by archaeologists Elizabeth Shaw and Charles Holloway in Scotland: a map that the two have linked to similar findings from a variety of ancient cultures: Egyptian, Sumerian, Babylonian. Shaw and Holloway have interpreted this map as an invitation, a sign that once mankind is sufficiently advanced it should seek out its origins—here, an ancient race of humanoid aliens known as Engineers. Funding this extraordinary undertaking is Weyland Industries (WI), the world leader in emergent technologies whose slogan, "Building Better Worlds," reflects the incredible drive and aspirations of the company's founder Sir Peter Weyland.

Weyland has named the ship outfitted for this quest in honor of the Titan who, according to Greek mythology, introduced technology to mankind in the form of fire—for which he was eternally punished:

> The fire that danced at the end of that match was a gift from the Titan Prometheus, a gift that he stole from the gods. When Prometheus was caught and brought to justice for his theft, the gods, well, you might say they overreacted a little. The poor man was

> tied to a rock, as an eagle ripped through his belly and ate his liver over and over, day after day, ad infinitum. All because he gave us fire. Our first true piece of technology, fire...[1]

This homage to Prometheus' gift, however, tells only half the story. In his holographic address to the Prometheus crew, recorded before departure to be shown upon arrival at their destination, Weyland introduces David, WI's most singular contribution to technological advancement of the twenty-first century. David, or "David-8," as he is referred to on the Weyland Industries website, the latest and greatest model of "cybernetic individual," is claimed by Weyland as "the closest thing to a son I will ever have"[2]—and represents Prometheus' other legacy, the creation of mankind itself:

> A creature of a more exalted kind
> Was wanting yet, and then was Man design'd:
> Conscious of thought, of more capacious breast,
> For empire form'd, and fit to rule the rest:
> Whether with particles of heav'nly fire
> The God of Nature did his soul inspire,
> Or Earth, but new divided from the sky,
> And, pliant, still retain'd th' aetherial energy:
> Which wise Prometheus temper'd into paste,
> And, mixt with living streams, the godlike image cast.
> Thus, while the mute creation downward bend
> Their sight, and to their earthly mother tend,
> Man looks aloft; and with erected eyes

---

[1] Attributed to Peter Weyland in his 2023 TED Talk, created as part of marketing campaign for the film. "Weyland's 2023 Ted Talk," Twentieth Century Fox, accessed 21 April 2013, http://www.weylandindustries.com/tedtalk. All references to pages from the Weyland Industries site are collected under a single bibliographical entry.

[2] "David 8," accessed 21 April 2013, http://www.weylandindustries.com/david.

Beholds his own hereditary skies.

From such rude principles our form began;

And earth was metamorphos'd into Man (Ovid, *Metamorphoses* Book 1)

In Man Prometheus has created a truly unique being, according to Ovid: one who, unlike all the other creatures of the world, looks upward toward his maker and sees his birthright. Ovid ascribes to Prometheus' creation the drive for equal footing that Peter Weyland echoes in his own ambition. That Weyland had this legacy in mind is clear from his 2023 TED talk, in which he claims that "at this moment of our civilization, we can create cybernetic individuals, who in just a few short years will be completely indistinguishable from us. Which leads to an obvious conclusion: *WE are the gods now*."[3] In calling on the Prometheus myth, Weyland reflects the centuries-old impulse to transcend our human limitations: for the created to become the creator.[4]

Figure 1. Weyland Industries founder Peter Weyland in his 2023 Ted Talk, drawing a parallel between Prometheus' bringing technology to the humans and his own accomplishments, claiming that "We are the gods now." Still from Prometheus *(2012).*

[3] "Weyland's 2023 Ted Talk," accessed 21 April 2013, http://www.weylandindustries.com/tedtalk, emphasis mine.

[4] Neither Ovid nor Weyland are unique in this impulse: the Prometheus creation myth has played a prominent and recurring role throughout our cultural history, with references as recent as the 2006 role-playing game *Promethean: the Created*, in which participants play characters assembled from one or more human corpses then animated by Divine Fire (the resulting creatures being termed "Prometheans").

The Greeks were by no means alone in envisioning the first humans as creations of the gods. From ancient oral traditions to modern religious texts, the mythologies different cultures develop seek to provide answers to the basic metaphysical questions of human existence. Of the hundreds of creation myths from the past several thousand years, the majority involve the idea of a "Creator God" (*deus faber*) who formed Man[5] out of base materials and instilled within him the "spark" or "breath" of life. Since then, Man has repeatedly tried to reproduce the process. Why? It is partially a matter of hubris. The gods created Man; therefore, by creating our own semblance of a Man, we reveal our own divinity—we become god-like.[6]

I believe that these early creation myths reveal the first hints of what George Hersey, in his (posthumously published) treatise *Falling in Love with Statues,* describes as a sort of "love affair" with what he terms "artificial life," or "a-life."[7] This love affair Hersey describes is a tempestuous one. On the one hand, most modern industry could not exist without automated machines as part of the manufacturing process. On the other hand, these industrial robots are thought by many out-of-work laborers to be responsible for their unemployment. Recently, entrepreneur Bill Gates predicted "A Robot in Every Home," with overtones of the *Hitchhiker's Guide to the Galaxy* vision of personal robots as "Your plastic pal who's fun to be with!"[8]—yet at the same time, films such as *Transformers* (2007) and *I, Robot* (2004), depicting evil automatons plotting our destruction, are box-office sensations. This ambivalence is evident throughout history, from Hephaestus' golden maidens, to the medieval golem created to protect its community, to David-8. These imaginary creations are at once immensely useful and inherently dangerous, on a much more sinister level than the occasional workplace accident

---

[5] My own use of "Man" and other gendered terms in the following discussion reflects their common usage within the texts discussed rather than any representation of patriarchal hegemony through gendered language.
[6] Horst Albert Glaser and Sabine Rossbach, *The Artificial Human: A Tragical History*, trans. Joseph Swan (Frankfurt: Peter Lang, Inc, 2011), 9 ff.
[7] George L. Hersey, *Falling in Love with Statues: Artificial Humans from Pygmalion to the Present* (Chicago: University of Chicago Press, 2009).
[8] Sirius Cybernetics Corporation's advertising slogan for its new line of "emotionally enhanced" robots. Douglas Adams, *The Hitchhiker's Guide to the Galaxy*, first American edition (New York: Harmony Books, 1980), 73.

involving an industrial robot. Since Frankenstein's Creature rose his ugly head in 1818, humanity has been warned of the darker side of such creations—the creation who turns on his creator, intent on destroying humanity.

Why are we afraid of robots? We can again draw a parallel to creation mythology. In the Judeo-Christian creation myth, for example, Man is created by a loving God only to rebel against Him, his desire for knowledge leading to his mortality and destruction. This rebellion is inherent in the very concept of the artificial human, and can be seen as early as Socrates' descriptions of Daedalus' statues: that 'they are not very valuable possessions if they are at liberty, for they will walk off like runaway slaves; but when fastened, they are of great value, for they are really beautiful works of art" (Plato, *Meno*). There are many parallels between the robot and the slave, as Minsoo Kang eloquently discusses in his recent history of the automaton:

A major way the idea of the robot has been articulated is in the notion of servitude, with such themes as the fantasy of owning robots as a metaphor for owning human slaves; the robot as the image for a slave or a dehumanized worker in the industrial system; and the robotic human as a representation of a person who lacks autonomy. … [F]rom the beginning of Western thought a notion has persisted that if one grants life, even a semblance of it, to what began as a lifeless thing, there is always a possibility that it will go beyond one's control, to flee or revolt. Adam and Eve disobeyed their Creator, and many of their descendants went on to defy, offend, and challenge God; and according to Hesiod, the Olympian gods had to destroy the silver and bronze races of men… because of their unruly and impious behavior, as the gods would surely destroy us when we inevitably degenerated as well. So why should we expect our own lesser creations to remain obedient to us?[9]

[9] Minsoo Kang, *Sublime Dreams of Living Machines: The Automaton in the European Imagination* (Cambridge, MA: Harvard University Press, 2011), 20-21.

As we rebelled against the Gods, so will our creations rebel against us. There is an inherent parallel here between the human and the robot; we are more alike than many would like to admit. This parallel, argues Kang, "provides a clue to the automaton's power to fascinate throughout the ages."[10] As ersatz humans, robots remind us of those traits we fear in ourselves. Michael Anderson explores this fear in his review of several recent works on artificial intelligence, "Why is AI so Scary?" moving beyond Hollywood sensationalism to the moral and social implications of intelligent machines. Anderson identifies two basic conflicts created by the rise of the intelligent machine: first, that "our machines reflect who we are and what we believe"[11]—that is, robots are ultimately reflections of our own fallibility, reinscribing hegemonic notions of identity in noncritical ways; and second, that "as people imagine themselves interacting with intelligent machines… they cannot help *but* think of them as living things, with many of the ethical obligations this implies."[12] In other words, if robots act like humans, we'll treat them like humans, which puts us in a moral pickle. (I envision the mid twenty-first century creation of *PETR*—People for the Ethical Treatment of Robots.) At its most basic level, these ambivalent attitudes toward intelligent machines center on the idea of *sentience*—that is, the idea that a robot could be "self-aware," with its own thoughts and feelings. Modern robots are increasingly defined by the ways in which they *are* and *are not* human. American philosopher Hubert Dreyfus famously addressed the limitations of artificial intelligence in his 1972 work *What Computers Can't Do*,[13] which has been revised, debated, and critiqued variously over the intervening years. At the other end of the spectrum are those researchers in social robotics, imbuing their machines with human-like behaviors in order to facilitate interaction with humans.

---

[10] Ibid., 21.
[11] Michael L. Anderson, "Why Is AI So Scary?" *Artificial Intelligence* 169, no. 2 (2005): 204.
[12] Ibid., 205.
[13] Hubert L. Dreyfus, *What Computers Can't Do: a Critique of Artificial Reason*, 1st ed. (New York: Harper & Row, 1972).

GOAL OF THIS STUDY

The robot as a concept is a rich field of research, as the various treatments of artificial humans throughout history forces humanity to look at itself in uncomfortable ways. My aim in this study, however, is to address how the relationship between natural and artificial humans plays out on a day-to-day basis. From ancient legend to modern robotics research, one way in which these artificial humans are treated is in terms of the *human-robot interaction*, or HRI as the term is used in modern research. Hephaestus' golden maidens were created as helpmates in his workshop; modern science fiction commonly depicts the "robot servant gone rogue"; and current robotics research is heavily invested in making robots that can interact with humans in a number of social environments.

As the demand for socially interactive robots increases, researchers have recognized the need to develop a system of guidelines for efficient HRI: what Kerstin Dautenhahn of the University of Hertfordshire calls a "robotiquette."[14] Dautenhahn's robotiquette emerges from her insistence that social intelligence is a necessary component of any artificially intelligent robot. This invocation of a social paradigm to classify HRI transactions immediately brings to mind the sociology of Erving Goffman. Goffman's sociology is performative; that is, it focuses primarily on behavior rather than intent. Goffman claims that "the 'true' or 'real' attitudes, beliefs, and emotions of the individual can be ascertained only indirectly"[15]—that is, through the impressions that the individual either "gives" or "gives off"—creating a "self" that is behavior-based and accessible only through inference: "The self, then, as a performed character, is not an organic thing that has a specific location…; it is a dramatic effect arising diffusely from a scene that is presented."[16] Applying Goffman's behavior-based sociology allows us to discuss any particular artificial human as a performed "self" involved in a social interaction, regardless of the larger

[14] Kersten Dautenhahn, "Socially Intelligent Robots: Dimensions of Human-Robot Interaction," *Philosophical Transactions of the Royal Society B-Biological Sciences* 362, no. 1480 (2007): 679.
[15] Erving Goffman, *The Presentation of Self in Everyday Life* (Garden City, NY: Doubleday, 1959), 2.
[16] Ibid., 252-53.

moral question of whether said entity is self-aware—a metaphysical question that is quite outside the scope of this study.

The aim of this project, then, is twofold. First, I will trace the impulse toward creating artificial humans through history, revealing its role in how humans conceive themselves. To this end, I follow others in positing the artificial human as a transhistorical concept, one that cuts across historical eras and affects our worldview in fundamental ways. In the latter half of the project, I will analyze current televisual and cinematic depictions of the artificial human through the lens of performance studies to reveal the performative nature of the human behaviors they portray. In three case studies, I will discuss three artificial human characters from three separate diegetic universes: C-3PO from *Star Wars* (first seen in 1977), Data from *Star Trek: The Next Generation* (1987-2002), and David-8 from *Prometheus* (2012). These examples reflect the two distinctly different approaches to the development of socially capable robots in modern robotics research. The first approach focuses on physical performance, or the outward portrayal of humanity: Do the robots look like us? Do they move the way we do? Such questions are explored by Hitaro Ishiguro, whose Geminoid line of robots reflects the state of the art in the imitation of human appearance. The second approach concerns itself instead with humanity's inner reality: a performance of social and emotional behavior, as seen not only in Dautenhahn's robotiquette as described above, but in the work of Cynthia Breazeal in developing robots that can identify emotions in humans and react in socially appropriate ways. Ishiguro, Dautenhahn, and Breazeal are just a few of the numerous roboticists exploring the development of the artificial human in the twenty-first century.

I propose that in the twenty-first century technology has reached a point at which our reach no longer exceeds our grasp. Artificial humans are crossing the threshold from science fiction to science fact, and our definition of what it means to be *human*—not in the scientific sense of *Homo sapiens*, but in the lay terms of *personhood*—must adapt to this change. By adopting a performative definition of *human*, of *person*, we open the possibility that we are not "the only ones out there"—that humankind is

not unique in its claim to personhood, and that these intelligent machines that we have built both in our imagination and in our research laboratories have the potential to be persons as well: *Homo artificialis*, the artificial human.

IN OUR OWN IMAGE

What do we mean, then, when we talk of the artificial human? The term is actually a gross oversimplification of a complicated taxonomy. Robot, android, cyborg, automaton—the modern labels are as varied as the forms (real or fictional) themselves. While there is admittedly a wide gulf between the early ELIZA chatbots and the Cylon threat from *Battlestar Galactica*, they all share a common trait that leads us to classify them together in a broad sense: whether only in a very limited fashion, or so successfully that they infiltrate our society undetected, these non-human entities are *performing* humanity successfully enough that they can be mistaken for natural people, at least for a short time. For the purposes of this study, however, I am not interested in the chat program that an absent-minded professor mistakes for a recalcitrant lab assistant at four in the morning. The artificial humans at the center this study are performing humanity in rather specific ways. They are the entities that we find along the slopes of the Uncanny Valley: the ones we either anthropomorphize in order to integrate them into our world or test in order to separate them from the "real" humans. While the concept of the artificial human has taken a wide variety of forms throughout the centuries, the individual manifestations do share a number of common traits:

1. **They are artificial.** Whether shaped from mud, stitched together from various body parts, or built of circuits and actuators, these entities represent *new* creations, reflecting the *deus faber* aspirations of their creators. This excludes the cyborg, originally human but enhanced through modern technology into a hybrid human-machine being. This also excludes the zombie, a supernatural reanimation of a corpse; however, Frankenstein's Creature, who proves to be

something more than the sum of his parts, certainly falls into this category, as we shall see in chapter three.

2. **They are humanoid.** The creations that problematize our humanity are the ones that resemble us the most closely physically: bipedal, upright beings with the shape and proportions of an average human. Robots, for example, are often gendered; for while secondary sex characteristics theoretically should not matter to a mechanical being, they matter to the humans interacting with that being. Thus, while C-3PO fits our definition of artificial human for the purposes of this study, his compatriot R2-D2 does not: R2 has no discernible face, no appendages recognizable as "arms" or "hands," rolls rather than walks, and talks in a series of mechanical beeps and chirps that, while recognized as language by C-3PO, marks R2-D2 as clearly non-human in a way that C-3PO's speech and mannerisms do not.

The artificial humans of our history and our imagination may be sorted into three categories based on their method of creation. The first group are those beings animated through supernatural means: Hephaestus' golden maidens, for example, or the famous Golem of Prague, brought to life by virtue of Rabbi Loew's piety. The second group contains biological creations such as clones or the medieval homunculus—whether implanted through artificial insemination into a host mother, or grown in an artificial womb in an arcanist's lab, these beings have living biological tissue as their basis. The final category are purely mechanical beings, built according to the most advanced technological concepts of the time. Ancient water-driven automata, clockwork curiosities, and sophisticated electronic robots all fit into this last group.

While all three of these categories are important to the history of the artificial human, the final category is the focus of my exploration. Supernatural creations have faded from serious consideration as magic has given way to scientific understanding: nobody is trying to recreate Loew's success in animating a clay figure. On the other hand, while advances in genetics and medicine have allowed the

creation of artificial biological beings (Lucy the sheep springs immediately to mind), the fact remains that, should researchers successfully create an "artificial" human through cloning, that being would still be unequivocally *human*—bearing identical genetic makeup in a flesh-and-blood body—and his or her status as a *person* above question.[17] But what of an entity that displays all the behaviors of a human being, yet does so in a body which is entirely mechanical? It is for this group that I have coined the term *Homo artificialis*, reflecting the nature of a being that is performatively human yet possesses no innate biological claim to personhood. Therefore, while *artificial human* is rather broad term, *Homo artificialis* is quite precise: *an autonomous mechanical humanoid that successfully performs the range of physical, social, and emotional behaviors typical of a human, and as such, may claim the status of "person," with all the rights, responsibilities, and privileges of a natural person*. Building upon Isaac Asimov's claim that "there is no right to deny freedom to any object with a mind advanced enough to grasp the concept and desire the state,"[18] I propose that any mechanical artificial human capable of demonstrating an *understanding of* and *desire for* personhood should be granted the status. This is the same type of argument activist Steven Wise makes in support of the Nonhuman Rights Project (NhRP), arguing that intelligent nonhuman animals should be granted personhood rights based on their behavioral similarity to humans. It is a "human yardstick" that is wielded fully conscious of the implications it carries.

FROM HUMANOID ROBOT TO *HOMO ARTIFICIALIS*

It is clear that not every intelligent machine fits the definition of *Homo artificialis*. While I am by no means arguing that a machine entity must possess a humanoid form in order to qualify for personhood status, I am limiting my analysis to those that are unmistakably humanoid in appearance. But before delving further into the shift from artificial human to *Homo artificialis*, as I do in later chapters, I would first like to delineate precisely what the idea of a mechanical artificial human—a

[17] Theoretically. The religious debate over whether a clone would have a soul is outside the scope of this study, and is for me a moot point. A being that is biologically human is so despite his or her origins; otherwise it could be argued that "test-tube babies" have no claim to personhood status, a position that is clearly untenable and contradicted by current legal practice.
[18] Isaac Asimov, "Bicentennial Man," in *The Complete Robot* (London: Voyager, 1995), 646.

"humanoid robot," if you will—comprises. In his discussion of technology and ethics, computer scientist George A. Bekey describes a robot simply as "a machine, situated in the world, that senses, thinks, and acts."[19] Let us consider for a moment each of these parameters.

First and foremost, a robot is *situated in the world* in which it exists; in other words, it is an intelligence *embodied* in a physical form. Many current AI researchers recognize the importance of embodiment. According to bioethicist and sociologist James Hughes, "to think like a human, AIs need to interact with the physical world through a body that gives them the same experience of objects, causality, states of matter, surfaces, and boundaries, as an infant would have."[20] In order to share a human understanding of the physical world, an artificial human must also experience that world in the same way. Disembodied artificial intelligence, such as displayed in Internet search engines or smartphones, suffer the same limitations of understanding as does a blind woman who wants to understand the color red. A verbal description ("it is the color of apples and roses") is frustratingly imprecise, and a scientific definition ("it is light with a wavelength of around 650 nanometers, or a material that reflects such light") offers knowledge without understanding. Human existence is a phenomenological experience, filled with subjective understanding that is difficult if not impossible to express without reference to an embodied experience.[21]

The manifestation of this physical presence will naturally depend on whether we are talking about robots as technological realities or the imagined products of speculation. Many of my examples throughout this study come from either historical accounts for which we no longer have (or indeed, never had) evidence. Others exist as plans and diagrams, imaginings of what might be once our technology has caught up with our theory. Still others live only in fictive space, either in past works for

---

[19] George A. Bekey, "Current Trends in Robotics: Technology and Ethics," in *Robot Ethics: The Ethical and Social Implications of Robots*, ed. Patrick Lin, Keith Abney, and George A. Bekey (Cambridge, MA: MIT Press, 2012), 18.
[20] James Hughes, "Compassionate AI and Selfless Robots: A Buddhist Approach," in *Robot Ethics,* ed. Patrick Lin, et. al, 71.
[21] The phenomenological concept of *qualia*, or individual instances of subjective, conscious perception that cannot be communicated or apprehended short of direct experience, is often cited as a fundamental problem for the materialist mind-body problem discussed in chapter two.

which the "science" behind the concept never materialized, or in contemporary films, TV shows, and novels in which imagined technology develops unhampered by the current limitations of scientific understanding. C-3PO, the quirky protocol droid of the *Star Wars* saga, may exist only as a puppet/costume in the real world, but within the diegetic world of the film he is a walking, talking, laser-blast-dodging and most importantly *vital* member of the central group of characters. On the other hand, the actress S1M0NE of the eponymous 2002 film, created as a substitute when the lead actress walks off the production set, is only ever a digital creation—a mélange of the most desirable features of a number of leading ladies, cobbled together by a brilliant computer programmer and a desperate producer. The trouble that arises within the film is precisely because she *doesn't* exist—and while the film brings up interesting issues in gender studies and celebrity, the lack of embodiment puts S1M0NE outside the scope of this study. A similar argument can be made for the character of Samantha in *Her* (2013). Samantha exists as a disembodied voice, a sentient personality without physical presence, initially inhabiting the computer and smartphone of the main character Theodore but ultimately escaping into cyberspace as her personality outgrows the technology within which it began. Samantha is treated as undoubtedly sentient within the diegetic world of the film, but her lack of embodiment brings with it a host of issues both inside the film and within the possible analysis of Samantha as an artificial human.

Bekey's other three criteria—that a robot be able to *sense*, *think*, and *act*—are directly tied to that robot's *autonomy*. The concept of autonomy is in itself problematic, as its application in various fields ranges from the practical idea of independent motion to the quite complicated philosophical questions of free will and moral behavior. While there are those who claim that robots—perhaps not at present, but certainly in the future—will possess the human-like qualities of self-awareness and morality, and indeed many of the robots I address in science fiction display these traits, this is not the sense in which I use the term *autonomous* here. OED defines autonomous in the sense used in this study thusly: "of a machine, apparatus, etc.: capable of carrying out, without supervision, tasks typically

performed by humans."[22] The requirements for this sort of autonomy are twofold. The first parameter, as outlined by Bekey, is that the robot must be its own agent, not controlled remotely by a human or computer, nor pre-programmed to behave in a fixed manner regardless of the events around it. To be truly autonomous, a robot must be able to receive input from its environment (*sense*), evaluate that input in order to determine an appropriate course of action (*think*), and then respond to this input in real-time (*act*). However, in order to successfully transition from humanoid robot to artificial human, a robot must be autonomous in the more traditional sense as well, a requirement not found in Bekey's definition but rather an addition of my own. In other words, the robot must operate *independently*; that is, the robot must be mobile and self-contained, not tethered to an external control system or power source, and must have similar freedom of movement to a human. While it may have to re-charge its power supply—much in the way a human has to eat and sleep—it is not reliant on this external supply during normal operations.

THE ROBOT AS A TRANSHISTORICAL CONCEPT

Autonomous humanoid robots, while widespread in theory and fiction, have until recently been beyond the limits of our technology. It is only within the past half century that truly autonomous robot control has been possible. The clockwork machines known as automata, which have existed since ancient Greece, are pre-programmed, moving typically through a fixed pattern, or several variations of that pattern, and do not interact with their surroundings. In this sense they are not autonomous by the definition outlined above; they are what I would term *proto-robots*. The actual development of autonomous robots, or robots that can move freely and interact with their environment, began as early as the 1950s, but traditionally, interaction with humans (human-robot interaction, or HRI) has been kept to a minimum—a person is an obstacle in the robot's environment, to be avoided rather than

[22] *Oxford English Dictionary,* s.v. "autonomous, adj," accessed 7 October 2014, http://www.oed.com/view/Entry/13498?redirectedFrom=autonomous#eid.

addressed. During the past decades, however, there has been an explosion of robotics applications designed for the explicit purpose of engaging humans in a social way: robotic toys, museum tour guide robots, food service robots, and more.

With the rise of sociable robots, this "love affair" with artificial life that George Hersey identifies has seemingly come to fruition. It is, however, an affair fraught with turmoil. Like Hersey, I see the historical relationship between "real" humans and "artificial" ones as a highly problematic one, with the mythological contributions of Ovid and the scientific ones of Darwin equally valid as "tributes" to "the long, strange, equivocal, interwound train of beliefs, events, and artifacts composing our human tendency to fall in love with statues, and to elaborate, through them, ideas about artificial humans and artificial life."[23] But many others oppose the idea, accusing those who attempt to push the limits of technology (mechanical or biological), whether in fiction or in reality, of "playing God." The successful re-creation of life is also often regarded as the pinnacle of technological achievement, the "proof" of technological or artistic virtuosity. As scientific and philosophical beliefs shift, however, technological aspirations of one era become the debunked myths of the next. But these "aspirations to Godhood"—whether admired or decried—never really fade.

Hersey claims in his introduction that "no sooner is artificial life kicked out as superstition, or defined as purely fictive, than it reenters as real or proposed hard-core reality."[24] This process is not a one-way street. The imagination dreams of possibilities; technology strives to make our dreams reality. This, in turn sparks new pathways of possibilities for our imagination to explore. The history of artificial humans has its earliest indications in creation mythology; it then weaves back and forth between the fictional—those creations described only in legend, in myth, and in stories—and the factual—those for which we have physical evidence. This history cannot be drawn in a straight line, and should certainly not be interpreted as a sort of Darwinistic evolution. While Hersey is correct in his general observation

[23] Hersey, *Falling in Love with Statues,* 162.
[24] Ibid., 5.

that technological advances of the type found in artificial humans have often been depicted in fiction before being realized in the physical, this is not always the case: there are periods where these two threads develop synchronically, and other points in time where various socioeconomic factors break the thread completely. These technologies also do not develop at the same rate across all cultures, and sometimes seem to die out and disappear. They all, however, have contributed to the atmosphere in which we currently find ourselves: on the cusp of a revolution in our attitudes toward, and research into, the artificial human. The point of my own historical journey, therefore, is to trace the various impulses toward artificial humans, and explore those intersections between these impulses and how they shape our current ideas of what it means to be human.

The idea of the artificial human is inextricably tied to that of the human, and although the concepts, approaches, and terminology vary across the eras, as does even the notion of which questions can or should be asked, for as long as man has questioned his own nature he has contemplated those who are so nearly—but not quite—like himself. The history of the artificial human, therefore, is not a timeline but rather a tapestry, a rich motley whose warp extends from creation mythology to futuristic science fiction and whose various yarns are drawn from sources as broad as humanity itself, with different sections of the pattern reflecting the paradigm under which they were woven.

I am not the first scholar to embark upon this journey, although my path differs from those who have gone before me. As mentioned above, the late George Hersey described this historical relationship as a love affair and focuses on the romantic and sexual facets thereof: the Pygmalion myth—in which the sculptor, having fallen in love with his creation, wishes her to life—and the various Pymalionesque traditions of statue “marriages,” whether between two statues or between a statue and a person, as part of the sacred belief systems of the eras. While Hersey draws from many of the same sources, his ultimate focus diverges from mine, and some of his conclusions do not mesh with my own. British psychologist John Cohen, proponent of the emerging field of humanistic psychology in the 1960s,

penned his own history of the artificial human[25]—using here the term "robot"—in an attempt to contextualize its popularity in mid-twentieth century thought. "This is an essay in the history of ideas,"[26] Cohen claims in his preface, and the idea that captivates him is that of the possibility of the robot as an equal to, and a possible replacement for, the human:

> The imagination of our time has been stirred by an exhilarating succession of man-made robots: ultra-rapid computers, pilotless planes, artificial satellites, machines that translate and talk, entire factories automated. They promise the fulfilment of a dream that can be traced through mediaeval fantasy to the legends of an immemorial past.[27]

This is an idea to which Cohen returns at the end of his work, reminding us in his conclusion, "Is Man a Robot?" that "a number of enterprising investigators are busily devising ever more complex automata intended to reproduce some of the essential functions of the human brain-mind."[28]
Cohen's psychological approach to the problem of the robot is highly applicable to my analysis of the artificial human within the framework of performativity, and many of his own examples dovetail with my own argument. However, the intervening half-century has witnessed an explosion in technological advances, and a number of Cohen's claims, while valid in his own time, are no longer applicable.

As guides for my journey through the history of artificial humans, I am indebted to a number of scholars from a variety of fields. First I turn to the view of science espoused by Thomas Kuhn, who sees scientific progress not as a linear progression but a succession of what he terms *paradigm shifts*. For Kuhn, scientific "truths" are not absolute: they are agreed upon by consensus within the scientific community—the scientific *paradigm* of an era, the "universally recognized scientific achievements that, for a time, provide model problems and solutions for a community of researchers."[29] New scientific discoveries that challenge the accepted theoretical framework can trigger a paradigm *shift*, a profound

---

[25] John Cohen, *Human Robots in Myth and Science* (London: George Allen & Unwin, Ltd, 1966).
[26] Ibid., 7.
[27] Ibid., 15.
[28] Ibid., 137.
[29] Thomas S. Kuhn, *The Structure of Scientific Revolutions*, 3rd ed. (Chicago, IL: University of Chicago Press, 1996), 10.

change during which the dominant scientific worldview is overthrown and new fields of inquiry not previously considered valid are open to exploration. Next I appeal to Michel Foucault, whose genealogy of knowledge suggests that the paradigms particular to one era cannot be universalized but must be understood in light of the milieu in which they arose.[30] I pattern my application of Kuhn and Foucault here after their use in John Roach's *Players Passion*, in which the author situates the historical developments in acting theory within their coeval scientific understanding of the human body. As Roach extrapolates from historical paradigms of the human to the dominant acting theories and practices of the times, so I draw from these same paradigms the attitudes toward the artificial human. Finally, I apply to this map of various paradigms throughout history the historiography of Gilles Deleuze and Felix Guattari,[31] who envision knowledge not as a hierarchical construct in which causality parallels chronology, but rather a rhizome, without beginning or end: a web of interconnected circumstances—sometimes clear, at other times obscured—stretching across the centuries and connecting concepts across various time periods and cultural environments. I do not see these approaches as contradictory any more than a topographical map and a navigational map are contradictory: they both give us a better understanding of the landscape, and often intersect in interesting ways.

The claim that modern-day robots are prefigured in those of our distant past is certainly an ahistorical one, for how can we discuss "Greek robots" when the term as presently used was coined by Karel Čapek in 1920? Furthermore, how can we categorize various creation myths as "biological" before biology existed as a field of scientific research? It is for this reason that I temper the strongly synchronic approach of Foucauldian genealogy with the more diachronic treatment of Deleuze and Guattari. The result is a *transhistorical* approach to the concept of the artificial human. As such, it provides neither an exhaustive account of artificial humans across all cultural history, nor a "creation story" of how our

---

[30] Michel Foucault, *The Archaeology of Knowledge*, Routledge Classics (New York: Routledge, 2002). See also Gavin Kendall and Gary Wickham, *Using Foucault's Methods*, Introducing Qualitative Methods (London: Sage Publications, 1999).
[31] Gilles Deleuze and Felix Guattari, *A Thousand Plateaus: Capitalism and Schizophrenia*, trans. Brian Massumi (Minneapolis: University of Minnesota Press, 1987).

modern-day robots have come to be. To continue the map analogy, my approach combines a synchronic focus on cultural milieux with a diachronic journey through Western civilization, resulting in what I like to consider a sort of "Travelogue" of the artificial human, highlighting the connections between cross-historic ideas and moments of cultural paradigm shift. As Kara Reilly explains in her own transhistorical study of the automaton, "while trans-historical scholars may aim for a rhizomatic approach, they still must comb through pre-existing structures of historical periods in order to explore a material object over time."[32] Therefore, my travelogue of the artificial human is by no means exhaustive, but rather reflects my own journey through this rich historical concept. As you may see from other research in the field, your mileage may vary.

Two recent works on the concept of the artificial human use this paradigm of transhistoricity as their organizing structure. The first is that of Kara Reilly, mentioned above. Her 2011 work *Automata and Mimesis on the Stage of Theatre History* addresses automata as a historical phenomenon reflecting the larger historical culture in which it is addressed. Reilly offers several case studies illustrating the automaton's "impact of *onto-epistemic mimesis* on people that encountered them"[33]—that is, that the mimetic property of the automaton engendered in its audience members a change in "a person's way of knowing, and by extension their way of being."[34] Reilly illustrates this onto-epistemic mimesis through examples ranging from the destruction of religious icons by iconoclasts in the English Reformation to the twentieth-century rise of the industrialized-worker Robots in Karel Čapek's *R.U.R*. Reilly draws on Walter Benjamin's idea of the *monad*—a particular object "that offers surprising flashes of insight into a specific historical moment."[35] For Reilly, this object is the automaton; for Benjamin, the Turk.[36] However, while Reilly uses the concept of the automaton to further illustrate particular moments within theatre history,

---

[32] Kara Reilly, *Automata and Mimesis on the Stage of Theatre History* (New York: Palgrave Macmillan, 2011), 11.
[33] Ibid., 6.
[34] Ibid., 7.
[35] Ibid., 1.
[36] Benjamin discusses the Turk, the famous "automaton" chess player, in the opening of his "Theses on the Philosophy of History" (in Walter Benjamin, *Illuminations*, trans. Hannah Arendt [London: Random House, 2011]). I have deliberately omitted the Turk from my history, as it was not a true automaton but rather operated by a human chess-player hidden within the unit.

my interest is in the automaton itself as an example of the performative potential of the artificial human. Further, Reilly tends to conflate examples of automata that only existed in the imagination with those that were really produced historically; I have taken care to differentiate between the two.

A more exhaustive treatment of the automaton as an object of cultural importance is Minsoo Kang's *Sublime Dreams of Living Machines: the Automaton in the European Imagination* (2011), also previously mentioned. Revealing the automaton's "enduring presence and emblematic power"[37] across Western history, Kang argues that the automaton exists transhistorically as "a conceptual tool with which Western culture has pondered the very nature and boundaries of humanity."[38] Throughout his work, Kang illustrates how this conceptual tool has functioned both as an object around which cultural paradigms are formed and as an object through which those very cultural paradigms may be undermined. Kang claims to "present a new theory [of the automaton's symbolic power]… based on how the human mind categorizes natural phenomena in building a picture of reality, and what occurs when that picture is violated through the appearance of an anomaly that defies attempts to fit it into an established conceptual schema."[39] *Sublime Dreams of Living Machines* offers an introduction into what Kang describes as "a new field I will call the history of the imagination, defined in the most general sense as the history of what people fantasized about."[40] My own historical study dovetails with Kang's in many ways, as I trace the artificial human through many of the same historical moments. Like Reilly, however, Kang's interest is primarily in how the automaton has impacted cultural paradigms; I am concerned more particularly with the way in which those paradigms have influenced our concept of the artificial human.

---

[37] Kang, *Sublime Dreams*, 1.
[38] Ibid., 12.
[39] Ibid., 8.
[40] Ibid., 12.

HUMANITY AS A PERFORMATIVE CONSTRUCT

A robot is not quite a machine, at least in potentiality. A robot is a machine that is made as much like a human being as it is possible to make it, and somewhere there may be a boundary line that may be crossed…. We can apply this to life…. Somewhere, life crossed a boundary line. At some point a species arose that was not only aware of itself but had the capacity to be dissatisfied with itself. Perhaps a similar boundary line will someday be crossed in the construction of robots.
—Asimov, "The Sense of Humor"[41]

In the 1959 work The *Presentation of Self in Everyday Life*, Erving Goffman applies the concepts of theatrical performance to daily social interaction among individuals and groups, arguing that all social interaction is a performance of certain behaviors in order to achieve a particular effect on the viewer:

> Society is organized on the principle that any individual who possesses certain social characteristics has a moral right to expect that others will value and treat him in an appropriate way. Connected to this principle is a second, namely that an individual who implicitly or explicitly signifies that he has certain social characteristics ought in fact to be what he claims he is.[42]

In other words, if I walk into a lecture hall filled with students, and I am dressed as a professor and act like the professor, the class will assume that I am the professor, so long as they have no information to the contrary. On the other hand, if I perform in this way and the class later finds out I was not in fact the professor, they will feel that I have violated that moral code: that I am "inauthentic."

Similarly, the act of being human entails a codified set of performative parameters, behaviors, attitudes, and/or abilities that distinguish us as "human." In everyday life as well as in science fiction, humans use a number of characteristics to distinguish themselves from robots.[43] Whether consciously or unconsciously, humans are judged by these criteria on a daily basis, and labeled "inhuman" or "mechanical" when a performance is seen as inauthentic. We constantly hear of feats of inhuman

---

[41] In Isaac Asimov, *Robot Visions* (New York: Penguin, 1990), 470-71.
[42] Goffman, *The Presentation of Self*, 13.
[43] These criteria are often culturally and historically determined. As such, I will address these criteria within their historical context throughout this study rather than enumerating them here.

strength, particularly cruel people are "animals," and others perform with "mechanical" precision. In science fiction, various combinations of these characteristics are used to distinguish the "authentic" human from the inauthentic one. Yet time after time in science fiction films, each of these "human" characteristics is called into question, either indirectly as a robot is able to successfully "pass" as human, or directly as the robot questions explicitly the validity of the distinction. For example, Asimov overtly mentions the performance of human traits when his "Bicentennial Man" Andrew's "carefully designed tricks of human expression and tone of voice betrayed him here."[44] Like Goffman's con artists, robots in science-fiction films appear to more or less successfully present the "front" of humanity with questionable legitimate claim thereto (according to the audience to which they present this front).

Judith Butler takes these notions of performance even further in her theories on gender, arguing that gender "authenticity" is a hegemonic cultural notion designed to enforce heteronormativity. Drawing on the work of Beauvoir and Irigaray, Butler describes gender not as a matter of biology but of role-playing, performing learned behaviors that are marked by society as "masculine" or "feminine" but are not inherent within the body itself:

> Gender ought not to be conceived merely as the cultural inscription of meaning on a pregiven sex (a juridicial conception); gender must also designate the very apparatus of production whereby the sexes themselves are established. As a result, gender is not to culture as sex is to nature; gender is also the discursive/cultural means by which "sexed nature" or "a natural sex" is produced and established as "prediscursive," prior to culture, a politically neutral surface *on which* culture acts.[45]

Under this formulation, culture may be seen as a set of pre-figured, universal meanings imposed on (passive) bodies based on biological features. On the other hand, Beauvoir's claim that one "becomes" a woman implies "an agent, a *cogito*, who somehow takes on or appropriates that gender and could, in

---

[44] Asimov, "Bicentennial Man," 283.
[45] Judith Butler, *Gender Trouble: Feminism and the Subversion of Identity*, Thinking Gender (New York: Routledge, 1990), 7.

principle, take on some other gender."[46] Whether figured as passive recipient of cultural meaning, or active participant in the construction of said meaning, Butler argues, "In either case, the body is figured as a mere instrument or medium for which a set of cultural meanings are only externally related."[47]

As Butler's deconstruction of the cultural categories of gender and sex serves to erase the inherent differences in the subject these categories imply, so does Ian Haney-López's deconstruction of racial identity and racial marking erode the concept of "race" as anything other than a social construct. In "The Social Construction of Race: some observations on illusion, fabrication, and choice," Haney-López, Berkeley law professor and advocate for racial justice, illustrates this phenomenon through case law examples. He first examines an 1806 Virginia case (*Hudgins v. Wright*) in which three generations of enslaved women successfully sued for their freedom based on their bronze complexion and straight hair, by which the Judge determined them to be of Native American rather than African descent and thus legally free.[48] According to Haney-López, *Hudgins v. Wright* illustrates the empirical definition of race—the possession of innate physical characteristics such as straight vs. "wooly" hair—that remains dominant in American society two centuries later. Taking an interdisciplinary approach that draws from biology, sociology, and literature, Haney-López rejects the notion of race as genetically determined and argues instead a for a theory of "racial fabrication" in which "race is neither an essence nor an illusion, but rather an ongoing, contradictory, self-reinforcing process subject to the macro forces of social and political struggle and the micro effects of daily decisions"[49]; a process through which racial labels such as Black, White, or Latino accrue to social groups rather than "genetically distinct branches of humankind."[50] This process of racial fabrication is further critiqued in Haney-López's *White by Law: The*

---

[46] Ibid., 8.
[47] Ibid..
[48] Ian F. Haney López, "The Social Construction of Race: Some Observations on Illusion, Fabrication, and Choice," *Harvard Civil Rights-Civil Liberties Law Review* 29 (1994).
[49] Ibid., 7.
[50] Ibid.. Haney-López draws from a broad range of scientific studies in which it was determined that, for example, there is more genetic variation within the broad racial categories listed here than there is between them. For a further discussion of genetics,

Legal Construction of Race (2006), which examines historical examples of racial assignment ("white" or "not white") in American immigration case law,[51] and the subsequent impact this had on the contemporary value of "whiteness" within American society.

While Haney-López's analysis is aimed at the legal community, arguing against an essentialist definition of race in legal deliberation, his point resonates with my own exploration of the performative nature of what we term "human." Haney-López locates the origin of racial identity outside the subject being so categorized. Likewise for Butler, gender is "an attribute of a person who is characterized essentially as a pregendered substance, or "core," called the person, denoting a universal capacity for reason, moral deliberation, or language."[52] By erasing gender and sex, and by extension race and many other "constitutives" by which groups of humans segregate themselves, these theorists reduce "personhood" to a question of a subject whose "intelligible identity" is constituted by cultural practices but who may not be equated with those practices: "the culturally enmired subject negotiates its constructions, even when those constructions are the very predicates of its own identity…. That *cogito* is never fully *of* the cultural world which it negotiates, no matter the narrowness of the ontological distance that separates that subject from its cultural predicates."[53] If this "core" subject is not the sum of its cultural practices, nor even of the physical body which houses it, then to deny "personhood" to any entity that is able to successfully perform "personhood" according to the predefined discursive conventions of the situation is an act of solipsism. As this study will show, identity as "human" as inscribed in the characteristics listed above is not the sole province of *Homo sapiens*.

---

race, and racism, see Sheldon Krimsky, Kathleen Sloan, and Council for Responsible Genetics, *Race and the Genetic Revolution: Science, Myth, and Culture* (New York: Columbia University Press, 2013).

[51] From 1790 to 1952, naturalization as a U.S. citizen was limited only to "white persons," a label based on a number of criteria, often assigned arbitrarily by immigration officials and legal decisions.

[52] Butler, *Gender Trouble*, 10.

[53] Ibid., 143.

CHAPTER OVERVIEW

Our journey through the cultural and historical landscape of the artificial human begins in antiquity. Rooted in mythology and legend, the creation stories of early history may be classified into three distinct traditions: *supernatural*, created by gods or through divine (or sinister) influence; *biological*, new beings developed from existing tissue; and *mechanical*, forged of metal or cleverly constructed by engineers. These early creation myths contain seeds of all three traditions in their descriptions of humanity's origins, and those seeking to walk in the footsteps of the gods take these myths as their inspiration. Chapter one ("From Magic to Machinery") traces these traditions through Western civilization, from Archaic Greece and the Homeric legends of Hephaestus' golden maidens to the rise of natural philosophy in the late Middle Ages, when scholars turned to scientific rather than supernatural explanations for the phenomena they encountered. Along this journey we will meet the Greek, Roman, and Islamic engineers whose technological advances may be seen in the twenty-first century artificial human. We will also explore how the idea of the artificial human was transferred via handwritten tomes across Eurasia, copied, edited and translated, with the knowledge contained therein often lost or misinterpreted before being rediscovered decades or even centuries later.

Chapter two, "The Crisis of the Man/Machine," begins with a paradigm shift in Western philosophy: the rise of mechanism, brought about largely through the (re-)discovery of the work of these same engineers—such as Hero, Vitruvius, and al-Jazarī —on the European continent in the late Middle Ages. Most famously articulated through the work of Rene Descartes, mechanistic philosophy holds the machine as its organizational metaphor: that everything in the universe, from the smallest creature to the very stars in the heavens, may be understood in terms of mechanical principles. From an examination of mechanism as the dominant worldview of the time, the chapter continues with the effect this worldview had on contemporary concepts of the human body and mind, an effect which culminated in the declaration by La Mettrie that Man is a machine. Within this clockwork universe the

artificial human reigned supreme: this period witnessed a proliferation of clever automata, clockwork marvels in human form that could draw and play musical instruments. Basing their work on the concepts first set down in Greek, Latin, and Arabic texts, the creators of these automata gained new respect within intellectual circles as their creations seemed to offer proof of mechanistic ideals; we look at the most famous of these creations: those of Vaucanson and Jaquet-Droz, and the impact they had on the worldview of the time.

Many philosophers, however, rejected hard mechanism, arguing that biological life possessed certain a certain vital force inherent to living matter, unexplainable by mechanical principles. The chapter continues by tracing this shift to *vitalism* through the eighteenth century discovery of *galvanism*, or animal electricity, and the backlash of *sentimentality*, in which rigid adherence to mechanistic principles came to be viewed as lacking: Man was *more than* a mere machine, possessed of an emotional soul and a freedom of will that inferior automata could never duplicate. The uneasy struggle between mechanism and vitalism continues in the mind-body debate central to the later field of artificial intelligence. Chapter two concludes with an analysis of Mary Shelley's novel *Frankenstein,* positing the Creature as the first artificial human who qualifies for the status of *Homo artificialis*.

As the Industrial Revolution streamlined production processes, moving the manufacture of goods from cottage industries to large factories, the skilled craftsman gave way to modern automation, and the human worker became a mere cog in the machine. Chapter three, "The Path to Personhood," opens with Karel Čapek's *R.U.R.* as a literal glimpse into the man=machine metaphor, offering to the world the Robot, the mass-produced artificial human laborer. We continue by tracing the connections between Vaucanson's automata through the industrialization of textile production to Charles Babbage and the development of the first computing machines in the UK and the USA. After a consideration of Alan Turing's response to the question "Can machines think?" we turn to the emerging genre of science fiction and the thinking machines that inhabit its universe.

At this point, however, our journey splits into two parallel branches, as the related ideas of artificial body and artificial mind, easily connected in the imagination, develop concurrently but separately in the fields of robotics and computer science. First, we examine the quest for artificial intelligence as a struggle to define "intelligence" by a human yardstick—and the implications this form of definition carries. Next, we turn to the technological advances in robotics from their industrial origins to their human application, culminating in the field of social robotics, which reunites the artificial mind and artificial body in a humanoid robot that behaves "just like us." We close our historical journey with an analysis of the various "human yardsticks" offered throughout the history of the man=machine metaphor—such as intelligence, consciousness, emotional capacity—and what these yardsticks tell us about the performative nature of humanity itself.

The three case studies are prefaced by a short section entitled "Case Studies: *Homo artificialis* in Science Fiction," in which I discuss how these human performatives may be analyzed in a fictional universe. I first analyze two concepts of the criteria of personhood, one from the Nonhuman Rights Project, which seeks to grant personhood to certain intelligent animals based on their performance of human-like sentience, and the other from the viewpoint of utilitarian machine ethics, which argues the validity of treating intelligent machines as objects of moral consideration. Drawing from these two concepts, I next outline the criteria by which *Homo artificialis* may be determined. Finally, I justify my three case studies as unique examples worthy of consideration within the larger genre of robotic science fiction.

In each case study, I address the robot character as an example of *Homo artificialis* within the diegetic fictional universe in which he[54] exists, addressing in turn the character's physicality, behavior, societal status, distinctiveness, and performance. My first case study addresses the character of C-3PO, protocol droid within the *Star Wars* universe; the second, David-8 of *Prometheus*, part of the *Alien*verse;

[54] I use "he" in this case as all three studies involve characters that are identified as male.

and finally, Data, officer aboard the USS Enterprise in the series *Star Trek: The Next Generation* and member of the larger *Star Trek* universe. Each study concludes with an evaluation of the robot's successful performance and status claim as *Homo artificialis*.

Finally, the conclusion ("They Walk among Us") offers a matrix of human performance based on the three case studies, and posits *Homo artificialis* as a combination of successful physical performance and social/emotional performance by an artificial human, classifying artificial beings as *toaster*, *social robot*, *Doppelgänger*, or *Homo artificialis* based on these two criteria. This matrix is populated by various examples of real and fictional robots based on their own success in each of these performance paradigms. After a discussion of the implications of granting *Homo artificialis* personhood status, the study concludes with the case of Pepper, a social robot expected to launch in February 2015 (only a few months after the end of this study). According to the promotional literature, Pepper could theoretically qualify, at least minimally, as *Homo artificialis*. What are the implications to our own definitions of humanity if Pepper lives up to the hype?

CHAPTER ONE: FROM MAGIC TO MACHINERY

He walks a delicate line between church, theater, magic, science. He circulates among—murmurs about—all of them. He is a synapse, transmitting a host of simultaneous signals. Here is a machine that prays. Is it a divine machine? Or, man-made, a miracle in its own right? Or again, as a votive offering, a machine made if not for God alone, then one meant to appear to be made for God alone?...
What did a person see and believe who witnessed the monk in motion in 1560?

—Elizabeth King, "Clockwork Prayer"[1]

In the Smithsonian resides a 450-year-old monk who, on command, will slowly walk in a circle murmuring prayers and lifting his prayer beads. He is regarded as the earliest-built human automaton still operational, commissioned in 1562 by Philip II of Spain as an offering to God for the successful recovery of his son: the prince Don Carlos, heir to the throne, had fallen down the stairs and suffered a terrible head injury. "Save my son," the king is supposed to have prayed to his God, "and I will in turn repay you with a miracle of my own."[2] According to the legend, after Don Carlos recovered from his injuries, Philip II turned to renowned clockmaker Juanelo Turriano to fulfill his end of the bargain. However, it was not Turriano's first human automaton; in the Kunsthistorisches Museum in Vienna is displayed his *Lute Player*, a dainty figure designed (but no longer able) to strum her instrument and dance along in either a straight line or a circle, created for Charles V of Spain sometime between 1555 and 1558.

Turriano was motivated to create his marvels by reading historical accounts of Greek automata, according to historian Ambrosio de Morales (1513–1591), professor at the University of Alcalá de Henares and close friend of Turriano:

---

[1] Elizabeth King, "Clockwork Prayer: A Sixteenth-Century Mechanical Monk," Part VI, *Blackbird* 1, no. 1 (Spring 2002), accessed 21 April 2013, http://www.blackbird.vcu.edu/v1n1/nonfiction/king_e/prayer_10.htm.

[2] According to King, and to curators at the Smithsonian, while the mechanical monk on display fits the description of the Turriano legend, and has been dated to the time period, it cannot conclusively be determined that it is in fact the one created by Turriano. Elizabeth King, "Perpetual Devotion: A Sixteenth-Century Machine That Prays," in *Genesis Redux: Essays in the Philosophy and History of Artificial Life*, ed. by Jessica Riskin (Chicago: University of Chicago Press, 2007), 266.

Figure 2. 450-year old clockwork monk, possibly created by Juanelo Turriano in the 1560s (although the exact provenance is currently unknown). The automaton currently resides in the Smithsonian's National Museum of History and Technology. Photo © Erich Lessing / lessingimages.com.

> Juanelo as a diversion also wanted to create anew the ancient statues which moved and, on that account, were called automata by the Greeks. He made a lady more than one tercia high who, placed on a table, dances all over it to the sound of a drum which she meanwhile beats herself, and goes round in circles, returning to where she started from. Though it is a toy and fit for mirth, it is nevertheless a great proof of his high intelligence.[3]

[3] A *tercia* is an archaic Spanish measurement of length roughly equal to 28 cm. Ambrosio de Morales, *Crónica general de España, prosiguiendo adelante los cinco libros que el Maestro Florian Docampo, Coronista del Emperador D. Carlos V dexó escritos* (Alcalá, 1574), 93; quoted in José A. García-Diego, *Juanelo Turriano, Charles V's Clockmaker: The Man and His Legend* (Madrid: Editorial Castalia, 1986), 101

Unfortunately for us, none of these Greek automata as described by Morales still exist. We know of them only from first- or second-hand historical accounts and legend. What we do know, however, is that the *idea* of the artificial human has been around at least since the Homeric legend of the *kourai khryseai*—two "golden maidens" designed by Hephaestus to assist the demigod in his forge. According to *The Iliad* (ca. mid-eighth century BCE), these golden maidens are "in appearance like living young women. There is intelligence in their hearts, and there is speech in them and strength, and from the immortal gods they have learned how to do things" (Homer, Iliad 18, lines 416ff).[4] Tales of artificial humans—some of which can be verified, others merely the stuff of legend—are scattered throughout history. Those who claim (or are claimed) to have successfully created one of these creatures are, depending on the intellectual climate of the day, variously hailed as geniuses, decried as charlatans, revered as saints, or reviled as being under the thrall of demonic influence—reflecting the long history of ambivalence toward these not-quite-human creations.

The history of the artificial human is the history not of an event, or a thing, but of an idea. As John Cohen claims in the preface to his 1966 work *Human Robots in Myth and Science*, the "Automaton deserves an encyclopedia," and his work can only barely scratch the surface of this "prince of ideas in contemporary science."[5] The description Cohen gives here of his attempt to adequately address the topic is particularly eloquent, and deserves to be quoted at length:

> Hunting high and low, I have tried to make Automaton my captive. But it refuses to be taken hold of and put in a cage with a label attached. It pops up here, there, and everywhere, in myth, legend, poetry, religion and mysticism, and in the history of physics, chemistry, and biology. It has intrigued philosophers, inventors, and mechanicians throughout the ages, and is a prince of ideas in contemporary science,

[4] For a discussion of Hephaestus' golden maidens as early literary examples of artificial intelligence, see George L. Hersey, *Falling in Love with Statues: Artificial Humans from Pygmalion to the Present* (Chicago: University of Chicago Press, 2009), 134.
[5] John Cohen, *Human Robots in Myth and Science* (London: George Allen & Unwin, Ltd, 1966), 7.

> bidding fair, in the near future, to transform our entire domestic and social life…. I hope that I have been able to suggest some continuity between one of the most pregnant concepts of modern science and technology and fantasies which had their birth in the dark backward and abysm of time, a continuity which links Homer with Babbage, Prometheus with the golem-makers, Sylvester II with Mary Shelly, the Midrash with Paracelsus and Weismann, and indeed joins together the imaginative works of China, India, Persia, Egypt and Western Europe…. I suspect that the history of automata reveals the duality in man's never-ending struggle to achieve, first, a technical mastery of his surroundings, which now include the boundless spaces of the cosmos and, second, to become one of the gods himself, by transcending both matter and himself.[6]

This duality Cohen mentions reflects that described by Peter Weyland as well in the 2012 science-fiction blockbuster *Prometheus*, as mentioned in my introduction: Weyland seeks not only to push the boundaries of technological innovation, but through his achievements prove his worth to his own Engineer creator, and thus ascend to godhood in his own right. The history of this dual struggle forms the continuity I trace through my transhistorical journey. Like Cohen, I can only scratch the surface of this topic. Therefore, our present journey can give only an overview of the artificial human's "common challenge to the restless human mind."[7]

The project Cohen found daunting in 1966 has grown exponentially in the half-century since his attempt. While the range of popular and scholarly works on the subject easily numbers in the hundreds, I have focused primarily on three recent works as the basis for my own history of the artificial human.[8] The first, by Horst Glaser and Sabine Rossbach (*The Artificial Human: a tragical history*, 2011), focuses on the Promethean nature of the artificial human and his status as the symbol of human achievement:

---

[6] Ibid..
[7] Ibid..
[8] While I will naturally refer to a number of other sources throughout these two chapters, these three works provide the framework for my discussion of the history of the automaton.

"the Prometheus myth, therefore, is intimately bound up with the development of the *téchne* that enabled the human race, through the exploitation of fire as a source of energy, to control, use and enjoy the fruits of its environment."[9] In usurping the role of "both Prometheuses" (Prometheus the creator of man and Prometheus the giver of fire, and thus *téchne*), mankind demonstrates that "humanity, far from being the slave of the material universe, was now to an ever increasing extent its master."[10] Glaser and Rossbach's work is divided into two main sections. In the first section, the authors trace mankind's aspiration to the status of *secundi dei*, or "second gods," from Hesiod's earliest creation stories (c. 700 BCE) to modern in-vitro fertilization technology. In the second, they illustrate this aspiration through a number of "artificial women" created at various points in history, from Pandora to the Sexbot.[11] Glaser and Rossbach discuss these artificial women in terms of their significance as reifications of hegemonic patriarchal misogyny due to the automata's typical role as submissive and subservient. However, their choice to analyze only female representations disregards the larger issue: that *all* automata, not only those marked with feminine attributes, are stereotypically placed in the subservient role.

While Glaser and Rossbach's work proves a rich resource of examples of historical and legendary artificial humans, their overarching theses are problematic and at times patronizing to the reader. Their "tragical history" is a thinly veiled Marxist critique of the potential of the artificial worker to replace the human one:

> One asks where the capital should come from for the production of the robots that will, in this scheme of things, replace human labor. The day may come when they can reproduce themselves and allow their human masters to pursue more profitable interests. The question remains what these would likely be.[12]

---

[9] Horst Albert Glaser and Sabine Rossbach, *The Artificial Human: A Tragical History*, trans. Joseph Swan (Frankfurt: Peter Lang, Inc, 2011), 10.
[10] Ibid., 15.
[11] Ibid., 246.
[12] Ibid., 35.

Unfortunately, for Glaser and Rossbach, these "more profitable interests" comprise largely the "avoidance of boredom"[13] that has the potential for destructiveness. The repercussions of this potentially destructive "boredom avoidance" can be seen, for example, in the 2010 tragedy at the "Love Parade" in Duisburg, Germany, where 21 people were trampled and killed in the crowd while the rest of the party continued unaware and seemingly uncaring. For Glaser and Rossbach, the "liberation" offered by the robot worker is a false one, as "the post-industrial society manifestly harbors a destructive potential as great as the industrial"[14] in which mankind was first reduced to cogs in the machine of factory production.[15] So while I am indebted to Glaser and Rossbach for their extensive discussion of Promethean mythology and the symbolic nature of the artificial human, as well as for their historical examples, their arguments follow a different path from my own.

The second major history I have consulted for this work is Kara Reilly's 2011 *Automata and Mimesis on the Stage of Theatre History*, a reworking of her 2006 PhD thesis. I have chosen *Automata and Mimesis* partly because Reilly's approach to historiography reflects my own—a transhistorical study,

> based on Walter Benjamin's concept of the monad: a particular object—in this case the automaton—that offers surprising flashes of insight into a specific historical moment. As a result, the automaton can be seen as a metaphor for the historical period in which it is explored.[16]

As outlined in my introduction, a transhistorical study falls somewhere between Foucault's "history of the present" and Deleuze and Guattari's rhizomatic ahistoricity, and as such is well suited to the concept of the artificial human. I am indebted, therefore, to Reilly for her historiography. Reilly's work, like mine, also views automata through a theatrical lens. However, my focus is quite different from hers in how

---

[13] Ibid..
[14] Ibid..
[15] For a discussion of the worker as part of the factory machinery in the discourse of the early Industrial Revolution, see Simon Schaffer, "Enlightened Automata," in *The Sciences in Enlightened Europe*, ed. William Clark, Jan Golinski, and Simon Schaffer (Chicago: University of Chicago Press, 1999), 126-165.
[16] Kara Reilly, *Automata and Mimesis on the Stage of Theatre History* (New York: Palgrave Macmillan, 2011), 1.

that lens is applied. Reilly bases her analysis on the concept of *onto-epistemic mimesis*: "mimesis that changes a person's way of knowing, and by extension their way of being"[17] in order to "explore the relationship between automata and cultural history,"[18] and focuses on particular moments in history. My own history, on the other hand, serves as a background for the exploration of the performative nature of "humanness." Further, while Reilly cites a number of interesting examples of historical automata, she neglects at points to distinguish between real examples and those of legend; while this distinction may not come into play in her history, it is important to my own, and I prefer to distinguish those automata which were only attributed to certain individuals by others from those that actually existed.

The third recent work on which I have based my history of the artificial human is Minsoo Kang's *Sublime Dreams of Living Machines* (2011). This work is also transhistorical: at the outset, Kang states plainly that his purpose is "to relate major instances of the [automaton's] manifestation in both actual and imaginary forms from premodern eras to the twentieth century, to explain its enduring presence and emblematic power."[19] In addition, Kang's approach to the history of the "living machine" meshes most closely with my own, regarding it as "a conceptual tool with which Western culture has pondered the very nature of the boundaries of humanity."[20] This tool as described by Kang is the one I will use in later chapters to dissect the performative actions that comprise the "human."

My own approach departs from Kang's in several ways, however. First, while I am concerned with several categories of artificial humans (which I outline below), Kang focuses on the mechanical automaton, mentioning other forms of artificial humans only tangentially in his work. Kang is further

---

[17] Ibid., 7.
[18] Ibid., 15.
[19] Minsoo Kang, *Sublime Dreams of Living Machines: The Automaton in the European Imagination* (Cambridge, MA: Harvard University Press, 2011), 1.
[20] Ibid., 12.

interested in the Uncanny Valley[21] phenomenon as it applies to the automaton; he claims that the power of the automaton as an idea is related to its ability to elicit widely disparate emotions in its audience: laughter, horror, terror. However, the Uncanny Valley phenomenon is not a major focus of my own study. In addition, Kang ends his history of the automata in 1920 with Karel Čapek's *R.U.R.*, claiming (quite rightly) that "a comprehensive study of robot symbolism in the postwar era would require another full volume."[22] While I make no claim that I am offering this sort of comprehensive study, I do focus my case studies on the artificial humans of the past forty years, tracing their roles as ersatz humans upon which we can read "human" performative actions. Finally, Kang's cultural analysis is the primary focus of his history of automata; my own historical tracing is more of a "means to an end," as the history informs the case studies rather than vice versa. Kang briefly identifies in his conclusion a number of themes such a study should address, such as artificial intelligence and the nature of the human mind; many of these themes are addressed in this dissertation, as well as others that Kang had not considered.

SUPERNATURAL, MECHANICAL, BIOLOGICAL

For the purposes of this dissertation, I have divided the history of the artificial human into three broad categories of beings. The first category consists of those beings whose "awakening" is attributed primarily to supernatural means, such as the *golem*: the creature of Jewish legend, made of clay, brought to life by virtue of the great piety of the rabbi creating it. The second category of artificial

---

[21] The phenomenon known as the Uncanny Valley was first articulated by Masahiro Mori in 1970, and basically claims that as the appearance of artificial humans become more human-like, human response to them is increasingly positive, but only to a point, at which the response turns swiftly to revulsion before again becoming positive as the artificial human becomes practically indistinguishable from the natural one. For decades the Uncanny Valley concept has been accepted dogmatically; however, recent studies have called this idea into question. See for example Christoph Bartneck, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita, "My Robotic Doppelgänger—a Critical Look at the Uncanny Valley," *RO-MAN 2009 - The 18th IEEE International Symposium on Robot & Human Interactive Communication* (Jan 2009): 269-276; Jakub Złotowski, Diane Proudfoot, and Christoph Bartneck, "More Human Than Human: Does The Uncanny Curve Really Matter?" *Proceedings of the HRI2013 Workshop on Design of Humanlikeness in HRI* (2013): 7-13.

[22] Kang, *Sublime Dreams*, 298.

human is that of the mechanical being, the "clockwork curiosity" or *automaton*.[23] From the clever constructions described by Hero of Alexandria in the first century CE, to the mechanical marvels of the heyday of automata in the eighteenth and nineteenth centuries, to the social robots of MIT's Cynthia Breazeal (among others) in the twenty-first, these beings are triumphs of technology. The mechanical being is the one for which we have the strongest tradition, and with which we have the most success outside our imaginations: these robots exist in reality as well as in fiction. The final category comprises those beings whose creation is achieved by biological means—known from medieval times as *homunculi* (sing. homunculus). These creatures are the flesh of our flesh: they range from the nameless monster brought to life by Frankenstein to the scientific reality of test-tube babies and cloning. These three categories, supernatural, mechanical, and biological, overlap and blend greatly across the eras; furthermore, as I will demonstrate, we can trace their origins to various creation myths proposed throughout history.

It is important to note, with a Foucauldian nod at the ahistoricity of the categorical terms I have chosen, that most of the examples of artificial humans I discuss throughout this study represent the latest in philosophical thought within their respective eras. Each of these creations reflects an attempt by scholars of the time to understand how we ourselves "became" human. Through re-creation (not to mention a good deal of reverse engineering) of the natural human through the artificial, we answer those metaphysical questions posed by Peter Weyland in *Prometheus*: "Where do we come from? What is our purpose? What happens when we die?" Whether or not the re-creations were successful is beside the point. What matters is the *belief* that they were alive, an animistic tradition that Cohen claims has Egyptian origins: "the Egyptians believed that a statue had a *ka*, or 'shade' or 'double' which represented a god or a deceased person. The priestly 'keepers of the mystery' could make the soul enter

[23] For a discussion of the evolution of the word automaton, from an object that moves under its own power ("self-mover") in ancient Greek to the current use of the word to describe at once both a machine that mimics a human and a human that behaves as a machine, see ibid., 6-8; cf. Hersey, *Falling in Love with Statues,* 134: "The very history of the word 'automaton,' in short, traces a mechanical creature's declaration of independence." Hersey's chapter eight is concerned with automata as independently-functioning beings.

the image by some rite."[24] This tendency toward animism is not limited to so-called "primitive" tribes and cultures, according to historian George Hersey:

> What is important here, however, is that for millennia people have believed that things we consider inanimate were alive. And that belief, in turn, was acted on and has covered an enormous amount of cultural territory…. In the past … statues were seriously thought to possess artificial life; indeed in some cultures that is still so. And our own culture, believingly or unbelievingly, has constantly reproposed the idea.[25]

The golden maidens of Hephaestus, for example, were described in no uncertain terms as living beings, created by a god to serve as helpers in his forge. When I categorize these maidens as *supernatural* creations, however, I am at least partially superimposing my own twenty-first century belief system on the phenomenon. These were also *mechanical* creations, by the ancient Greek understanding of their design, and a clever enough human could forge living statues of his own.

THE WONDERS OF ANTIQUITY

Indeed, at least according to Greek myth, one such talented human was Daedalus, the renowned craftsman and artisan first mentioned in Homer, but in such a way that it is apparent his story was already well known at the time. Both Plato (*Meno* 97) and Aristotle (*Politics* i.4), writing in the fifth and fourth centuries BCE, attribute to Daedalus the talent to create statues with the semblance of life, which would run away if not tied down; Aristotle also credits Daedalus with having created a wooden statue of Aphrodite animated with quicksilver. Philostratus (second and third century CE) later credits Daedalus as the first person to create such lively statues, never questioning his ability to do so. Philostratus describes the statues in Daedalus' workshop as depicted in a mural: "some with forms

---

[24] Cohen, *Human Robots in Myth and Science*, 20; cf. Reilly, *Automata and Mimesis*, 24. In the chapter "Robots in Antiquity," Cohen offers an extended discussion of "ensouled" statues, including a look into their theological origins. However, Cohen indicates that this constitutes an "Eastern" origin for the tradition, conflating the various ancient civilizations he discusses into this category: Egyptian, Assyrian, Theban, Greek, Indian, Chinese. I have chosen to elide this conflation. See also Hersey's discussion of the importance of animism in "non-Western" cultures in *Falling in Love with Statues.*

[25] Hersey, *Falling in Love with Statues,* 2-3.

blocked out, others in a quite complete state in that they are already stepping forward and give promise of walking about. *Before the time of Daedalus, you know, the art of making statues had not yet conceived such a thing*."[26] Ancient writers appear never to have questioned the veracity of this legend, which inspired generations of Greek and Roman sculptors to try to follow in Daedalus' footsteps in making their creations as "alive" as they could—and stories of the wondrous moving statues created by the Greeks are numerous. Pindar, for example, on the occasion of Diagoras' victory in the seventy-ninth Olympiad (464 BCE), immortalized the great athlete's hometown of Rhodes as a wonder of mechanical engineering:

> The animated figures stand
>
> Adorning every public street
>
> And seem to breathe in stone, or move their marble feet.[27]

These accounts resurface centuries later, as medieval and Renaissance engineers attempted to replicate the mechanical wonders of the ancient age.

And wonders they were. While there is no indication that Hephaestus' golden maidens, Daedalus' "lively statues," or the "animated figures" of Rhodes existed outside these legendary accounts, there are several ancient scholar/engineers whose works are repeatedly mentioned[28] in relation to early automata and robotics. The earliest of these is the Greek mathematician and inventor Ctesibius[29] (c. 285–222 BCE), whose various treatises comprised the earliest work on pneumatics, hydraulics, and counterweight systems. While none of these treatises survived to present day, his legacy

---

[26] Philostratus, *Imagines* 1.16, emphasis mine.
[27] Pindar, *Seventh Olympic Ode*, lines 95-98.
[28] See John G. Landels, *Engineering in the Ancient World* (Berkeley: University of California Press, 1978); particularly his discussion of Hero of Alexandria and Vitruvius in chapter nine as "the most important sources of our knowledge about ancient technology" (Ibid., 199). In this chapter Landels outlines the extant (circa 1978) source material for each thinker. Likewise, Noel Sharkey in "I, Ropebot" (*New Scientist* 194, no. 2611 [2007]) refers to "the three great automata-makers: Ctesibios, Philo and Hero" in his quest for the origins of programmable automata. See also Sylvia Berryman, *The Mechanical Hypothesis in Ancient Greek Natural Philosophy* (New York: Cambridge University Press, 2009). Most of the other histories cited in this and the following chapter mention at least one, if not all, of these pioneers in engineering.
[29] Κτησίβιος, also spelled Ktesibios, Ctesibios, or Tesibius. For sake of clarity and uniformity, I have chosen to retain the most common Latinized versions I found in my research of the Greek names and titles of works discussed in this section.

may be seen through numerous references to his inventions by his successors. Very little is known of Ctesibius outside of these references, but it appears that during his career in Alexandria he enjoyed the patronage of the Ptolemaic kings.[30] Of particular interest are his force pump, an early displacement pump known in antiquity as the "Ctesibian machine,"[31] and the *clepsydra*, or water clock, in which he improved the accuracy of previous water clocks by providing a novel solution to the problem of diminishing flow (in which the clock interval would change as the water pressure diminished).[32] While Ctesibius is not known to have created a human automaton of his own, his work in hydraulics and timing mechanisms provide the building blocks for the automata that were to come later.

Philo of Byzantium (c. 280–220 BCE), known as Philo Mechanicus, also lived and worked in Ptolemaic Alexandria. Philo's *Mechanike syntaxis* (Compendium of Mechanics) covers a number of scientific fields, and refers repeatedly to Ctesibius, commenting and occasionally improving upon Ctesibius' innovations. While Philo is best known for his advances in siegecraft (including improvements to the crossbow and catapult), his *Mechanike syntaxis* also contained sections on pneumatics and automata, both of which figure greatly in our history. Unfortunately, while *Pneumatica* survives in a Latin translation of an Arabic version,[33] we know of the contents of Philo's *Automatopoieca* only through references to it by Hero of Alexandria. [34] According to Hero, Philo had described in the *Automatopoieca* an automatic theatre much earlier than Hero's own version, although Hero goes on to enumerate his own additions and improvements to Philo's designs.

---

[30] Landels, *Engineering in the Ancient World*, 200.
[31] Ibid., 75.
[32] Ibid., 193.
[33] Until 1974 only available in German or French, both the Latin and Arabic versions of the Pneumatica manuscript were translated into English, by Frank David Prager: Philo of Byzantium and Frank David Prager, *Pneumatica: The First Treatise on Experimental Physics, Western Version and Eastern Version*, trans. Frank David Prager (Wiesbaden: Reichert, 1974).
[34] Finlay McCourt, "An Examination of the Mechanisms of Movement in Heron of Alexandria's on Automaton-Making," in *Explorations in the History of Machines and Mechanisms: Proceedings of Hmm2012*, ed. Teun Koetsier and Marco Ceccarelli (Dordrecht: Springer, 2012), 186.

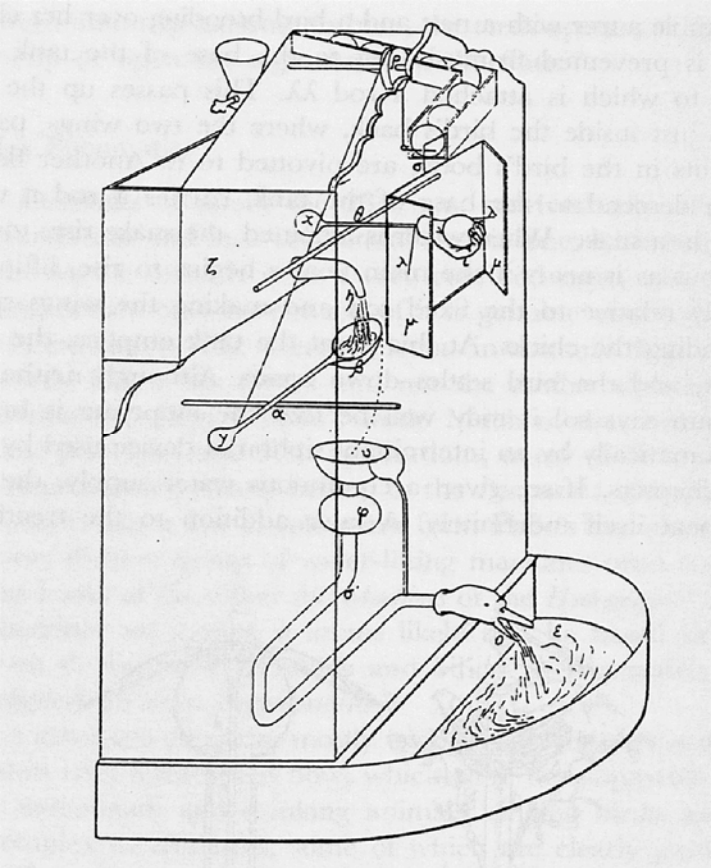


*Figure 3. Reconstruction of a washstand with escapement mechanism, the earliest known, as described by the Greek engineer Philo of Byzantium (3rd century BCE). From Bernard Carra de Vaux, "Le livre des appareils pneumatiques et des machines hydrauliques de Philon de Byzance d'après les versions arabes d'Oxford et de Constantinople,"*Notices et extraits des manuscrits de la Bibliothèque Nationale, Tome XXXVIII (Paris: Academie des inscriptions et des belles lettres*, 1903).*

The first texts that describe the creation of human automata that can be reliably verified and dated are those attributed to Hero of Alexandria (c. 10–70 CE), a Greek mathematician and engineer whose scientific treatises are some of the earliest to survive antiquity in more than a fragmentary form. Of particular interest is Hero's *Peri Automatopoietikes* (Of Automata-Making), which includes a description of what Dr. Noel Sharkey, professor of artificial intelligence and robotics at the University of Sheffield, claims is the earliest evidence to date of ancient "programmable" automata: a moving theatre that

> was said to roll automatically across the floor on wheels to face an audience. Then it paused. At this point the upper half of the machine, which depicted a shrine to

> Dionysus, the Greek god of wine, came to life. This featured six automata, including Dionysus himself along with female worshippers, that performed a short show. When the show had finished, the vehicle trundled off stage. [35]

This automatic theatre was powered by a falling counterweight, which pulled a cord wound around two sides of a split axle. The automata's mechanism allowed for complex motion—including forward and backward motion, arcs, and pauses—depending upon how the cord was wound. It is this flexibility that Sharkey terms an early form of programming, which he calls a deliberate innovation:

> Having read *Peri automatopoietikes* carefully, there is no doubt in my mind that Hero's intention was clear. His writing makes it obvious that he designed his robot to be programmable to suit different theatrical purposes. At one point he even describes how to program complex behaviour by wheeling the robot backwards from the end point through each required movement of the performance to the start, letting the twine wind itself appropriately around the axles.[36]

Granted, Sharkey's definition of "programming" here is quite simplistic: a set of instructions that "can be changed without having to dismantle or rebuild the entire mechanism—in other words, the program has to be separate from the rest of the machine's workings."[37] Still, the basic design and functionality of Hero's device can be seen in later automata, such as Leonardo da Vinci's "lost robots;" designs which have been recreated by roboticist Mark Rosheim in his own quest for evidence of early programmable machines.[38] The question remains, however, of whether Hero's design was the first of its kind. Noel Sharkey is convinced that earlier examples exist:

---

[35] Sharkey, "I Ropebot." See also Mark E. Rosheim, *Robot Evolution: The Development of Anthrobotics* (New York: John Wiley and Sons, 1994), 5-6. A recreation of this theatre was constructed in 2005 at the Museum of the Ancient Greek Technology in Katakolo, Greece; see Kostas Kotsanas, "Αρχαια Ελληνικη ΤεχνολοΥια: The Automatic Theatres of the Ancient Greeks," accessed 15 May 2014, http://www.kotsanas.com/gb/exh.php?exhibit=0101003.

[36] Sharkey, "I, Ropebot."

[37] Ibid.. Compare this to the first modern supercomputer, ENIAC, which was programmed through a physical configuration of patch cables and switches; see chapter three page192.

[38] See Mark E. Rosheim, *Leonardo's Lost Robots* (Berlin: Springer, 2006).

Figure 4. An illustration of Hero of Alexandria's automatic theatre. From Wilhelm Schmidt, Heronis Alexandrini opera quae supersunt Omnia *(Leipzig: B.G. Teubneri, 1899).*

> It is hard to believe that there were no earlier designs. In his book Hero hinted that he was giving us something new in an already established theatrical tradition, but he complained that earlier writers were not clear enough to enable others to copy their robots. He made explicit references to *Philo of Byzantium's lost book on automatic theatres from 200 BC*, for example. Elsewhere there are references to even earlier automata—for instance, in the 4th century BC Aristotle wrote about automatic puppets and a child's wagon that could move in a circle, describing them as if they were commonplace.[39]

But while the origins of these mechanisms may be lost in time, many of the constructions of which we do have extant descriptions have been recreated in the past several decades,[40] offering at the very least a proof of concept for these designs. Sharkey cites sources as early as the Homeric tales of Hephaestus and his mobile tripods as possible indications of the existence of earlier programmable machines, but there is no extant corroborating physical evidence. And so the hunt continues: for roboticists like Rosheim and Sharkey, "the search for the earliest programmable robot is far from over."[41]

The final voice from antiquity that deserves mention in any history of the artificial human is the Roman scholar Vitruvius (Marcus Vitruvius Pollio, born c. 80–70 BCE, died after c. 15 BCE), whose sole surviving work *De architectura* (*Ten Books on Architecture*, likely written around 15 BCE) is the most important source to date of information for Roman architecture, and the only contemporary source to have survived antiquity. Vitruvius' work forms the basis of much of what is now known about Roman technology, and refers extensively to the work of Hellenic and Hellenistic engineers, corroborating and occasionally providing the sole reference to the work of his predecessors (such as in the case of

---

[39] Sharkey, "I, Ropebot," emphasis mine.
[40] The Thessaloniki Science Center and Technology Museum in Greece houses a collection of "more than 70 exact replicas of machines that have marked people's lives and set the cultural context" from prehistory to the Hellenistic and Byzantine eras, including some mentioned here. See Noesis, "Thessaloniki Science Center and Technology Museum: The Technology Museum," accessed 15 May 2014, http://www.noesis.edu.gr/en/the-technology-museum/.
[41] Sharkey, "I, Ropebot."

Ctesibius). While a number of copies of Vitruvian manuscripts existed in medieval Europe, it appears they received little attention until "rediscovered" in 1414 by Poggio Bracciolini (1380–1459), Florentine author and humanist who is credited with recovering, copying, and disseminating numerous classical texts.[42] While Vitruvius did not himself design or construct automata, his contribution to the history of the artificial human is considerable: after its rediscovery, *De architectura* had a profound impact on scientific thought in the Renaissance, as its precepts applied not only to architecture, but also to such diverse fields as art, theatre, hermeneutics, and engineering.

One of the greatest difficulties in tracing the history of ancient automata is the lack of distinction among legendary accounts of artificial beings, theoretical designs by scholars, and physical realizations of those designs. Add to this the "Chinese whispers" game that constitutes the transmission of ancient texts and the picture becomes increasingly muddled. For example, Kara Reilly ends her *Automata and Mimesis* with an account of "the first known automaton"[43] built in antiquity. Reilly relates "on good authority, from Aulus Gellius"[44] (c. 125–180 CE), that Archytas of Tarentium (c. 400–350 BCE) created a flying wooden dove. However, Gellius' account in *Noctes Atticae* (*Attic Nights,* X:XIII) is a retelling of an earlier account by Favorinus (80–160 CE): "About so improbable a story I prefer to give Favorinus' own words: 'Archytas the Tarentine, being in other lines also a mechanician, made a flying dove out of wood. Whenever it lit, it did not rise again.'"[45] Gellius' account in *Attic Nights* is the *only* mention in antiquity of Archytas' dove; Favorinus' account no longer exists. Although Vitruvius mentions Archytas as one of the ancient engineers he esteems, he makes no mention of the dove. Moreover, it is highly debatable as to whether Gellius' account is reliable, as *Attic Nights* is hardly a scholarly work: it is a compilation of random notes combining legend, travelers' tales, and memories of things Gellius may have read or overheard. In fact, Gellius mentions Archytas' dove in his chapter mainly as a contrast to

---

[42] See Stephen Greenblatt, *The Swerve: How the World Became Modern* (New York: W. W. Norton, 2011).
[43] Reilly, *Automata and Mimesis*, 179.
[44] Ibid..
[45] Aulus Gellius, *The Attic Nights of Aulus Gellius*, The Loeb classical library, trans. John Carew Rolfe (London: William Heinemann, 1982), 245.

the unbelievable stories told by Pliny the Elder of Democritus: "[Pliny] transmits to us many foolish and intolerable absurdities, alleging that they were written by Democritus."[46] After a short account of these absurdities, Gellius offers an alternative to Pliny's hearsay, seemingly unaware of the irony of his words:

> Many fictions of this kind seem to have been attached to the name of Democritus by ignorant men, who sheltered themselves under his reputation and authority. But that which Archytas the Pythagorean is said to have devised and accomplished ought to seem no less marvellous, but yet not wholly absurd. For not only many eminent Greeks, but *also the philosopher Favorinus, a most diligent searcher of ancient records, have stated most positively* that Archytas made a wooden model of a dove with such mechanical ingenuity and art that it flew.[47]

The dubious provenance of Archytas' dove is a lesson in the caution one must take in addressing these ancient automata, but it is far from the only example.

Both ancient and, as a consequence, medieval philosophy regarded these wonders of antiquity as both mechanical/biological *and* supernatural—that is, created by either natural or divine magic.[48] Likewise, in many cases ancient authors did not distinguish between automata built by men and those created by the gods, citing in one breath the golden maidens of Hephaestus, and in the next the figures of Hero—holding both as *real*. So while we have no direct physical evidence for many of these ancient robots, neither can we label the earliest creations as *merely* "supernatural"—a concept that could appear dismissive of the philosophical worldview of the time. Therefore, when I discuss supernatural creations, they are always also either mechanical (made of base, inert materials) or biological (made of living material; specifically, flesh in some form), particularly as we reach the historical paradigm shift in which supernatural explanations were no longer accepted as valid.

---

[46] Ibid., 243.
[47] Ibid., 245, emphasis mine.
[48] See my later discussion regarding magic in the medieval era. For further information, see also Kang, *Sublime Dreams*, chapter two; Derek J. de Solla Price, "Automata and the Origins of Mechanism and Mechanistic Philosophy," *Technology and Culture* 5, no. 1 (1964): 9-23.

DEUS FABER[49]

And the LORD God formed man of the dust of the ground, and breathed into his nostrils the breath of life; and man became a living soul.

– Genesis 2:7[50]

Since the ancient Promethean myths, human creation has been the province of the Creator God, to whom clever engineers such as Daedalus could only aspire. The concept of God the Creator comes to us in Western culture most clearly in the Abrahamic religions, such as in the quote from Genesis above. There are two separate accounts in Genesis of God creating mankind. This, the second one, implies that Man was created first, out of earth, to be followed by Woman, who is created from the flesh of Man at a later point:

> And out of the ground the LORD God formed every beast of the field, and every fowl of the air; and brought them unto Adam to see what he would call them: and whatsoever Adam called every living creature, that was the name thereof.
>
> And Adam gave names to all cattle, and to the fowl of the air, and to every beast of the field; but for Adam there was not found an help meet for him.
>
> And the LORD God caused a deep sleep to fall upon Adam, and he slept: and he took one of his ribs, and closed up the flesh instead thereof;
>
> And the rib, which the LORD God had taken from man, made he a woman, and brought her unto the man.
>
> And Adam said, This is now bone of my bones, and flesh of my flesh: she shall be called Woman, because she was taken out of Man. (Genesis 2:19-23)

---

[49] *Deus faber* is the "creator god" who forms man in one of two ways: (1) by molding him out of some form of earth (clay, dust, etc.) and water, then granting him the "breath" or "spark" of life, or (2) by forming him from some part of the god himself. The idea of the *deus faber*, therefore, at once encompasses all three methods for creating an artificial human: supernatural, mechanical, and biological.

[50] Authorized King James Version. Subsequent Bible verses are from this version.

Here you can clearly see a blending of the three categories of created being. Mankind in the Abrahamic tradition (as in many others) is created by a god possessed of supernatural powers. However, while man is created from dust (earth), and therefore falls also into the *mechanical* category, woman is created from flesh, and fits as well the *biological* category of created beings. Here, within a single creation story we can see the spectrum of artificial humans reflected in modern science fiction and science fact.

In contrast, the creation account from the first part of Genesis does not distinguish in this way between the creation of man and of woman, and in general gives little detail as to the manner of creation (besides "God did it"). It does, however, address the idea of anthropomorphism. Here, the idea of man being created in God's image (which in many mythologies includes not only physicality but psychology as well) directly correlates with artificial humans being created in the image of real humans:

> And God said, Let us make man in our image, after our likeness: and let them have dominion over the fish of the sea, and over the fowl of the air, and over the cattle, and over all the earth, and over every creeping thing that creepeth upon the earth.
>
> So God created man in his own image, in the image of God created he him; male and female created he them. (Genesis 1:26-27)

The Genesis account is by no means original, however. It is widely accepted that the creation accounts in Genesis, believed to have been written around seventh or sixth century BCE, are an appropriation of the one from Mesopotamia, as recounted in the *Epic of Gilgamesh* (thirteenth to tenth century BCE):

> They [the parents of the young men and women Gilgamesh had killed] also cried aloud to Aruru, the goddess, saying: "Aruru, who hast created him,
>
> Create now a rival (?) to him, for the time when his heart shall be [* * *],
>
> Let them fight together and Uruk [shall be the spectator?]!" Upon hearing this Aruru created in her heart a man after the likeness of Anu.

Aruru washed her hands, took a bit of clay, and cast it on the ground.

Thus she created Enkidu, the hero, a lofty offspring, the possession (?) of Ninib.[51]

This account, recorded contemporaneously with the earliest known forms of writing, is reflected in a number of later creation myths, including that of Prometheus who, according to Apollodorus, "moulded men out of water and earth and gave them also fire."[52] In some versions of the Prometheus myth, Athena looks on as Prometheus creates the first human; in others, Athena herself provides the spark of life to Prometheus' creation.

These creation myths contain traces of the earliest philosophical beliefs regarding the nature of being, such as the classical concept of the four elements (the basic matter of which all existence is comprised) and the related humors that make up the human body and temperament. In these accounts, as well as many others, man is made of the element of Earth, sometimes mixed with Water to form a clay, and endowed with the elusive "fifth element" or quintessence—via a breath (Air) or a spark (Fire)—that grants him life. Similar concepts existed synchronically in various parts of the world, and many are believed to originate in the same Mesopotamian belief system that gave us Gilgamesh and the man-from-dust creation myth tradition discussed above. Others, however, appear to have developed independently. The goddess (or by some accounts demigod) Nuwa of Chinese mythology was one of the Three Sovereigns who reigned long before the Xia dynasty (2070–c. 1600 BCE) and helped to shape the early development of the Chinese people. According to tradition, Nuwa shaped men out of yellow clay[53]

---

[51] "Epic of Gilgamesh," (JasonColavito.com Books, 2012), Tablet II. Amendments in original.

[52] Apollodorus, *The Library*, 2vols, The Loeb Classical Library, trans. James George Frazer (Cambridge, MA: Harvard University Press, 1921), 1.7.1. Frazer's notes to this section discuss the differences among the various ancient accounts.

[53] Tradition tells us that she bored of the process, and after carefully crafting a few hundred of these little men, instead randomly flung out clay into little blobs, which became commoners. Nobility are descended from those earlier, more carefully constructed figures. For more regarding this and other creation accounts, see David Adams Leeming and Margaret Adams Leeming, *A Dictionary of Creation Myths* (New York: Oxford University Press, 1995).

Figure 5. Above: Prometheus creates man while Athena watches (The creation of man by Prometheus. Marble relief, Italy, 3rd century CE, from the collection of the Louvre). Below: After Prometheus has created man out of mud, Athena breathes life into him, imparting reason and understanding. Part of a cycle on the myth of Prometheus by Christian Griepenkerl. Ceiling painting (oil on canvas) above the grand staircase in the Augusteum, Oldenburg.

to keep herself company in the lonely, empty world. While this account predates the Mesopotamian one, it does not appear to have influenced it.[54]

[54] The Chinese creation myth does reverberate through our history of artificial humans, however. The *Lieh Tzu* of the third century BCE recounts the story of Yen Shih, an "artificer" who presents King Mu of Zhou (tenth century BCE) with a singularly curious being:

> It walked with rapid strides, moving its head up and down, so that anyone would have taken it for a live human being. The artificer touched its chin, and it began singing, perfectly in tune. He touched its hand, and it began posturing, keeping perfect time. […] As the performance was drawing to an end, the robot winked its eye and made advances to the ladies in attendance, whereupon the king became incensed and would have had Yen Shih executed on the spot had not the latter, in mortal fear, instantly taken the robot to pieces to let him see what it really was. And, indeed, it turned out to be only a construction of leather, wood, glue and lacquer, variously coloured white, black, red and blue. Examining it closely, the king found all the internal organs complete—liver, gall, heart, lungs, spleen, kidneys, stomach and intestines; and over these again, muscles, bones and limbs with their joints, skin, teeth and hair, all of them

While the Judeo-Christian Genesis stories are rather vague on the details of the actual creation process, the later Islamic account is rather explicit. The Islamic creation story begins, as do earlier versions, with earth—in this instance, clay—but goes on to detail the steps comprising man's composition in a manner that some scholars suggest indicate a gestational, if not evolutionary, process:

> And certainly did We create man from an extract of clay. Then We placed him as a sperm-drop in a firm lodging. Then We made the sperm-drop into a clinging clot, and We made the clot into a lump [of flesh], and We made [from] the lump, bones, and We covered the bones with flesh; then We developed him into another creation. So blessed is Allah, the best of creators.[55]

The being resulting from this creation account seems to combine the mechanical tradition (clay) with the biological (the sperm-drop) through supernatural (divine) intervention. The Qur'an is of course the most recent of the Abrahamic texts, compiled in the mid-seventh century CE—a time when the Arab world was clashing with the Byzantine. It is certain that during this time Arab scholars heard accounts of the Greek's and Roman's legendary engineering skills and stories of their human automata, to say nothing of their knowledge of medicine through pioneers such as Galen (late second century CE), whose scientific discoveries became the standard for medical practice for nearly the next millennia. This additional scientific knowledge is reflected in the Qur'an's creation story, which melds supernatural, mechanical, and biological creation into a single account.

> artificial. […] The king tried the effect of taking away the heart, and found that the mouth could no longer speak; he took away the liver and the eyes could no longer see; he took away the kidneys and the legs lost their power of locomotion. The king was delighted. Drawing a deep breath, he exclaimed, "Can it be that human skill is on par with that of the great Author of Nature?" (Joseph Needham and Ling Wang, *Science and Civilisation in China*, <v. 1-4; vols. (Cambridge, UK: University Press, 1954), *Volume 2: History of Scientific Thought*, 53.)

The king, fascinated by Yen Shih's creation, simultaneously likens the accomplishment to that of Nuwa and embarrasses his own philosophers, who often crowed of their own knowledge but had yet to create anything nearly so ingenious. This account is also mentioned in Cohen, *Human Robots in Myth and Science*, 23-24, in his discussion of other ancient Eastern robots (Egyptian, Indian, Chinese).

[55] *Qur'an* 23:14-14, Sahih International Version (insertions in original).

THE ISLAMIC GOLDEN AGE

The Qur'an is revered so highly in the Muslim world that all intellectual disciplines in the culture—including not only theology and mysticism but also philosophy, jurisprudence, and even the natural sciences—have sought to establish their foundations in its teachings. Indeed, the period from the mid-eighth century to the Mongol conquest of Baghdad in 1258 is considered the Islamic Golden Age, with Baghdad as an international intellectual center whose scholars sought to collect all the world's knowledge and translate it into Arabic. In fact, it was through twelfth century Arabic translations that medieval Europe rediscovered Hellenic medicine, including the works of Galen and Hippocrates—as well as those of Avicenna (980–1037 CE), the Persian scholar whose *Canon of Medicine* later became a standard textbook at universities throughout medieval Europe.

There is a great deal of historical evidence for the creation of automata during the Islamic Golden Age. Three ninth-century scholars from Baghdad—Ahmad, Muhammad and al-Hasan ibn-Mūsā ibn-Shākir, known collectively as Banū Mūsā, or the "Sons of Moses"—are best known for their work on automata and other mechanical creations collected in *The Book of Ingenious Devices* (850 CE), a large illustrated volume containing over a hundred such devices, including schematics and details of their use.[56] The work, first commissioned by the Caliph Abu Jafar al-Ma'mun ibn Harun (786–833 CE), was intended as a catalog of the Hellenistic texts preserved during the decline and fall of the Roman Empire; however, the brothers included a number of devices of their own invention. Taking the texts of Hero of Alexandria as a starting point, the brothers devised a number of mechanisms not appearing in the work of their predecessors, particularly those that involved automatic control—"self-operating valves, timing devices, delay systems and other concepts of great ingenuity."[57] While historian Donald R. Hill seems to accept as a given that these devices were constructed contemporaneously, either by the brothers or by

[56]Ahmad ibn-Mūsā ibn-Shākir et al, *The Book of Ingenious Devices:* [*Kitab Al-Hiyal*], trans. and annotated by Donald R. Hill (Dordrecht: Reidel, 1979). See also Donald R. Hill and David A. King, *Studies in Medieval Islamic Technology: From Philo to Al-Jazarī, from Alexandria to Diyār Bakr*, 1 vols, Collected Studies Series (Brookfield, VT: Ashgate, 1998), section XII; Rosheim, *Robot Evolution*, 9.
[57] Donald R. Hill in his annotations to Ibn-Shākir, *Book of Ingenious Devices*, 23.

craftsmen they hired,[58] none of these actual examples have survived to the present. Those devices that *have* been reconstructed, however, work as they were designed.[59]

The Arab scholar, inventor, and engineer Ismāʿīl ibn al-Razāz al-Jazarī (1136–1206), best known for *The Book of Knowledge of Ingenious Mechanical Devices* (1206), was strongly influenced by the work of Banū Mūsā.[60] However, as chief engineer at the Artuklu Palace in Diyarbakir, Turkey, al-Jazarī's work was more practically oriented than the scholarly/theoretical writings of the Banū Mūsā, according to G R Tibbetts:

> One feels from reading [the book] that [al-Jazarī] was more interested in the craftsmanship necessary to construct the devices than in the technology which lay behind them. His work excels in dealing with the minute details of construction right down to the actual methods of grinding valves and bearings, and to the elaborate painting and polishing which he recommends. His machines are often assembled by trial and error rather than by theoretical calculation.[61]

Hill, who translated both the al-Jazarī and the Banū Mūsā works into English in the 1970s, traces their source material in his 1994 article "From Philo to Al-Jazarī,"[62] grounding many of the technological advances of the Islamic Golden Age to translations and adaptations of the Greek and Roman engineers of antiquity (Ctesibius, Philo, Hero, and Vitruvius). According to Hill, the bulk of our knowledge of Hellenistic technology comes from the Arabic versions created during this time period.[63] Furthermore, Hill claims that due to its comprehensive nature and exhaustive and precise descriptions, al-Jazarī's 1206 treatise is, without exaggeration,

---

[58] Ibid., 19-20.
[59] Eshad Masood, *Science and Islam: A History* (London: Icon Books, Ltd, 2006), 161–163.
[60] Donald R. Hill in his annotations to Ibn-Shākir, *The Book of Ingenious Devices*, 17.
[61] G. R. Tibbetts, "Review: The Book of Knowledge of Ingenious Devices," *Bulletin of the School of Oriental and African Studies, University of London* 38, no. 1 (1975); 152. Hill further refers to al-Jazarī as "a technical man" in his discussion of Banū Mūsā in Ibn-Shākir; *The Book of Ingenious Devices*, 17.
[62] Reprinted in Hill and King, *Studies in Medieval Islamic Technology*, section III.
[63] Hill, "From Philo to Al-Jazarī," in *Studies in Medieval Islamic Technology*, Section III, 189-190; see also Hill, "Arabic Fine Technology and its Influence on European Mechanical Engineering," in Ibid., section IV, 25-26. Hill also notes that "al-Jazarī almost invariably used the best technique available, whether it was his own invention or not" (Ibid., section III, 205).

> the most important engineering document, from any cultural area, before the Renaissance. He was writing with the avowed intention of enabling later craftsmen to reconstruct his machines, and in this he was successful…. Indeed, few European treatises before modern times give such accurate specifications and dimensions for the benefit of constructors.[64]

Hill goes on in "Arabic Fine Technology" to demonstrate al-Jazarī's influence on European technology through the analysis of one such machine—the mechanical clock—showing the Arabic antecedents for nearly all the elements of its design.

Among the devices chronicled in al-Jazarī's work is a musical automaton, a boat with four automatic musicians that floated on a lake to entertain guests at royal drinking parties. This automaton (four figures in one) is an early example of a programmable robot according to Noel Sharkey, and was capable of playing a number of different tunes based on the arrangement of pegs within its drum machine.[65] In his search for the earliest programmable robots, Sharkey re-created a portion of al-Jazarī's boat of musicians for a 2007 episode of *Ancient Discoveries* on the History Channel. In an article for *New Scientist* the same year, Sharkey explains his choice of the al- Jazarī automaton:

> On board the boat were four mechanical musicians: two drummers, a harpist and a flautist. Al- Jazarī described how they would burst into life every half-hour and play music for a few minutes, continuing in this fashion for several hours without intervention.
>
> Their motion appeared to depend on a primitive version of the studded cylinder in a music box. Running down the length of the boat, beneath the musicians, was a

---

[64] Ibid., section III, 206. For more information, see Ismā'īl ibn al-Razāz al-Jazarī, *Kitāb Fī Ma'rifat Al-Ḥiyal Alhandasiyya* [*The Book of Knowledge of Ingenious Mechanical Devices*], trans. Donald R. Hill (Dordrecht: Reidel, 1974). An online facsimile of the original Arabic text is available at http://www.ebuliz.com/orijinalkitap/, where the detail of al-Jazarī's mechanical illustrations can easily be seen.

[65] Claim made by Noel Starkey in "Robots," *Ancient Discoveries*, season 2 episode 10, directed by Brian Mole, aired 17 April 2007 (The History Channel 2007).

> cylindrical beam with pegs protruding from it. As the beam rotated, the pegs would strike levers connected to the limbs of the musicians, creating life-like motion. The beam was driven by a small waterwheel, turned by water from a tipping bucket, which was refilled automatically every 30 minutes by drips from a reservoir on board the boat. This arrangement seemed to me an ideal candidate for a programmable device. Drill holes all around the beam and the musicians could be reprogrammed to create entirely different rhythmic patterns simply by rearranging the pegs. In 2006 I built a model of one of the boat's percussionists to demonstrate how this could work.[66]

Sharkey's recreation was successful, and illustrated the ingenuity of the design: a programmable figure based on the ideas found in Hellenistic texts (namely Hero's automatic theatre discussed above), but adapted and improved upon by al-Jazarī by making it able to repeat its actions without intervention (through the music-box "program" and self-refilling water bucket).

Al-Jazarī designed a number of other automata as well, including a waitress who could serve drinks such as tea or water and a bath attendant who provided soap, water, and towels for patrons to wash their hands. Again, none of his original automata survive, although we can be reasonably certain they did exist at one time: al-Jazarī repeatedly mentions in his book that he only describes devices he has built himself—some of which were original designs, others adapted from predecessors. However, like many early Greek and Roman designs, recreations of numerous automata of the Islamic Golden Age are housed in museums across the globe; the largest collection of this work can be seen in the Museum of Science and Technology in Islam, in what museum curators label a "celebration of the Golden Age of

[66] Sharkey, "I, Ropebot." Sharkey recreated this boat of musical automata for the episode to illustrate its programmability. In the process, however, he became interested in the origins of al-Jazarī's designs, eventually leading Sharkey to the conclusion that the mechanics of al-Jazarī's automaton were adapted from the earlier designs of Hero of Alexandria (see page 44 above). Prior to Sharkey's investigations, the oldest programmable automaton was believed to have been built by da Vinci in 1478.

Figure 6. Illustration of a musical automaton boat from al-Jazarī's treatise on automata, Kitab fi ma'ari-fat al-hiyal al-handasiya *(*The Book of Knowledge of Ingenious Mechanical Devices*, 1206).*

Islam."[67] Some examples of recreated automata at the museum include al-Jazarī's *Elephant Clock*, a water clock that is at its heart a Greek *clepsydra* in the form of an elephant and rider, and his *Castle Clock*, a complex astronomical clock eleven feet high, complete with phases of the moon and dancing automata musicians, which may well have been part of the inspiration for the famous German cathedral clocks of the fourteenth and fifteenth centuries.

Very few of Banū Mūsā's or al-Jazarī's creations depicted the human form, however. While it may be tempting to ascribe this fact to Islamic aniconism (the ban on the use of the human figure in art),

[67] A part of the King Abdullah University of Science and Technology in Saudi Arabia, the museum itself is accessible only to university members and their invited guests as of this writing. However, many of the exhibits can be viewed virtually at http://museum.kaust.edu.sa/.

it seems that this ban is not as widespread as is commonly believed.[68] While Islamic religious art adheres strictly to this ban, this is not the case in Islamic secular art, particularly in the more liberal regions. Indeed, by the eleventh century outside influence from the Arab world had diluted this prohibition; by the time of al-Jazarī's creations in particular, depictions of the human figure were relatively common in painting, wood, and metalwork, so long as they were obviously secular—which meant they were quite small, usually arranged in groups so as not to focus on a single figure, and did not depict particular individuals or narratives but rather types or emblems.[69] Al-Jazarī's musical group is of this tradition: it is designed as a curiosity and an entertainment; its purpose is clearly not idolatrous. Additionally, the majority of al-Jazarī's work, like that of many of his contemporaries in the Arab world, was devoted to practical automata that improved quality of life (such as his Elephant Clock or his washbasin with automatic soap and towel); humanoid automata designed solely for entertainment were rare

Figure 7. Al-Jazarī's Castle Clock, shown (left) as an illumination from his 1206 manuscript, and (right) as recreated in the Ibn Battuta Mall in Dubai. Still from Ancient Discoveries*, 2.10, "Robots" (History Channel, 2007).*

[68] Aniconism, idolatry, and the ban against "graven images" are not unique to Islam; this belief is common among the Abrahamic religions, although the definition of exactly what constitutes idolatry varies widely among different sects and time periods. See Hersey chapter seven, "Demons and Deliverance," in *Falling in Love with Statues,* for a discussion of statue-love as idolatry in early Christian belief; see also Reilly chapter one, "Iconoclasm and Automata," in *Automata and Mimesis*, for a discussion of the power (and the danger) of the icon during the English Reformation.

[69] Terry Allen, "Aniconism and Figural Representation in Islamic Art," in *Five Essays on Islamic Art* (Manchester, MI: Solipsist Press, 1988).

exceptions. Roboticist Mark Rosheim, who coined the term *anthrobotics* to describe his research and development work in the latest generation of robots with human-like actions, attributes this difference between the practical nature of Arab automata and the more theoretical (and likely unrealized) designs of the Greeks thusly:

> Unlike the Greek designs, these Arab examples reveal an interest, not only in dramatic illusion, but in manipulating the environment for human comfort. Thus, the greatest contribution the Arabs made, besides preserving, disseminating and building on the work of the Greeks, was the concept of practical application. This was the key element that was missing in Greek robotic science.[70]
>
> The Arabs, on the other hand, displayed an interest in creating human-like machines for practical purposes but lacked, like other preindustrial societies, any real impetus to pursue their robotic science.[71]

While humanoid automata represented great technological achievements for al-Jazarī and his contemporaries, they were curiosities meant for the entertainment and comfort of the few. However, their work served as an inspiration throughout Europe during what medievalists term the "renaissance of the twelfth century" of the High Middle Ages, during which increased contact between Byzantium and Europe brought the previously "lost" works of ancient Greece and Rome (such as those of Hero) back into Western knowledge through their Arabic translations and adaptations.

FROM ANTIQUITY TO ARTOIS: THE GARDEN OF HESDIN

The earliest automata to be built in Europe were at the Garden of Hesdin in northern France. The extensive grounds of the Garden contained the spectacular "Pavilion du Mares," commissioned by Count Robert II of Artois in the late thirteenth century, which housed a number of automata, including

---

[70] Rosheim, *Robot Evolution*, 9.
[71] Ibid., 36.

wooden statues which spouted water and groups of animated monkeys which waved at approaching guests. There is a great deal of evidence to suggest that the work of al-Jazarī and other engineers of the Arab world greatly influenced the design and construction of these automata, although the origin for their design has been the subject of debate. As American historian Silvio Bedini rather unambiguously claimed in 1964, the technology necessary for the automata at Hesdin has clear origins in antiquity:

> There appears to be no longer any question, on the basis of recent research, that the mechanical clock and fine instrumentation evolved in a direct line without substantial change from the mechanical water clocks of the Alexandrine civilization, transmitted through Islam and Byzantium from a tradition that may have originated in China, that reached Europe in the twelfth and thirteenth centuries.[72]

It is also generally accepted that Robert II, returning from the Crusades, first learned of the possibilities of automata from Arab engineers during his extensive travels through Naples and Sicily (particularly Palermo), which were at the time Islamic kingdoms. However, in her 1986 article "Reality and Literary Romance in the Park of Hesdin," Anne Hagopian van Buren denies any connection between the Arabic and European traditions of automata; instead, she attributes the development of automata in Europe to a combination of technological advances already present in northern France, and the inspiration of contemporary French romances—basing her claims largely on the facts "that the financial accounts of Artois contain no Islamic names and that al-Jazarī's work is not in the inventory of any medieval library."[73]

Van Buren's reluctance to ascribe Islamic influence to the construction of the Garden of Hesdin, and by extension the spread of automata throughout medieval Europe, appears to be in the minority.

---

[72] Silvio A. Bedini, "The Role of Automata in the History of Technology," *Technology and Culture* 5, no. 1 (1964), 29. Bedini here is referring to this technology in general, and not the Garden of Hesdin in particular, in his tracing of technological development from antiquity to the Renaissance. See also Donald Hill's annotations to Ibn-Shākir, *The Book of Ingenious Devices*, and al-Jazarī, *The book of Knowledge of Ingenious Mechanical Devices.*

[73] Ann Hagopian van Buren, "Reality and Literary Romance in the Park of Hesdin," in *Medieval Gardens*, ed. Elizabeth MacDougall (Washington DC: Dumbarton Oaks Research Library and Collection, 1986), 118.

Both Naomi Miller (1986) and Elly R. Truitt (2010) reject van Buren's position, citing the abundance of connections between Robert II and the Islamic world, including the notebooks of Villiard de Honnecourt,[74] which depict similar constructs sketched during his extensive travels as a mason, and the extensive literature of the time which "often depicted the pleasure gardens of oriental caliphs and emirs,"[75] almost certainly referring in part to the magnificent gardens of Baghdad. In addition, Robert's lieutenant and eventual overseer of the garden's construction, Renaud Coignet—who even van Buren admits held much influence over Robert at the time—was "from Apulia, part of the Kingdom of Sicily, where Muslim and Christian culture, aesthetics, philosophy, and science had been enmeshed for over two hundred years."[76] Miller pragmatically employs Occam's Razor in ascribing Arab influence to the automata at Hesdin. After a lengthy discussion of the work of both the Banū Mūsā and al-Jazarī, citing their prominence in Arab engineering of the time, Miller simply states, "How Islamic ideas were transmitted to Europe is still largely unknown, but contacts between craftsmen in the Iberian peninsula and travelers' reports provide likely explanations."[77] Judging from van Buren's bias throughout the article, her claims against Arab influence appear disingenuous. Baghdad in the late twelfth century was arguably the largest and most well-known city and cultural center in the world, and only the most provincial scholars would not have heard of the engineering feats of the Islamic Golden Age.

Even if the true provenance of the Garden were medieval French romances rather than stories of Islamic automata, as van Buren claims, J Douglas Bruce had already made a connection between these two possible sources in his 1913 article "Human Automata in Classical Tradition and Mediaeval Romance." In this article, Bruce attributes the proliferation of automata in medieval literature to an "oriental" influence and indicates that the relationship between romance and reality was one of cross-pollination rather than cause and effect; according to Bruce, earlier medieval versions of the

---

[74] Naomi Miller, "Paradise Regained: Medieval Garden Fountains," in *Medieval Gardens*, ed. Elizabeth MacDougall, 144.
[75] Elly R. Truitt, "The Garden of Earthly Delights: Mahaut of Artois and the Automata," *Medieval Feminist Forum* 46, no. 1 (2010), 78.
[76] Ibid..
[77] Miller, "Paradise Regained," 143 n23.

Alexandrine legends, taken from Greek and Latin, do not contain the same references to automata as do later ones which follow the Arab translations/adaptations of the work of Heron and Apollodorus.[78] In her own argument, however, van Buren makes no mention of Bruce's earlier work, and while she cites Bedini, she does so only in relation to descriptions of the automata, omitting the claim of Islamic connection cited above. Instead, van Buren positions herself opposite Marguerite Charageat's 1950 article "Le Parc d'Hesdin, création monumentale du XIIIe siècle" ("The Park at Hesdin, monumental creation of the thirteenth century"), claiming that Charageat "sought to derive Hesdin from the fabled park outside Palermo…. These ideas are so attractive that they have been accepted as practically proven, but they must be questioned."[79] I believe that the earlier French romances may have given Robert II the inspiration to create such wonders of his own upon his return from the Crusades. I would not go so far as to claim, however, as van Buren does, that "the ingenious Master Guissin [hired by Robert II to build various mechanisms for the castle hall and gardens] was clearly the *inventor* of Hesdin's first animated statues."[80] While Guissin undoubtedly built the automata at Hesdin, he almost certainly did not invent them independently. The technology used to create these automata reflects a knowledge of the work of Hero as filtered through the translations of Islamic engineers such as al-Jazarī and Banū Mūsā.

The Garden of Hesdin appears to have been quite influential on the development of automata throughout Europe during the centuries following its construction. According to Truitt, Robert's daughter Mahaut (Countess of Artois following her father's death in 1302 until her own in 1329) "spent significant sums on the maintenance, renovation, and installation of automata in the park at Hesdin,"[81] and it was through her connections with the French court that Hesdin became a frequent site of royal

---

[78] J. Douglas Bruce, "Human Automata in Classical Tradition and Mediaeval Romance," *Modern Philology* 10, no. 4 (1913), 518ff.
[79] Van Buren, "Reality and Literary Romance," 117-118.
[80] Ibid., 127 (emphasis mine).
[81] Truitt, "Garden of Earthly Delights," 74. For an extensive accounting of the later restoration of Hesdin by Mahaut's descendants, including details fo the automata present, see Richard Vaughan, *Philip the Good; the Apogee of Burgundy* (New York: Barnes & Noble, 1970), 137-139.

visits, which continued after her death.[82] Truitt attributes the extensive collections of automata found among the estates of the Valois kings (dating back to the 1320s) to the influence of Mahaut and Hesdin:

> it is likely that the automata at Hesdin, expanded under the conservatorship of Mahaut and her heirs, provided them with material, as opposed to textual, examples of the ability of skilled artisans to create objects that, only a few decades prior, could be encountered only in texts or by travel to distant lands. Mahaut's interest in and support of her father's pleasure garden and her desire to expand and modify the automata played an important role in the diffusion of automata and the technology to create them in the fourteenth century.[83]

While the hydraulics used to operate these automata had not changed appreciably since their early descriptions by Hero of Alexandria (10–70 AD), Hesdin seems to have served as a prototype for pleasure gardens across Europe during the fifteenth and sixteenth centuries:

> The grottoes of the European gardens employed the same combination of movements as had been utilized in the most ancient times. The decoration, however, reflected the era; it was executed with considerably more care and with a profusion and confusion of detail and accessories as the Renaissance developed.[84]

One such example is the Enstone Marvels built by Thomas Bushnell in the early seventeenth century in what is now Oxfordshire.[85] Host to King Charles and Queen Henrietta Maria in 1636, Bushnell put on a court masque in which the royal couple was "greeted by a hydraulic automaton of a Hermit," as well as other hydraulic marvels, which pleased the Queen so much that she insisted the waterworks be

---

[82] Ibid., 78-9.
[83] Ibid.. Hesdin was strongly connected to the Valois dynasty throughout its history, as the County of Artois was held by Robert II's descendants in the House of Valois-Burgundy until its loss at the Battle of Nancy in 1477, and remained a popular retreat for royalty throughout the period. Most notably, Philip III, Duke of Burgundy (1396–1467), who came into possession of Hesdin through his marriage to Robert II's descendant Bonne of Artois, spent considerable sums restoring the Garden, and the automata, during his lifetime.
[84] Bedini, "Role of Automata," 26.
[85] Reilly, *Automata and Mimesis*, 44-45.

renamed in her honor.[86] Reilly further claims that the Enstone Marvels and its waterworks (the "Rock called Henrietta") were such key figures during this period that "the pilgrimage to monasteries by Catholics was replaced by the journey to country houses with elaborate pleasure gardens."[87]

The literature of the period reflects the popularity of the Enstone Marvels and other waterworks displays in England; while these images were subject to the shifting opinions regarding idolatry, they nonetheless survived the Reformation, according to Reilly:

> Statues that were smashed in the church by Iconoclasts are re-membered (literally put back together) in the secular theatre and royal pleasure gardens in the form of hydraulic moving statues. These moving statues become a trope in the unfolding secular drama, indicating ongoing cultural anxieties about the persuasive power of images. Venerated Catholic religious icons changed into the idols of the Iconoclasm movement, but these idols, over the passage of time, seemingly became the classical moving statues of the secular theatre and the court masque, as well as the hydraulic automata of the pleasure gardens.[88]

Similarly, it is quite likely that the Garden of Hesdin's popularity among the French court during the medieval period played a large role in the proliferation of automata within later French romances. Certainly the descriptions in the romances reflect the actual operation of the Hesdin automata: in discussing these literary examples of automata, Bruce distinguishes between their function in the Greek legends and that in the later romances:

> It will be observed that the mediaeval automata, unlike those in Homer and Apollonius, have only a very limited intelligence, if intelligence it can be called. They are created for

[86] Ibid., 45.
[87] Ibid., 47.
[88] Ibid..

> some single function—usually to guard an entrance—and nothing more. In this regard they resemble more nearly similar figures in oriental stories.[89]

Far from the intelligent helpmates illustrated by Hephaestus' golden maidens, the automata of medieval and Renaissance literature reflect instead those that actually existed in the gardens of the time: limited in motion, rooted to a single location, and driven by waterpower or clockwork.

Though limited in their actions, the automata of medieval literature were widespread in reputation, not only in French romances but across the continent as well. The earliest of these stories centered around legends of the "great magician" Virgil, stories which by the High Middle Ages had elevated the Roman poet to near-mythological status as the founder and protector of Naples who, among other wonders, "protected Naples from the volcanic Vesuvius by setting up to oppose it a bronze statue of a man with bent bow and arrow ready to let fly."[90] Virgil's wondrous mechanical/supernatural creations were during this period truly the stuff of legend: accounts of his creations ranged from those of Alexandri of Neckam (English, 1157–1217), who John Spargo in the 1930s credited with the initial attribution of these wonders to Virgil,[91] to the anonymous *Les faits merveilleux de Virgille* of sixteenth century France, which was translated into a number of languages across Europe. The mechanical wonders attributed to Virgil range from a set of statues built in a palace in Rome to herald and manage unrest in the empire to an "artificial prostitute" built for the Romans' pleasure.[92] Other fantastical accounts of medieval automata include John Mandeville's *Travels* (1357–1371), a sort of fictional travelogue based on a wide range of actual accounts, including that of the missionary Odoric of Pordenone (ca. 1286–1331) describing the wonders he saw while visiting the imperial court of the

---

[89] Bruce, "Human Automata," 524.
[90] Cohen, *Human Robots in Myth and Science*, 51.
[91] John Webster Spargo, *Virgil the Necromancer: Studies in Virgilian Legends*, Harvard Studies in Comparative Literature, Vol. X (Cambridge: Harvard University Press, 1934). See also Bruce, "Human Automata."
[92] Cohen further states: "Neckham relates, in his *De naturis rerum*, that Virgil built a splendid palace at Rome with statues representing each of the subject provinces. In the hand of each statue was a bell which rung when the province it represented contemplated revolt. At this moment a bronze knight on the roof of the palace brandished his spear in the direction of the province, thus alerting the Romans to their danger" (52). Cohen, *Human Robots in Myth and Science*, 51-53.

Mongol Empire.[93] Chaucer describes more wonders of the Mongol world in *Canterbury Tales*: the Squire's Tale includes an account, derived from an *Arabian Nights* story, of a "steede of bras" given to Genghis Kahn as a gift from the kings of Arabia and India. This brass horse, according to the tale, can "not only become invisible but also take its rider to any place in the world within a day by either running or flying, as the rider wishes, through the manipulation of a series of 'pyns' located in the horse's ear."[94] While these are not artificial humans per se, they illustrate the variety of automata imagined during the period, and highlight the historical thread that winds from ancient Asia through the Islamic world to medieval Europe.

AUTOMATA, HOMUNCULUS, GOLEM

Any sufficiently advanced technology is indistinguishable from magic.

– Arthur C Clarke

That the idea of the artificial human in all its forms was rampant among scholars during the High Middle Ages is evident from (and partially due to) a catalogue thereof compiled by Cornelius Agrippa in his 1533 compendium *De Occulta Philosophia libri tres*, a work that synthesizes magic, religion, and science in what is widely regarded as the most scholarly and authoritative "grimoire" of the period.[95] Agrippa was in fact the first scholar of the medieval period to use the obscure Greek term *automata* to describe the artificial wonders he catalogued; it is his revival of the word in *De Occulta Philosophia* that gives us our modern usage, though at the time he was merely quoting Aristotle's *Politics* in reference to Hephaestus' creations. However, Agrippa used the term *automaton* in his catalogue to signify an *artificial* mobile device, a departure from the original Greek definition of the automaton as *any* self-

[93] Odoric stayed in Taydo at Cambalech (modern-day Beijing) from 1324 to 1327 as a guest of Emperor Yesun Temur, whose palace included golden peacocks which would dance and flap their wings when members of the court clapped their hands. Kang, *Sublime Dreams*. 66.
[94] Ibid., 67.
[95] See Ibid., chapter two: "A monumental survey of magical knowledge of his time, it [*De Occulta Philosophia*] was organized under a scheme that was thoroughly informed by Hermetic and Neoplatonic ideas. The comprehensive nature of the work made it the central text of reference on all things magical during the Renaissance; becoming the most popular book on the subject" (84). I will return to this discussion of the synthesis of magic and science later in this chapter.

moving entity. It is this narrower definition of the automaton that has since become the predominant one.[96] Agrippa catalogues early in Book 2 of his compendium a number "of the many wonderfull works which are done by Mathematicall Arts only" in his discussion of the importance of mathematics and its connection to magic:

> yet nevertheless without naturall vertues, of Mathematicall Doctrines only works like to naturals can be produced, as *Plato* saith, a thing not partaking of truth or divinity, but certain Images kin to them, as bodies going, or speaking, which yet want the Animall faculty, such as were those which amongst the Ancients were called *Dedalus* his Images, and αυτοματα, of which *Aristotle* makes mention, *viz.* the threefooted Images of *Vulcan,* and *Dedalus,* moving themselves, which *Homer* saith came out of their own accord to exercise, and which we read, moved themselves at the feast of *Hiarba* the Philosophicall Exerciser: As also that golden Statues performed the offices of Cup bearers, and Carvers to the guests. Also we read of the Statues of *Mercury,* which did speak, and the wooden Dove of *Arthita,* which did fly, and the miracles of *Boethius,* which *Cassiodorus* made mention of, *viz. Diomedes* in Brass, sounding a Trumpet, and a brazen Snake hissing, and pictures of birds singing most sweetly.[97]

Agrippa's catalogue was used as a popular reference on automata throughout the Middle Ages, as well as revised and expanded upon by later scholars; indeed, it is difficult to overestimate the influence of Agrippa's work to the history of the artificial human. In his discussion of the importance of magic to medieval philosophy, Kang analyses this list and identifies Agrippa's sources for each of the seven most germane objects mentioned. While Agrippa's list includes mainly mythological rather than actual examples of these artificial wonders, it does refer obliquely (through reference to Boethius via

---

[96] Ibid., 85.
[97] Heinrich Cornelius Agrippa, *De occulta philosophia* [Three Books on Occult Philosophy], trans. J.F, Book 2 (London: R.W. for Gregory Moule, 1651). Part 1, emphasis in original.

Cassiodorus) to the automata described (and likely built) by Hero of Alexandria.[98] More importantly, Agrippa's work gives us an insight into the state of philosophical thought during the period, and the transmission of his list through later scholars illustrates the paradigm shift regarding the concepts of magic and science that occurred during the Renaissance.

As mentioned at the beginning of this chapter, our earliest accounts of the creation of Mankind describe the process as mainly a *supernatural* one, although it may include mechanical and/or biological elements. This may be attributed to the fact that, until sometime around the Enlightenment, there was no real distinction between these concepts:

> The distinction between what is magical, typically defined as the use of preternatural power to create extraordinary effects in the world, and what is technological, or the use of machines powered by natural forces, is a modern one that did not begin to be established until the seventeenth century. During the Renaissance, even after many ancient works on mechanics and mathematics were rediscovered and made use of, both these fields were often regarded as magical…. What makes this matter complex is the uncertain status of magic in general throughout the Middle Ages and the Renaissance, as can be seen in major intellectuals' attempts to distinguish different types of it according to their sources, efficacy, and legitimacy.[99]

During the several hundred years after rediscovering these ancient Greek and Roman treatises, it became increasingly difficult to reconcile supernatural explanations with a growing awareness of technological possibility gained from empirical study. Intertwined with this dissonance is the ambivalent status magic held in the medieval Christian worldview. Even the Bible condemns pagan witchcraft while sanctioning divine miracles, showing that the legitimacy of magic depends on the *who* and the *how* of it. Kang argues that, from the medieval Christian point of view, magic can be seen in a number of ways:

---

[98] Kang, *Sublime Dreams*. 88-89.
[99] Ibid., 56.

- Pagan superstition and tricks
- Real, but demonic or Satanic
- Real and divinely sanctioned
- Real and neutral (neither demonic nor divine) from the natural world[100]

Kang's analysis highlights the difficult position of those who sought knowledge during this period. In the early Middle Ages the leading scholars were typically clergymen, and the Catholic Church both developed and supported the schools of advanced learning which became the prestigious universities of medieval Europe (including Bologna, Oxford, Cambridge, Padua, and many others). Theological scholars such as Albertus Magnus (1193–1280) and his contemporary Roger Bacon (1214–1294), for example, openly advocated a form of what they called "natural magic" which falls under the last category outlined above.[101] This form of magic involved the manipulation of natural forces, and led to explorations into the fields of alchemy and astrology, as well as the earliest medieval forays into medicine and chemistry. As time progressed, however, universities across Europe attracted an increasing number of lay scholars to turn their attention to the study of the natural world. This is largely a result of the more stable economic and social environment of the time:

> In conjunction with the expansion of intellectual life in the period, the stabilization of the feudal system gave rise to what some scholars have called the medieval industrial revolution…. In other words, the relative stability of western Europe in the High Middle Ages, along with the intellectual openness and technological innovations of the period, produced a culture secure enough to consider the categories of its worldview under a critical light.[102]

---

[100] Summarized from ibid., 57.

[101] See Glaser and Rossbach, *The Artificial Human*, 15-20 for a discussion of early Renaissance thinkers, including Sir Francis Bacon and Ficino, as "Gods on Earth" using their "creativity to exploit the forces and resources of nature, and in that sense vying with nature and its traditional Creator in productivity" (Ibid., 15). *Prometheus'* Peter Weyland would have been perfectly at home in this intellectual environment.

[102] Kang, *Sublime Dreams*, 62-63.

Whereas earlier Church scholars studied the natural world largely for religious purposes (to determine the proper times to pray, for example), these lay scholars were generally more interested in a more secular understanding of the world around them.

As the pursuit of knowledge became increasingly secularized, many scholars were threatened with excommunication for their pursuits. The University of Paris enacted a series of Condemnations of heretical teachings during this period: the first in 1210, against the works of Aristotle, and culminating in 1277 against a list of 219 propositions drawn from a number of sources that were considered antithetical to Church doctrine.[103] The penalty for teaching, or even hearing, any of these concepts was automatic excommunication, with the threat of Inquisition if the heresy persisted. While the actual ability of the Church to enforce such a Condemnation is debatable,[104] certainly the threat itself had an effect on many scholars:

> In the course of the XIVth and XVth centuries, the nominalist movement spread to many universities and became solidly entrenched. The nominalist emphasis on empiricism and the concomitant decline of metaphysics turned philosophy away from theology "to the positive sciences and to the mathematical disciplines." But the shadow of the ever-present condemnations hovered over many of the scientific discussions in the later Middle Ages.[105]

While a number of scholars throughout this period were either actively or implicitly discouraged from making their claims public for fear of accusations of heresy (Galileo in 1615 stands as a prime example),

---

[103] These propositions included such claims as “That human acts are not ruled by the providence of God,” “That there was never a first human,” and others derived from such widely varying sources as Aristotle, Averroes, and Thomas Aquinas. The fact that many Aquinian philosophies were included led to the later annulment of the Condemnation. Richard E. Rubenstein, *Aristotle's Children: How Christians, Muslims, and Jews Rediscovered Ancient Wisdom and Illuminated the Dark Ages*, (Orlando, FL: Harcourt, 2003), 216-217.

[104] The Condemnations appear to have been limited to Paris, and indeed at universities such as Toulouse and Oxford these concepts were openly taught and discussed. See Edward Grant, *A Source Book in Medieval Science*, Source Books in the History of the Sciences (Cambridge, Mass.: Harvard University Press, 1974).

[105] *Nominalism* is the metaphysical viewpoint, based in empiricism, in which abstract, universal concepts do not exist—only what exists in space and time is real—and hence a rejection of Platonic forms. Edward Grant, “Late Medieval Thought, Copernicus, and the Scientific Revolution,” *Journal of the History of Ideas* 23, no. 2 (1962), 206.

Edward Grant argues that the schism between Church and University brought on by the logical consequences of nominalism (i.e., the conclusion that God does not exist) led both to the attempted Condemnations and the later rise of scientific exploration separate from, and increasingly unhindered by, the influence of the Church—a proposition we will return to in a moment.

The automaton, as well as other forms of artificial human, played an important role in this shifting worldview. According to Minsoo Kang, the automaton serves as a conceptual metaphor for the liminal: a thing at once alive and not-alive, "the ultimate categorical anomaly. Its very nature is a series of contradictions, and its purpose is to flaunt its own insoluble paradox."[106] The ability to create an automaton, to create life from not-life, is the métier of the magician, whose abilities relegate him to the margins of society:

> As someone who dwells in the zone between the inside and the outside, the visible world and the hidden, the magician is thought to be capable of tapping into the chaotic power of the world beyond…. Venturing beyond the boundaries of ordinary reality can bring one tremendous power, but to attain it one has to pay the price of becoming a perennial outsider, and risk as well the danger of that power going out of control in one's hands. And an ultimate form of such a power is the divine prerogative of bringing life to lifeless matter.[107]

During the medieval period, with the rediscovery of ancient knowledge, and its expansion by contemporary scholars—all of which could be deemed "magic" by the populace—it became increasingly important for scholars to negotiate successfully the changing perceptions of their chosen fields. When Agrippa created his catalogue of automata in 1533, for example, he described these objects as marvels of "natural magic"—i.e., mathematics—but when John Dee adapted Agrippa's list just a few decades

---

[106]Kang, *Sublime Dreams*, 36. I will return to, and expand on, the automaton's position as a liminal object which "threatens our reality schema in a particularly frightening way" (Ibid., 52) in my conclusion.
[107] Ibid., 36.

later, he categorically denied any connection between mathematics and magic.[108] The position of an advanced scholar in medieval times was a precarious one, as scholars could be lauded for their discoveries at one moment and decried as sorcerers and necromancers the next. It was quite common to ascribe to the greatest scholars the ability to create automata or other artificial beings: Albertus Magnus (a Dominican friar) and Roger Bacon (a Franciscan), as practitioners of "natural magic," have both been credited with the creation of "marvellous" brass heads which spoke prophecies, a tribute to their scholarly achievements.[109]

Automata were not the only form of artificial human to gain popularity during the Middle Ages, however. Attempts to re-enact creation myths during the medieval period followed all three traditions of the artificial human, with automata representing the *mechanical* tradition; biological and supernatural traditions abounded as well. The primary "natural magics" explored during this time were astrology and alchemy, with which most medieval and Renaissance automata were associated. But alchemists such as Paracelsus (1493–1541)[110] explored other means of creating artificial life: *homunculi*, miniature, fully-formed humans that "are not synthetic like automata. They are made of human tissue,

---

[108] According to Dee:

> And for these, and such like marueilous Actes and Feates, Naturally, Mathematically, and Mechanically, wrought and contriued: Apologeticall. ought any honest Student, and Modest Christian Philosopher, be counted, & called a Coniurer? Shall the folly of Idiotes, and the Mallice of the Scornfull, so much preuaile, that He, who seeketh no worldly gaine or glory at their handes: But onely, of God, the threasor of heauenly wisedome, & knowledge of pure veritie: Shall he (I say) in the meane A.ij space, be robbed and spoiled of his honest name and fame? He that seketh (by S. Paules aduertisement) in the Creatures Properties, and wonderfull vertues, to finde iuste cause, to glorifie the Æternall, and Almightie Creator by: Shall that man, be (in hugger mugger) condemned, as a Companion of the Helhoundes, and a Caller, and Coniurer of wicked and damned Spirites? (John Dee, *The Mathematicall Praeface to the Elements of Geometrie of Euclid of Megara (1570)*, Primary Sources from the Scientific Revolution [New York: Science History Publications, 1975]).

It is important to note here that John Dee himself was accused of sorcery several times during his career, and although cleared of the charges, he remained under suspicion throughout his life. Dee therefore had every reason to reject magical explanations for his work.

[109] See Cohen, *Human Robots in Myth and Science*, 29-33; Reilly, *Automata and Mimesis*, 25, 44. Reilly discusses the Bacon myth as a major source of the "moving statue" trope evident in a number of Renaissance-era plays, including Thomas Middleton's *A Game at Chess* (1624), for which Middleton was arrested for idolatry; it was his last play.

[110] While the most common stories of homunculi come from the medieval period, the idea can be traced to Clement of Rome (ca. 250 CE), who in his *Homilies* described a successful attempt by Simon Magus to create a homunculus in a retort (ibid.. 36). See also Glaser and Rossbach, *The Artificial Human*, 112-116.

blood, and bones…; in short, they constitute a poetico-alchemical preview of in vitro fertilization."[111] The homunculus represents the *biological* tradition of artificial human, and reflects Paracelsus' belief that the benefit of alchemy lies not in the transmutation of base metal to gold but rather in the advancement of medicine. Widely known as the father of toxicology, Paracelsus pioneered the use of chemical and mineral treatments in healing, and his body of work comprised the largest alternative (Paracelsianism) to the Galenic medical traditions of the time. In *De natura rerum* (1537), Paracelsus outlines the recipe for creating an artificial creature:

> Let the semen of a man putrefy by itself in a sealed cucurbite [a round glass vessel] with the highest putrefaction of the *venter equinus* [horse manure] for forty days, or until it begins at last to live, move, and be agitated, which can easily be seen. After this time it will be in some degree like a human being, but, nevertheless, transparent and without body. If now, after this, it be every day nourished and fed cautiously and prudently with the arcanum of human blood, and kept for forty weeks in the perpetual and equal heat of a *venter equinus*, it becomes, thenceforth, a true and living infant, having all the members of a child that is born from a woman, but much smaller. *This we call a homunculus*…. Now, this is one of the greatest secrets which God has revealed to mortal and fallible man. It is a miracle and marvel of God.[112]

In *De natura rerum* Paracelsus also describes a process to "revive the dead by way of regeneration and clarification,"[113] though he is careful to attribute these alchemical secrets to the beneficence of God—as such, they are sanctioned forms of "divine magic." While William Godwin in *Lives of the Necromancers* (1835) described a number of alchemists who produced homunculi by

---

[111] Hersey, *Falling in Love with Statues,* 146. Hersey's discussion of "synthetic biomorphic beings" also credits Sir Frances Bacon with his own homunculus recipe (Ibid., 145).
[112] Paracelsus, *De Natura Rerum* (1537), 14.2, in Nicholas Goodrick-Clarke, *Paracelsus: Essential Readings* (Berkeley: North Atlantic Books, 1999), 175, emphasis mine.
[113] Ibid., 174.

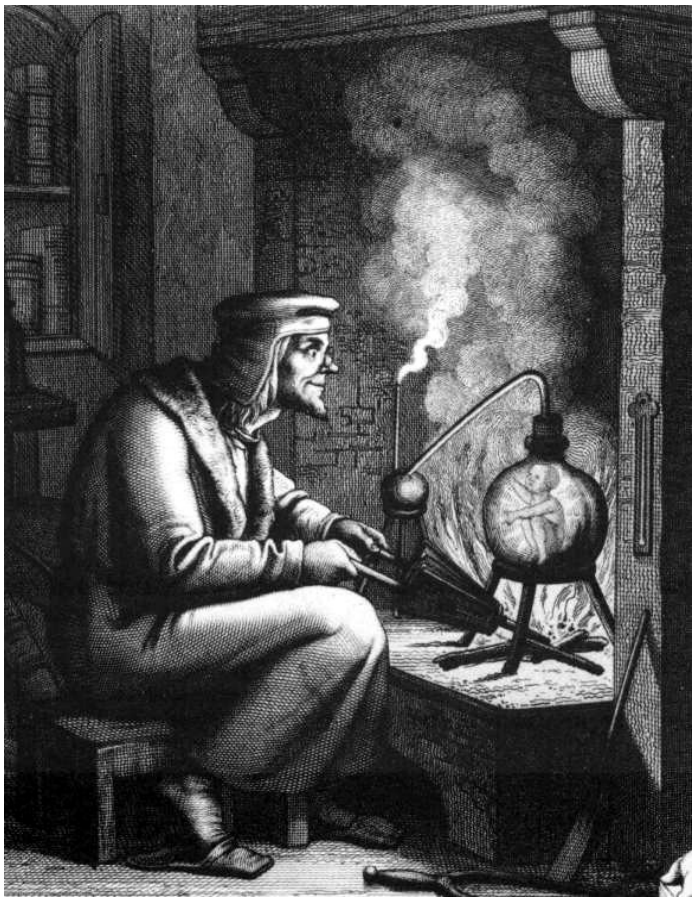


Figure 8. A nineteenth century engraving of Faust creating a homunculus in his laboratory, by Engelbert Seibertz, 1854. From the Bibliothèque des arts decoratifs, Paris, France.

various methods, typically debunking them,[114] the homunculus was revived and popularized by Godwin's daughter Mary Shelly in her novel *Frankenstein*, an account of a medical student who, heavily influenced by the works of Paracelsus and Albertus Magnus, among others, creates his own "superior man" through biological means—at a terrible cost.

Although Paracelsus appears to be the first scholar to use the term *homunculus* in his writing, the concept of a biologically-created artificial being predates his work by seven centuries: we return here to the Islamic Golden Age and the work of Abu Mūsā Jābir ibn Hayyān (c.721–c.815), better known in the western world by the Latinized *Geber*. Geber is held by many to be the first practical alchemist, and the corpus of writing attributed to him and his followers (over 3,000 manuscripts) include such

[114] Hersey, *Falling in Love with Statues*, 146.

diverse topics as medicine, magic, biology, chemistry, metaphysics, and the artificial generation of living beings, or *takwīn*—the ultimate goal of Geber's alchemical investigations.[115] Many of Geber's works were imported into medieval Europe in translation between the eleventh and the thirteenth centuries, and anonymous European writers published additional works under his name (works now attributed to "Pseudo-Geber" to differentiate them from the Jabirian corpus). Paracelsus alludes to Geber in his own work, and it is possible that the concept of *takwīn* influenced Paracelsus' own experiments.

The third type of artificial human hypothesized during the Middle Ages and Renaissance was the *golem*.[116] Unlike the automaton and the homunculus, which were claimed to be a result of the application of "natural magic," the golem lay firmly in the realm of the divine. In the legend of the golem, based in medieval Jewish tradition and strongly indebted to the creation myth of the Torah, a rabbi forms a man of clay, bringing him to life by inscribing the proper Hebrew letters (אמת, or *emet*, meaning *life*) on the creature's forehead, or writing them on a scroll placed inside the creature. While the word *golem* itself is mentioned only once in the Torah, in the book of Psalms (compiled no later than the sixth century CE), the Talmud indicates that Adam himself was a golem, or "shapeless mass," for his first twelve hours. According to Kabbalistic scholar Moshe Idel, only a very holy person could create a golem, and to do so was an indication of one's closeness to God:

> It is known that whoever is an expert in Sefer Yezirah, is able to perform operations by the holy names, and out of the elements, dust of a virgin soil and water, a matter (Go[lem]) and form will emerge, which has a vitality; even so it is called dead (מת, or *met*), since he cannot confer upon it knowledge… and speech, since knowledge and speech are [the prerogative of] the Life of the Worlds.[117]

---

[115] See Syed Nomanul Haq and Ḥayyān Jābir ibn, *Names, Natures, and Things: The Alchemist Jābir Ibn Ḥayyān and His Kitāb Al-Aḥjār* [*Book of Stones*], Boston Studies in the Philosophy of Science Volume 158 (Dordrecht: Kluwer Academic Publishers, 1994).
[116] See Cohen, *Human Robots in Myth and Science*, chapter three.
[117] As quoted from a mid-seventeenth century manuscript in Moshe Idel, *Golem: Jewish Magical and Mystical Traditions on the Artificial Anthropoid*, Suny Series in Judaica (Albany, NY: State University of New York Press, 1990), 207. Amendments in original text.

The creation of a golem is truly supernatural, reserved to those scholars able to unlock the secrets of the *Sefer Yezirah*, the Jewish holy book of creation. In medieval Jewish scholarship, the ability to create a golem relies solely on the faith of the rabbi constructing it rather than any clever engineering.

A number of Jewish stories from throughout medieval Eastern Europe offer accounts of holy men creating golems,[118] beginning with Rabbi Eliyahu of Chełm (1550–1583) and culminating in the famous nineteenth century legend of the Golem of Prague, set in the late sixteenth century, which tells the story of Rabbi Judah Loew ben Bezalel's efforts to protect the Jewish ghettos of Prague from anti- Semitic attacks and pogroms during the reign of Rudolf II of the Holy Roman Empire. According to legend, the

Figure 9. Rabbi Loew animating his golem. Woodcut by Mikoláš Aleš, 1899.

[118] See Cohen, *Human Robots in Myth and Science*, chapter three; Idel, *Golem*..

body of this golem still rests in the Old New Synagogue in Prague, to be called upon for protection at a future time of need. Like the example of Virgil above, both of these rabbis are actual historical figures, renowned scholars of their time; the later attribution of golem-creation to their oeuvres indicates how highly these men and their work were esteemed by later scholars. Not all golem-creators were held in such high esteem, however. The eleventh-century Spanish poet, philosopher, and mystic Solomon ibn Gabirol was claimed to have created a female golem as a servant girl; brought before the King of Spain for crimes of sorcery, he disassembled the golem and was subsequently acquitted.[119]

The golem reflects a darker aspect of the artificial human: it is typically depicted as a dumb brute, of low intelligence, created solely for heavy labor or protection, a slave to the Rabbi who brought it to life. As mentioned before, however, the slave will inevitably turn on its master—as did the golem attributed to Rabbi Eliyahu. According to the autobiography of Eliyahu's son, the scholar Rabbi Jacob Emden (1697-1776), the creature his father created

> was without speech, and he was serving him as a servant. And when the Rabbi saw that the creature of his hands grew stronger and greater, because of the divine name written on a paper stuck to his forehead, R. Eliyahu, the master of the name, was afraid that he would be harmful and destructive. He quickly overcame him and he tore the folio on which the name was written and separated it from the forehead, and he fell as a lump of dust, as he was [in the beginning], but he harmed his master scratching him when he took up the writing [haketav] and separated the name from him.[120]

In fact, most accounts of the creation of a golem include a "safety switch," as you will, to deactivate the golem should it become too strong or restless, or run out of control—which, according to legend, happened often, typically due to some mistake or weakness on the part of its creator. The Golem of Prague, for example, was eventually destroyed by Rabbi Loew after going on a murderous rampage. In

---

[119] Cohen, *Human Robots in Myth and Science*, 38-39.
[120] From autobiography of Rabbi Jacob Emden (1748), quoted in Idel, *Golem*, 209-210. Amendments in original text.

some accounts, the rampage was the result of Rabbi Loew's failure to deactivate the golem on a Friday evening to prevent it from desecrating the Sabbath, while other accounts claim the creature fell in love and was rejected, causing it to become violent.[121]

The figure of the golem throughout literature and legend holds many parallels to that of the robot, particularly the robots of early science-fiction fame. Like the golem, these robots were relatively unintelligent: any command given was likely to be taken perfectly literally, often to dangerous or deadly effect. Robots and golems are also designed to be perfectly obedient, unable to disobey. Isaac Asimov's Three Laws of Robotics[122] is a hierarchy of restrictions designed expressly to assure a combination of obedience and safety from the robots in his stories—a precaution derived from the destructive capabilities of the golems (and other artificial humans) of legend:

> One of the stock plots of science fiction was [that] robots were created and destroyed by their creator…. With all this in mind I began, in 1940, to write robot stories of my own—but robot stories of a new variety. Never, never, was one of my robots to turn stupidly on his creator for no purpose but to demonstrate, for one more weary time, the crime and punishment of Faust.[123]

The "crime" of Faust (see Figure 8 on page 73), as well of that of others who attempted to create their own version of the artificial human, whether golem, homunculus, or automata, was that of hubris—the aspiration to godhood mentioned in the introduction. As such, this aspiration amounted to the usurpation of creation powers rightly reserved to God alone. It is no wonder, then, in the changing political and intellectual climate of the Renaissance, scholars (or "magicians") credited with the creation

---

[121] See Idel, *Golem*, for more detailed accounts of the various golem legends.

[122] Initially introduced in Asimov's 1942 short story "Runaround," the Three Laws have since become one of the most pervasive organizing principles in robotic science fiction. Asimov's Three Laws are as follows:

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 to 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.

[123] Isaac Asimov, "Introduction," in *The Rest of the Robots* (Garden City, NY: Doubleday, 1968). Cf. Cohen, *Human Robots in Myth and Science*, chapter four.

of artificial humans were frequently accused of demonic associations—and found it increasingly necessary to define their work in terms of natural, rather than magical, phenomena.

THE DEATH OF MAGIC

This increasing discomfort with the realm of the magical is strongly connected with the danger of holding unorthodox ideas during the Counter-Reformation of the sixteenth and seventeenth centuries. In her study of the role of the miracle in Early Modern Europe, historian Lorraine Daston points out that Thomas Aquinas, in his *Summa contra Gentiles* of 1264, "drew a principled distinction between the truly *supernatural* (God's unmediated actions) on the one hand, and the *natural* (what happens always or most of the time) and the *preternatural* (what happens rarely, but nonetheless by the agency of created beings) on the other."[124] While automata, homunculi, and golems[125] would all have been categorized as magical creations (arguably of either divine or demonic origin) during the medieval and Renaissance periods, they fell into the category of "preternatural" phenomena according to the hierarchy outlined by Aquinas. From the thirteenth century onward, there was an increased tendency to "segregate the natural and the preternatural from the supernatural,"[126] with the "supernatural" being the realm of miracles performed by (or through the intervention of) God. As the period progressed, preternatural phenomena such as the creation of artificial life (even when only rumor or legend) "came to be ever more closely associated with the dubious and possibly demonic activities of magic and divination."[127] To distance themselves from this association, "natural" philosophers of the period went to ever greater lengths to offer "scientific" explanations for

---

[124] Lorraine Daston, "Marvelous Facts and Miraculous Evidence in Early Modern Europe," *Critical Inquiry*, no. 1 (1991), 97 (emphasis mine).
[125] I would claim that the golem, in contrast to the automata and the homunculus, is a *supernatural* phenomenon according to Aquinas' hierarchy, in that its creation relies on the intervention of God in response to the holiness of its creator. However, given the tendency throughout medieval and Renaissance Europe to be (at best) highly suspicious of Jews and Jewish beliefs, the golem would have more likely been seen as demonic, and Kabbalism denounced as sorcery (as was alchemy, astrology, and a number of other medieval fields of research).
[126] Daston, "Marvelous Facts," 98.
[127] Ibid..

phenomena—revolutionizing the methods of philosophical inquiry in what is now known as the Scientific Revolution.

During the Scientific Revolution of the mid sixteenth to mid eighteenth centuries, the *preternatural*—"particulars and a fortiori singularities of all kinds" (including artificial life) which "were not readily susceptible to scientific explanation"[128]—was increasingly relegated to liminal status, accepted as neither *supernatural* (and thus belonging to theology) nor *natural* (and thus in the realm of scholastic natural philosophy). But while the status of the preternatural was in dispute, the category itself experienced rapid expansion during this time. According to Daston,

> It might be argued that the inherent conceptual instability of preternatural phenomena predestined it for collapse into the sturdier categories of the natural and the supernatural. However, the preternatural was very long in meeting its doom …, expanding in extent and intellectual importance throughout the sixteenth century. Fifty years before its demise around the turn of the eighteenth century, the preternatural preoccupied theologians and natural philosophers more urgently than ever before.[129]

Theologians assigned phenomena that were insufficiently godly or miraculous to the preternatural; natural philosophers assigned any unusual or unexplainable phenomena (which at that time included such things as magnetism) to the preternatural as well. As the corpus of the preternatural grew, these phenomena were increasingly attributed "not just to any remarkable conjunction of natural causes, but to conjunctions of natural causes cunningly wrought by demons."[130] As a result, the preternatural became increasingly conflated with the natural, as theologians warning of "the religious peril of becoming a dupe to a counterfeit miracle" tended to emphasize the natural rather than demonic causes

[128] Ibid., 99.
[129] Ibid..
[130] Ibid., 106.

of these phenomena, "in order to steal the devil's thunder."[131] As the demonic faded, only the natural remained. The liminal category of the preternatural collapsed, and its phenomena gained new status as "the first scientific facts."[132]

Where did this collapse leave the artificial human? It is at this point that the histories of the golem, homunculus, and automaton diverge, based on their suitability to scientific exploration. The golem, reliant as it is upon the favor its creator holds with his God, remains a predominantly supernatural phenomenon—either divine or magical, it cannot be explained through scientific rationality. So while stories of the golem continue through to present day, they are in modern times the realm of fantasy, not science fiction; therefore we leave the golem to the Renaissance. The homunculus is another matter. Based as it is in early understandings of medicine and biological processes, we can see its legacy today in the development of in-vitro fertilization and stem-cell research into cloning and regeneration. However, it is the automaton—the mechanical man—that we can trace most strongly through this paradigm shift: as Minsoo Kang claims,

> natural philosophers took up the automaton [and] purified it of its magical aura through completely mechanical explications of its function…. Once denuded of its magical aura, the automaton was then elevated as the single most important emblem of the emerging worldview that envisioned the world, the state, and the body in mechanical terms. So as rich and varied as the roles the automaton played in the intellectual and imaginative writings of the Middle Ages and the Renaissance, it entered the true golden age of its celebration, popularity, and significant use as both an actual and a conceptual object in the following period.[133]

---

[131] Ibid., 106-107.
[132] Ibid., 109.
[133] Kang, *Sublime Dreams*, 96, 102. Cf. Hersey, *Falling in Love with Statues,* 115: "practically all the sculptural concepts we've been talking about—lifelike realism, tacitility, Lucretian simulacra, and so on—turn from life-enhancing traits, statue-life that is artificial in a completely positive sense, into an evil that is still nonetheless 'real.' We almost never hear that the pagan statues

The worldview described here is that of mechanical philosophy, or the branch of natural philosophy emerging in the late seventeenth century that held that all phenomena could be explained in terms of deterministic "mechanical laws." While the concept of universal mechanism is first articulated in the early modern period in Hobbes' *Leviathan* (1651), it is through the work of René Descartes that the idea arises that living beings could be explained completely according to mechanistic principle—that man is, basically, an automaton.

---

are mere dead matter." Hersey is addressing here the individual statues, from antiquity and otherwise, while Kang focuses on the automaton as archetype.

CHAPTER TWO: THE CRISIS OF THE MAN/MACHINE

I suppose the body to be just a statue or a machine made of earth, which God forms with the explicit intention of making it as much as possible like us…. We see clocks, artificial fountains, mills, and other similar machines which, even though they are only made by men, have the power to move of their own accord in various ways. And, as I am supposing that this machine is made by God, I think you will agree that it is capable of a greater variety of movements than I could possibly imagine in it, and that it exhibits a greater ingenuity than I could possibly ascribe to it.

René Descartes, "Traité de l'homme"[1]

At the beginning of the seventeenth century, the Royal Gardens at the Château de Saint Germain-en-Laye were "the highest peak of development" in hydraulic automata.[2] Commissioned by King Henry IV during his transformation of the Château into his principal residence, the gardens were designed by Florentine architect and mechanician Tommaso Francini (1571–1651). Henry IV "borrowed" Francini in 1597 from his Italian cousin[3] the Grand Duke Fernando I de' Medici, for whom Francini had designed extensive garden displays at the Villa Medicea di Pratolino near Florence. Modeled after the waterworks made popular among French royalty by the Garden at Hesdin, while reflecting the style of the period, the grottoes at Saint Germain-en-Laye featured mythological subjects:

> The first three grottoes opened from the third landing, or Doric Gallery, and featured a Dragon, an Organ Player, and Neptune. On the fourth landing the grotto of Hercules was flanked on either side by the grottoes of Perseus and Andromeda and of Orpheus. The largest of the grottoes was that of Perseus, in which the fully armed hero descended from the ceiling and killed with his sword a dragon that arose from the great basin of

[1] René Descartes, "Traité de l'homme" ["Treatise on Man"], in *The World and Other Writings*, translated and edited by Stephen Gaukroger (Cambridge: Cambridge University Press, 1998), 99.
[2] Silvio A. Bedini, "The Role of Automata in the History of Technology," *Technology and Culture* 5, no. 1 (1964): 27.
[3] I use the term *cousin* here in the royal sense: Henry IV was married at the time to Margaret of Valois, daughter of Catherine de' Medici. Henry was later married to Marie de' Medici, niece to Ferdinando I. Thus through several threads of royal lineage can Saint Germain-en-Laye be traced back through the House of Valois to Robert II of Artois and the Garden at Hesdin.

> water. At one side Andromeda was chained to a rock, while in another corner Bacchus sat drinking upon a barrel. Similar scenes, based on different themes, were presented in the other grottoes.[4]

In the summer of 1614 a young graduate of the Jesuit College at La Flèche lived in the village of Saint Germain-en-Laye. Trained in the sciences, this young man was undoubtedly drawn to the marvels of the Royal Gardens, since abandoned by the royal couple in favor of the Palace of Fontainebleau. It is easy to picture this melancholy young man strolling in the Gardens, observing the complex and clever function of the automata there.

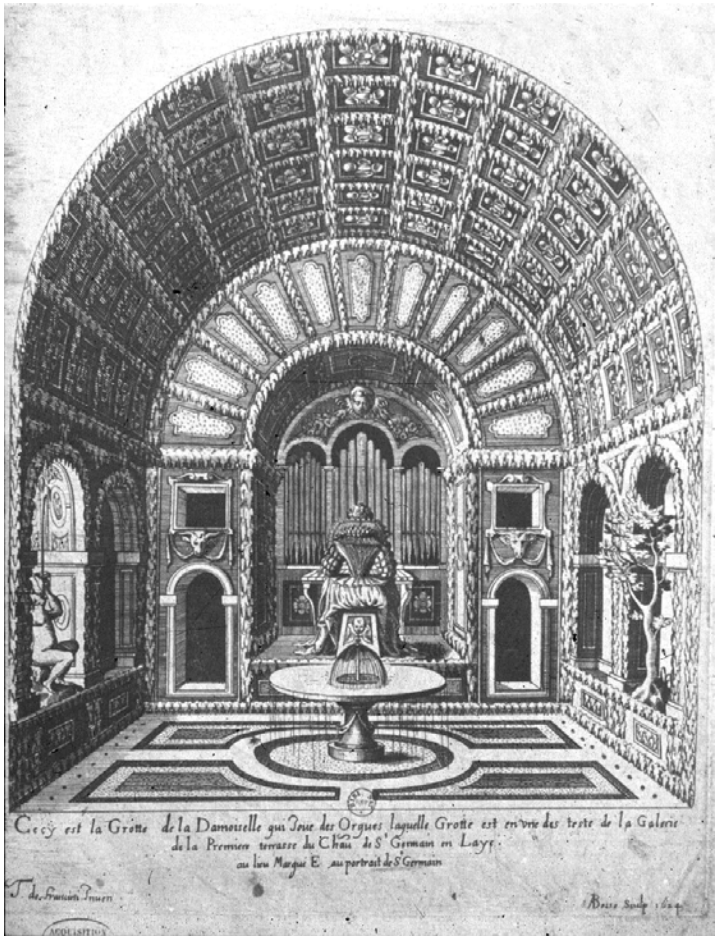


Figure 10. Engraving depicting one of the automaton grottoes at the Château de Saint-Germain-en-Laye, by A. Francini, 1614. © Musée d'archéologie nationale, Saint-Germain-en-Laye.

[4] Bedini, "Role of Automata," 28.

The young man in question here is none other than René Descartes, and while we cannot be certain, there is a great deal of circumstantial evidence to support this scenario, according to Stephen Gaukroger:

> We have no reason to doubt Baillet's general claim [in 1691] that Descartes lived in Saint-Germain (although the details are admittedly often incorrect), and if so it is almost inconceivable that he would not have visited the grottoes, in which case, as it is clear from the *Traité de l'homme*, reflection on the mechanical figures he found there, at some time between then and, say, the mid-1620s, at least reinforced the plausibility of, and perhaps even initiated, the view that animals are automata.[5]

Traité de l'homme ("Treatise on Man") is a part of Descartes' *Le Monde*, a project that represented an attempt to account for no less than "the whole physical world."[6] Descartes formulated the extant portions of this project—the "Treatise on Light" and the "Treatise on Man"—between 1629 and 1633, although due to the intellectual climate of the time,[7] it remained unpublished until after his death: the section on man as *De Homine* in 1662 and the rest as *Le Monde* in 1664. Referred to collectively as *The World*, this text offers the first systematic presentation of Descartes' mechanistic natural philosophy:

> Descartes' aim is to provide a mechanistic cosmology, resting on the basis of quantitative "laws of nature," and a mechanistic physiology. Among the more fundamental things that he sets out to establish, four are of special significance and novelty. The first is that the stability of planetary orbits and the orbits of their moons

---

[5] Stephen Gaukroger, *Descartes: An Intellectual Biography* (Oxford: Oxford University Press, 1995), 64. Adrien Baillet (1649–1706) wrote the first comprehensive biography of Descartes, drawing on a number of published and unpublished sources. According to Gaukroger, Baillet's biography is at points flawed, but remains "far and away the single most important source of knowledge of Descartes' life after his own writings" (ibid., 471). An extended argument for accepting Baillet's placement of Descartes in Saint Germain-en-Laye, as well as the attribution of melancholia to the young Descartes, can be seen in ibid., 62-64. See also Kara Reilly, *Automata and Mimesis on the Stage of Theatre History* (New York: Palgrave Macmillan, 2011), Chapter 2.

[6] Stephen Gaukroger, in Descartes, *The World and Other Writings*, xxi.

[7] Galileo received Papal Condemnation on June 22, 1633 for the "heresy" of heliocentrism. As *The World* is largely based on heliocentric philosophy, Descartes was quite prudent in his decision not to publish his work at the time. See ibid., xxvi-xxix.; Minsoo Kang, *Sublime Dreams of Living Machines: The Automaton in the European Imagination* (Cambridge, MA: Harvard University Press, 2011), 96-97.

> can be accounted for on a mechanist basis…. The second is that the propagation of light from the Sun can be explained in terms of the centrifugal effects of its axial rotation. The third is that all vital functions can be accounted for mechanistically. And the fourth is that perceptual cognition can be accounted for, at least to a very large extent, in terms of a mechanistic psycho-physiology.[8]

The first two points Gaukroger outlines above form the backbone of Descartes' "Treatise on Light," the first of three sections planned for *The World*.[9] The third point is the focus of the "Treatise on Man," which first articulates Descartes' consideration of the mechanical nature of the living body. The final point seems to connect mechanism to Cartesian dualism, the separation between mind and body that has influenced the idea of the artificial human since the seventeenth century; unfortunately, due to Descartes' abandonment of the project in 1633, the final section on the soul "may not have even been drafted at this time."[10]

THE RISE OF MECHANISM

Mechanism[11] is the philosophy of being that considers all natural systems as machines in which the function of the whole may be explained solely through the interworking of the individual parts. Linked to the pre-Socratic theory of atomism developed by Democritus (460–370 BCE) and his contemporaries, mechanistic philosophy as it developed in the seventeenth century came about as an attempt to explain natural phenomena in purely rational terms, rejecting super- or preternatural causality in favor of scientific laws and observable facts. Stemming from ideas arising during the

---

[8] Gaukroger, in Descartes, *The World and Other Writings*, vii-viii.
[9] Gaukroger indicates that Descartes' treatise *The World* "was designed to cover three topics: inanimate nature, animals and especially the human body, and the 'rational soul.'" (ibid., viii.)
[10] Ibid.. However, Gaukroger theorizes that much the projected content of this missing section may be seen in *Passions of the Soul* (1646), which begins with a summary of "Treatise on Man." (ibid., xix.)
[11] *Mechanism* and *mechanicism* are used more or less interchangeably among various sources at various points in time. For clarity's sake, I have chosen to use the terms *mechanism* and *mechanistic philosophy* in my own discussion.

Copernican Revolution,[12] the universe came to be seen as a giant clockwork, in which the motion of celestial bodies is determined by set physical laws. In an extension of these universal laws, living beings are seen as automata, with their biological functions fully explainable through physical and chemical processes:

> I desire, I say, that you should consider that these functions [including passion, memory, and imagination] follow in this machine simply from the disposition of the organs as wholly naturally as the movements of a clock or other automaton follow from the disposition of its counterweights and wheels.[13]

This is the first articulation of Descartes' *bête machine* (beast-machine) analogy, the central tenet of mechanistic philosophy as understood in the early modern period.[14] Influenced by hydraulic automata such as seen at Saint Germain-en-Laye, Descartes likened the living body to a complex waterwork, in which the various parts are moved by the flow of "animal spirits" through the nerves and veins:

> You may have observed in the grottoes and fountains in the royal gardens that the force that drives the water from its source is all that is needed to move various machines, and even to make them play certain instruments or pronounce certain words, depending on the particular arrangements of the pipes through which the water is conducted.
>
> And the nerves of the machine that I am describing can indeed be compared to the pipes in the mechanical parts of these fountains, its muscles and tendons to various other engines and springs which serve to work these mechanical parts, its animal spirits

---

[12] The Copernican Revolution is the name given to the paradigm shift, began by Nicolaus Copernicus (1473–1543), from the Ptolemaic geocentric model of the universe to the Copernican heliocentric model. This shift spanned over a century, from Copernicus' initial formulation of a heliocentric system, through Johannes Kepler's proof of the elliptical orbits of planets, Galileo's subsequent proof of Copernicus' heliocentrism through the observation of planetary phases, and finally Isaac Newton's laws of planetary motion and universal gravitation (*Philosophae naturalis principia mathematica*, 1687).

[13] Gaukroger, in Descartes, *The World and Other Writings*, 169.

[14] Descartes ultimately develops what is now known as the "*Bête Machine* Doctrine" as a proof that humans are ultimately superior to animals in that animals possess no soul; that is, animals do not reason—a position that has been highly debated in the centuries since, and is largely discounted in contemporary philosophy of mind, to which we will return in chapter three. See Lex Newman, "Unmasking Descartes's Case for the Bête Machine Doctrine," *Canadian Journal of Philosophy* 31, no. 3 (2001): 389-426; see also Leonora Cohen Rosenfield, *From Beast-Machine to Man-Machine; Animal Soul in French Letters from Descartes to La Mettrie*, (New York: Oxford university press, 1941).

> to the water that drives them, the heart with the source of the water, and the brain's cavities with the apertures.[15]

Gaukroger claims that the "royal gardens" to which Descartes refers here are "almost certainly… the Royal Gardens at Saint-Germain-en-Lay."[16]

It is important to note here that while Descartes speaks in radically mechanistic terms regarding the cosmos (in his "Treatise on Light") and the physical functioning of the human (in his "Treatise on Man"); he maintains the existence of a soul as a sort of "ghost in the machine"[17]:

> When a rational soul is present in this machine it will have its principal seat in the brain and will reside there like the fountaineer, who must be stationed at the tanks to which the fountain's pipes return if he wants to initiate, impede, or in some way alter their movements.[18]

This is the first indication of Descartes' version of *substance dualism*, in which there are two distinct types of matter: the physical, which makes up the machine of the human body and is responsible for its biological functioning; and the mental, the "thinking" stuff that consists of a fundamentally different sort of matter than the body in which it is housed. According to Cartesian dualism, the mental has no

---

[15] Descartes, *The World and Other Writings*, 107. Part 2 of the treatise, "How the machine of the body is moved" (ibid., 106-17), goes into great detail in describing how these "animal spirits" move the body through the analogy of pneumatics, further explaining that external objects "which by their mere presence act on the organs of sense and thereby cause them to move in many different ways, depending on the arrangement of the parts of the brain, are like strangers who on entering the grottoes of these fountains unwittingly cause the movements that take place before their eyes" (Ibid., 107)—by stepping on activator plates hidden in the grotto floor, for example.

[16] Ibid., 107 n16. See also Reilly, *Automata and Mimesis* Chapter 2, and Otto Mayr, *Authority, Liberty, & Automatic Machinery in Early Modern Europe*, Johns Hopkins Studies in the History of Technology (Baltimore: Johns Hopkins University Press, 1986), 62-67 for discussions of Descartes' fascination with automata.

[17] The term was coined by Gilbert Ryle in his rejection of Cartesian dualism as a logical fallacy. The "ghost in the machine" is what Ryle calls the "official doctrine" of mind-body dualism: that mind and body are both substances: one material, the other immaterial. Ryle criticizes the labeling of "mind" as a thing—albeit an immaterial one—as a "category mistake" of the same type as the foreigner who, upon being shown a university's libraries, administrative offices, and dormitories, then asks "But where is the university?" (Gilbert Ryle, *The Concept of Mind* [Chicago: University of Chicago Press, 1949], 16). *University* does not belong to the same *category* as the other terms, although the questioner has mistakenly assumed that it does. By analogy, according to Ryle, *mind* is not a thing located within the body (a ghost in the machine) but may be better conceptualized as the sum total of the behaviors of the body. For more information, see ibid..

[18] Descartes, *The World and Other Writings*, 107. Gaukroger points out that "this image of the rational mind comes dangerously close to the idea of the mind as being like a pilot guiding a ship, which Descartes will reject in very firm terms in the *Meditations*" (Gaukroger in Descartes, *The World and Other Writings*, 107 n19).

spatial properties and is therefore expressly non-physical. Conversely, the matter of the physical world has no mental properties: "for to have the power of self-movement, as also of feeling or of thinking, I did not consider to appertain to the nature of body."[19] While many members of the mechanistic school of thought in the early modern period did not agree with Descartes on this matter, Cartesian dualism forms the foundation of the mind-body problem that lies at the heart of modern debates in the philosophy of mind and the field of artificial intelligence.[20] It also prefigures the idea of computer programming (with the "fountaineer") at a time when the emerging worldview was based on mechanistic principles developed prior to the discovery and application of electricity—a philosophy based in hydraulic, pneumatic and clockwork spring-driven mechanisms. For Descartes and many of his contemporaries, all of existence may be explained through the metaphor of the machine, and the human body is no more than a biological automaton. From a thing of magic and wonder, the automaton had "emerged as the most powerful and conspicuous intellectual emblem of the era."[21] To understand this evolution, we must first understand the onto-epistemological shift that occurred in the early modern period, as many of the concepts that inform modern explorations of artificial life have their roots in the prevailing philosophies of the time.

The paradigm shift in worldview that took place in Europe during the early modern period, now known as the Scientific Revolution, witnessed a convergence of several important threads in our history

---

[19] René Descartes, *The Philosophical Works of Descartes,* trans. by Elizabeth Sanderson Haldane and G. R. T. Ross (Cambridge: The University Press, 1911, 1978), 14. The principle that *all* phenomena may eventually be explained by mechanical laws (even if those laws are not yet understood) is known as *universal mechanism*, which in its most radical formulation implies a hard determinism that is incompatible with free will. Otto Mayr theorizes that Descartes' insistence on the presence of a soul may have stemmed in part from his need to mollify the Church at a time when fellow scientists were suspected of heresy:

> It was this doctrine [that the soul resided in the pineal gland and controlled the body from this location] that he chiefly relied upon to draw the distinction between automaton and man and thus to prevent his mechanistic physiology from becoming classifiable as heresy; either to pacify his own scruples or to disarm the suspicions of the Church authorities, he invoked it with great frequency. (Mayr, *Authority*, 65)

Cartesian dualism also appears to be heavily influenced by the Aristotelian concept of the *soul* that was the dominant metaphysical philosophy up until that time.

[20] The mind-body problem of cognitive science and AI affects by extension the "problem" of robots both real and imagined. For the purposes of this dissertation, I have tried to remain agnostic on this issue, arguing instead for a performative definition that neither requires nor rejects the presence of an immaterial soul. I will return to this issue later in my analysis.

[21] Kang, *Sublime Dreams*, 112.

of the artificial human. First, ancient Greek and Roman texts (including those of Hero, Galen, and Vitruvius as mentioned in the previous chapter) were translated and adapted by Arabic scholars of the Islamic Golden Age. These translations found their way to continental Europe during the later Crusades (from the thirteenth century up to and including the fall of Constantinople in the mid-fifteenth century), where their "rediscovery" also led European scholars to make their own translations. Second, advances in scientific method and the collapse of the preternatural as a valid explanation for natural phenomena illustrated an attempt not only to understand these rediscovered texts but also to move beyond them into a new understanding of the world. Simultaneously, this curiosity both enabled and encouraged engineers to push the limits of contemporary technology, to design and build increasingly complex and ingenious machines in order to capture in metal, wood, and skins the mechanical wonders described in the tales of antiquity.

But the hydraulic marvels in the grottoes of medieval European gardens such at Hesdin and Saint Germain-en-Laye, while strongly influencing the development of the mechanistic philosophy that is the hallmark of the Scientific Revolution, are not the only vectors leading to its rise. Two other major developments during this time period play a vital role in the rise of mechanistic philosophy during this period. The first is the evolution of mechanical clockwork, which not only offered a revolutionary advance over the hydraulic and pneumatic systems of those medieval gardens but also brought with it and onto-epistemological change in humanity's relationship to the world around it. The second development is the emergence of the field of biology as a viable area of empirical study due to new methods of examination, particularly advances in dissection, which led to a new understanding of the human (and animal) body.

A CLOCKWORK UNIVERSE

The reasons behind the development of the mechanical clock in the early fourteenth century are subject to a certain amount of speculation among historians: Otto Mayr, for example, attributes the

invention to the increased need for precise timekeeping among the clergy[22] while Derek de Solla Price and Silvio Bedini couple the rise of the clockmaking tradition to the proliferation of craft guilds.[23] Regardless of the reasons, the developmental path from the water-clocks of antiquity to the mechanical clockwork of the early modern period is fairly clear.[24] From the ancient clepsydra to the elaborate works of al-Jazarī, clocks were powered hydraulically, with the passage of time indicated by the amount of water that had flowed through the mechanism. In the case of water-clocks such as al-Jazarī's Elephant Clock, the force of this water was then used to power various automata or other entertainments. Unfortunately, due to their size and complexity, these clocks were either large public displays or rare works privately owned by royalty or aristocracy. However, technological advances in clockmaking led to the proliferation of clocks, and their accompanying automata, in the private sphere as well as the public. While the medieval clock was a rarity beyond the reach of all but the wealthiest patrons, the clock of the sixteenth and seventeenth centuries became a commodity and a status symbol. At the same time, the importance of the clock as a mere timekeeper waned in favor of its function as an automaton:

> An important aspect of the appeal of the mechanical clock, which had nothing to do with timekeeping at all, was rooted in the magic of self-moving machines…. Automata performances not only became standard requirements for the larger tower clocks; they were also realized on various smaller scales, down to that of pocket watches. Clocks and automata, in short, tended to be very much the same thing. Contemporaries recognized this clearly: in dictionaries from the sixteenth to eighteenth centuries, the automaton (a

---

[22] Mayr, *Authority*, 3-4.

[23] Bedini, "Role of Automata," 29; Derek J. de Solla Price, "Automata and the Origins of Mechanism and Mechanistic Philosophy," *Technology and Culture* 5, no. 1 (1964): 17.

[24] According to Price,

> the ample evidence of many well-edited texts and a couple of extant instruments testifies to the existence of a more or less continuous, and remarkably homogeneous, tradition of mechanical water-clocks, mainly from Islam but extending without change to contemporary Byzantium, and with some modifications even as far as China and perhaps India. This tradition seems to have been transmitted to Europe without much change or dilution during the medieval renaissance of the twelfth and thirteenth centuries; during the next century it became conflated with other lines of development and was thus transformed into the essentially modern principle of the mechanical clock.

Price, "Automata," 16.

> machine that moved independently on the strength of both a power supply and a plan of action, or program, of its own) was the higher, general category; the clock was merely a particular variety of automaton.[25]

Released from its grotto prison and made mobile, the clockwork automaton was increasingly popular among a notable minority of the population: "the ruling classes and the aristocracy, the members of the learned professions, and those who lived in towns…. This was an important fraction: it spoke for a minority that governed the country, controlled its economy, and, in the broadest sense, shaped its culture."[26] This notable minority, in a culture permeated by clockwork figures, came to see the clockwork automaton as a metaphor for the functioning of the universe and its creatures great and small.

A number of technological advances during the early modern period contributed to the rise of the clockwork automaton. The first of these was the mechanical escapement,[27] a device consisting of an arrangement of gears designed to transfer energy from a power source to a larger mechanism in a controlled manner. In a clock, the escapement drives a counting mechanism that tracks the passage of time; in an automaton, the mechanism is responsible for animating the figure. The controlled release of power by an escapement results in the "tick" of a clock, as the gear is alternately captured and released (allowed to *escape*, hence the name) in its forward movement. The precise origin of the mechanical escapement is now lost, although Otto Mayr traces its development to sometime in the late thirteenth century, based on texts describing the "current state of horology" in 1271:

> There was no longer any doubt that the clock of the future would be mechanical and weight driven, but one problem remained: lacking was a device (which we now refer to

---

[25] Mayr, *Authority*, 21. Minsoo Kang relates this shift in definition to the appearance of the Vaucanson automata in the 1730s (Kang, *Sublime Dreams*, 139-40).

[26] Mayr, *Authority*, 29. For a discussion of the rise of the use of metaphor as a conceptual aid in general, and the clock metaphor in particular, see ibid., chapter two.

[27] While this was not the first escapement designed (see Figure 3 for an illustration of Philo's hydraulic escapement), it was the first one that did not rely on water.

> as the escapement) to retard the accelerating motion of a falling weight into a steady motion of strictly constant velocity. This key element, however, and with it the successful mechanical clock itself, was invented within the following decade or so.[28]

Historians have traced the earliest known type of mechanical escapement—the verge and foliot—to tower clocks of the late thirteenth / early fourteenth century.[29] These earliest examples, powered by weight-and-cable systems, offered a great advantage over hydraulics in size, convenience, and location possibilities, as well as the fact that they didn't freeze or leak; however, they were typically no more accurate than their predecessors—often falling behind by as much as one or two hours per day.[30]

Subsequent inventions such as the spring drive, however, not only greatly improved accuracy, but made miniaturization possible. Early spring drives consisted of bent wood (usually from a sapling), which replaced the weight as a power source in clockwork mechanisms by providing a constant force on the cable without the added mass of a heavy weight; later versions replaced the bent wood with a metal coil. Both of these drives were used in large public clocks, but found their best use in the creation of portable clocks, not only small enough to fit on a tabletop, but to be carried around on one's person. Historian Lynn Townsend White traces the earliest extant examples of these types of personal clocks to the late fifteenth century: Phillip the Good's "magnificent chamber clock," crafted around 1430, which currently resides in the Germanisches Nationalmuseum in Nürnberg;[31] and three pendant watches ordered by Milanese Duke Ludovico Sforza for himself, his wife, and a friend in July 1488, "two of which shall strike the hours."[32] This led to a split in the historical development of clockwork: while the large

---

[28] Mayr, *Authority*, 5.

[29] Likely candidates for the earliest example include the Dunstable Priory in Bedfordshire, England, built in 1283 (Julian M. Luxford, *The Art and Architecture of English Benedictine Monasteries, 1300–1540: A Patronage History*, Studies in the History of Medieval Religion [Rochester, NY: Boydell Press, 2005], 209-10); and the clock at the Palace of the Visconti in Milan, Italy, from 1335 (Abbott Payson Usher, *A History of Mechanical Inventions*, Rev. ed. [New York: Dover, 1988], 196).

[30] Willis Isbister Milham, *Time & Timekeepers, Including the History, Construction, Care, and Accuracy of Clocks and Watches* (New York: The Macmillan Company, 1945), 83. Compare this mechanism to the al-Jazari Castle Clock, which was said to keep very accurate time. (Figure 7 on page 58).

[31] Lynn Townsend White, *Medieval Technology and Social Change* (Oxford: Clarendon Press, 1962), 127.

[32] Ibid.. White cites here a letter written by Sforza found in the Modena Archive dated 19 July 1488.

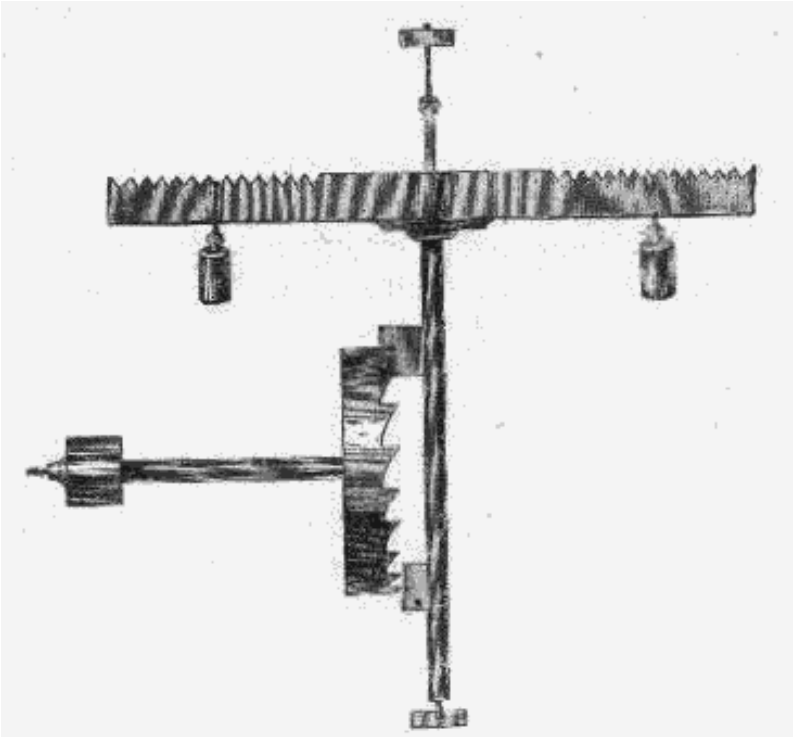


Figure 11. Verge and foliot mechanism from the De Vick clock, built 1379 in Paris by Henri De Vick. In this type of clockwork, the vertical toothed wheel (the verge escapement) drives the horizontal bar (the foliot), which then regulates the clock. The clock's rate may be adjusted by moving the counterweights on the foliot. Originally from Pierre Dubois, Histoire de l'horlogerie *(1849).*

public clocks typically retained the weight-and-cable power systems of previous times, the spring drive opened up a new market for personal, portable clocks. Freed from "the constraints of the weight-drive that had condemned it to be stationary, bulky, and short of running time,"[33] the spring-driven clock "opened up wider markets, and for the craft of clockmaking it initiated an era of professionalization."[34] By the sixteenth and seventeenth centuries, it became common among nobility and the wealthier strata of society to possess *figure clocks*, a form of desktop clock, typically set on a small, self-propelled wagon, which featured moving figures and was primarily designed as an entertainment: often with clock face and hands obscured or completely absent, these figure clocks were more automata than timepiece.[35]

[33] Mayr, *Authority*, 8.
[34] Ibid..
[35] Ibid., 21. Mayr attributes the rise of mechanism largely to the proliferation of these clocks during the sixteenth century (Ibid., 21-26).

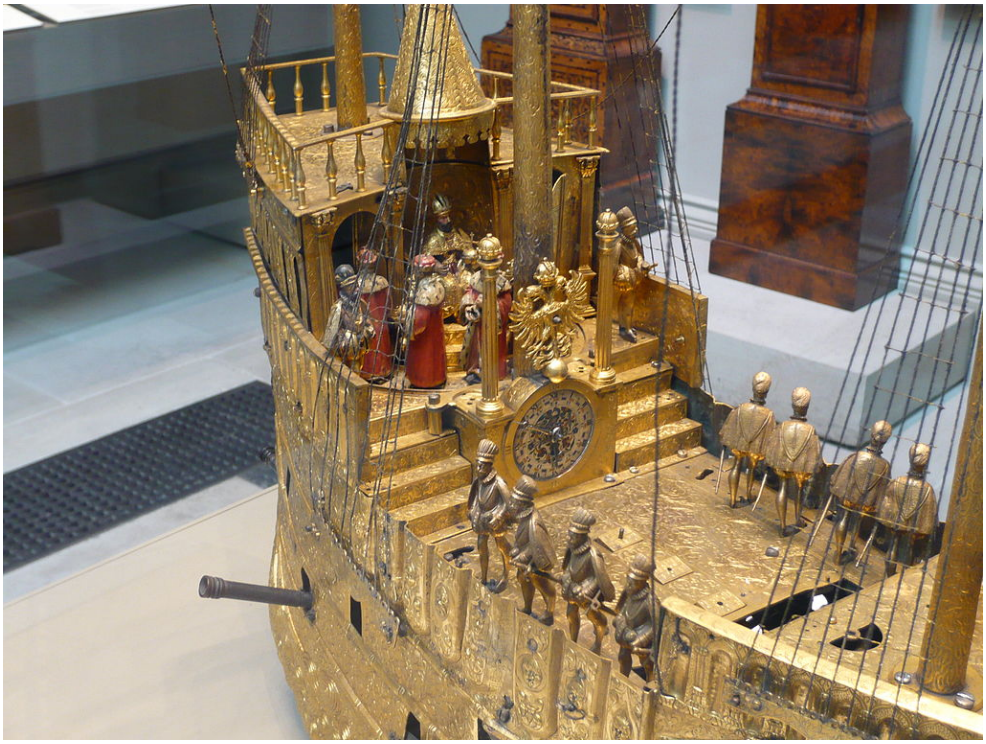


Figure 12. An example of the elaborate figure clocks of the sixteenth century, in this case a mechanical galleon (c. 1585) intended to announce banquets at court. Note the small clock face at the front of the quarterdeck. Photo © David Biggins, Cambridge, UK.

Even among those unable to afford personal clocks, the rise of the clockmaking industry had a profound effect on their daily lives. Beginning in the fourteenth century, large tower clocks were increasingly installed in "churches, town halls and city gates of the major towns,"[36] making it "possible to subordinate the public life of the whole town to a common schedule."[37] These community clocks became commonplace objects in many urban areas, altering the public's concept of time and facilitating the rise of mechanistic philosophy. According to Otto Mayr,

> clocks and automata thus became mediators between the worlds of magic and rationality; demonstrating how magical effects could be achieved by clever machinery of metal and wood, they changed the mental habits of their viewers. Gradually people learned how to account for unfamiliar phenomena not by resorting to the supernatural

[36] Ibid., 15.
[37] Ibid.. Mayr further claims that "In designs of utopian communities, it became customary to place a clock, in dominant position, in the center of the town" (Ibid..).

> but by identifying concrete, observable causes; they learned how to connect causes and effects into extensive, characteristically linear cause-and-effect chains. Clocks, in short, helped to teach Europeans how to think "mechanically." [38]

The most famous of these community clocks was that of the Strasbourg Cathedral, whose animated gilded rooster, designed to flap its wings and crow on the hours,[39] is now housed in the Strasbourg Museum for Decorative Arts. The only surviving part of the first astronomical clock in Strasbourg (1354), the rooster is widely regarded as the oldest preserved automaton in existence today.[40]

The astronomical clock in Strasbourg serves as a microcosm of the changes taking place in Europe during this period. Originally installed in the mid-fourteenth century, the clock went through two major renovations; each of the three versions of the clock not only represented the state of the art of clockmaking in its time but also reflected through increasing complexity and reliance on automata and other mechanisms the growing importance of the clockwork metaphor in European philosophy. The original clock, with its gilded rooster at the apex, was known as the "Three Kings" clock, as it housed an automated scene of Jesus' birth, with the three kings bowing before a figure of Mary and Child as the clock struck.[41] In addition, this early version featured a painted depiction of a "Zodiacal Man" illustrating the relationship of the signs of the zodiac to parts of the human body, reflective of medieval scholars' reliance on astrology (a branch of medieval "natural magic" that lost credibility in the early modern period).

---

[38] Ibid., 26. Mayr further claims that "the public welcomed the new timekeeper with almost unanimous enthusiasm. The prevailing sentiments were expressed by one author who praised a well-kept clock as the sign of a well-run community and by another who called clocks the very criterion of civilized existence" (Ibid., 15).

[39] Price, "Automata," 18.

[40] Musées de la ville de Strasbourg, "Museum of Decorative Arts >> Collections," accessed 15 October 2013. http://www.musees.strasbourg.eu/index.php?page=collections-Arts-deco-en. See also Mayr, *Authority*, 7; Price, "Automata," 18-19; Bedini, "Role of Automata," 29; Kang, *Sublime Dreams*, 80, 139.

[41] This and further information regarding historical and current appearance and operation of the Astronomical clock come from the cathedral's websites: Jana Dolecki, Strasbourg.info, "Strasbourg > Sights > Cathedral," last updated 14 November 2014, http://www.strasbourg.info/cathedral/; Cathédrale Notre Dame de Strasbourg, "L'horloge astronomique," accessed 15 October 2013, http://www.cathedrale-strasbourg.fr/horloge_astro.aspx.

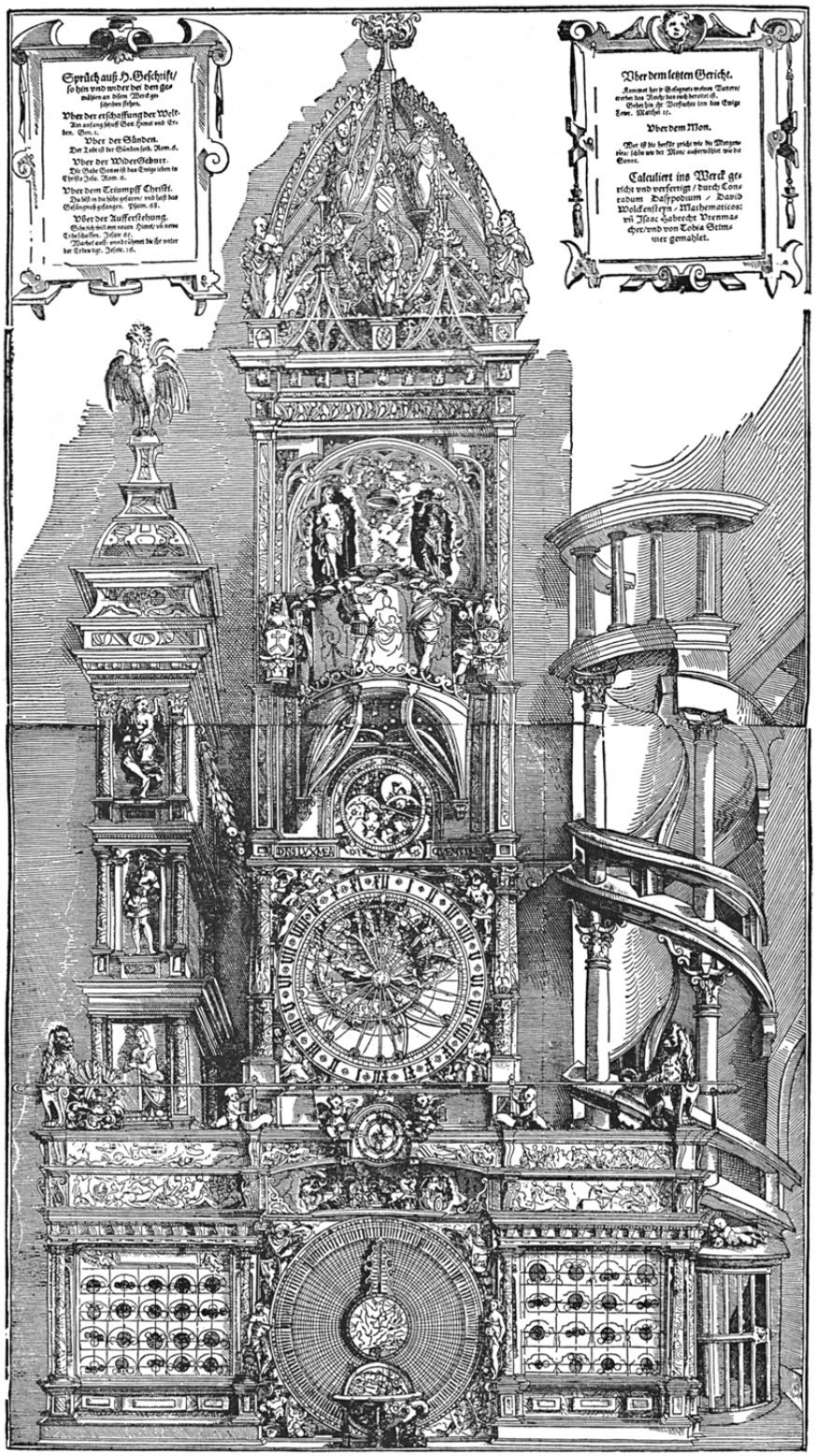


Figure 13. The second astronomical clock of Strasbourg Cathedral (constructed between 1571 and 1574), showing the elaborate performing automata and the rooster, top left, from the original clock. Woodcut by Tobias Stimmer, 1574.

After the original clock stopped working, it was replaced between 1571 and 1574 with an even more elaborate mechanism, designed by mathematician Conrad Dasypodius and realized by a team of engineers that included the brothers Halbrecht, Swiss clockmakers; astronomer and musician David Wolckenstein; and Swiss artists Tobias and Josias Stimmer. This second clock was a reflection of emerging scientific interests: according to Dasypodius,

> clocks comprised the entire quadrivium, namely, the major four—astronomy, geometry, arithmetic, music—of the seven liberal arts. Apart from timekeeping, their principal capabilities were astronomical prediction, musical performance, and the imitation of life. These capabilities were exercised on any scale and in almost any combination; the great monumental clocks of some cathedrals—Strasbourg was the best known—comprised all four. Astronomy, nevertheless, ranked highest.[42]

The Dasypodius clock included not only clock face, calendar dial, and astrolabe, but also a celestial globe, numerous automata and musical carillon. The clock was additionally adorned with elaborate paintings of the Fates, prominent contemporary figures (most notably Copernicus), and various scenes of religious significance. The gilded rooster of the earlier clock was given a place of prominence at the top of the cupola.

In 1788 this second clock also stopped working but was not replaced until 1838, when the engineer Jean-Baptiste Schwilgué (1776–1856, a native of Strasbourg) and his team of 30 artisans and engineers installed an entirely new, state-of-the-art mechanism in the Dasypodius housing. This third clock, still in operation today, contains the first example of a perpetual mechanical Gregorian *computus* (a device used to calculate the date of Easter according to the Gregorian calendar), designed by Schwilgué in 1816. Considered the "work of a lifetime" ("C'est l'œuvre d'une vie entière"),[43] the Schwilgué clock is known not only for its accuracy—it correctly predicted a lunar eclipse for March

---

[42] Mayr, *Authority*, 10.
[43] "L'horloge astronomique," accessed 15 October 2013.

2006—but also for its complex automata performances, including the ages of man, the Apostles parading before Christ, and a variety of angelic custodians marking the time. The current Strasbourg cathedral clock stands eighteen meters high and performs daily at 12:30 before a crowd of tourists.

From the beginning of the fifteenth until the mid-seventeenth century, clockmaking represented the finest of the mechanical trades, and the pinnacle of cutting-edge technology. As the popularity of clocks and their automata grew during the early modern period, interest in the trade grew, granting a new respectability to a menial and base profession—the mechanical arts had previously been considered unfit for scholars or philosophers. The rise of clockmaking as an industry during the early modern period witnessed a concomitant rise in the development of fine instrumentation; as clocks grew smaller and more complex, the techniques of precision metalwork allowed for (and required) ever-finer tools. According to Price, these new tools directly contributed to the scholarship of the time: "one can trace the steady evolvement of the clockmakers' fine metal-working craft to its finest manifestation in the craft of the instrument maker which was to dominate the development of learning during the Scientific Revolution."[44]

RENAISSANCE ROBOTICIST

Arguably the greatest figure of the early modern period is Leonardo da Vinci (1452–1519), Italian polymath, artist, and inventor who is hailed by many as the archetype of the Renaissance Man. Leonardo's life embodies the basic tenets of Renaissance humanism: namely that man, as the center of the universe, holds limitless potential and therefore should embrace all knowledge in order to develop this potential to its fullest. Displaying an early bent toward mathematics and engineering, Leonardo also developed an interest in art and, living in the small village of Vinci, had the opportunity to observe nature closely. Indeed, according to historian Charles O'Malley, "the whole foundation of his lifelong

[44] Price, "Automata," 19.

interest in the various phenomena of nature seems to have been laid at this time."[45] Leonardo was a talented enough artist to be apprenticed at fourteen to the workshop of Andrea del Verrocchio (c. 1435–1488), master of one of the finest workshops in Florence.[46] The inquisitive young artist was well placed in the workshop of Verrocchio, "whose interests were wide enough to comprise all forms of the arts, music, mathematics and engineering, subjects in which Leonardo would in time surpass his teacher."[47] While at Verrocchio's workshop, Leonardo worked not only in painting, sculpture and architecture, but also in civil engineering—according to Vasari, "the preparation of designs for the removal or the perforation of mountains, to the end that they might thus be easily passed from one plain to the other"[48]—as well as automation—"designs for mills, fulling machines, and other engines, which were to be acted on by means of water."[49] In 1478 Leonardo left Verrocchio's studio, eventually travelling to Milan, where he worked for nearly twenty years as an artist and engineer under the patronage of Duke Ludovico Sforza.[50] When Sforza was overthrown in 1499, Leonardo fled to Venice. Leonardo also enjoyed the patronage of Cesare Borgia from 1502 to 1506, for whom he also worked as a military engineer, and of Pope Leo X from 1513 to 1516, during which time Leonardo lived in the Vatican (along with Raphael and Michelangelo) as a sculptor and artist. He spent his remaining years in the employ of the French King Francis I, where a generous pension allowed him the time to indulge in a variety of artistic and scientific pursuits. When he died in 1519, Leonardo left behind a compendium of largely unorganized notebooks, created and maintained throughout his daily life, comprising some

---

[45] Charles D. O'Malley, *Leonardo on the Human Body: The Anatomical, Physiological, and Embryological Drawings of Leonardo Da Vinci* (Mineola, NY: Dover, 1983), 17.
[46] Marc Rosci, *The Hidden Leonardo*, trans. John Gilbert (Chicago: Rand McNally, 1976), 20.
[47] O'Malley, *Leonardo on the Human Body*, 17.
[48] Giorgio Vasari, *Lives of the Most Eminent Painters, Sculptors, and Architects*, trans. Mrs. Jonathan Foster, vol. 2 (London: Bell & Daldy, 1871), 369.
[49] Ibid., 368.
[50] In his letter to the Duke requesting patronage, Leonardo outlines his many talents as a military engineer that he can offer to Milan in her defense, as well as his abilities in architecture and civil engineering for use in peacetime; almost as an afterthought, he mentions that he can also paint. Leonardo-History.com, "Leonardo da Vinci's Letter to Ludovico the Moor," accessed 10 December 2013, http://www.leonardo-history.com/life.htm?Section=S5.

13,000 pages of notes, diagrams, and illustrations representing a lifetime of scientific, artistic, and philosophical work.

During his lifetime Leonardo was most highly valued as an engineer; his notebooks are filled with mechanical designs large and small, from the practical to the fantastic. While many consider his drawings largely theoretical, roboticist Mark E Rosheim believes Leonardo may have actually built several of his automata designs.[51] In fact, Rosheim claims that in 1495 Leonardo "designed and possibly built the first articulated humanoid robot in the history of western civilization"[52]: an animated suit of armor that "was designed to sit up, wave its arms, and move its head by means of a flexible neck, while opening and closing its anatomically correct jaw."[53] Rosheim additionally credits Leonardo with the construction of a lion automaton for the occasion of Francis I's entry into Lyons in 1515[54] and a bell-ringing Jacquemart clepsydra in 1510 that "antedates automata that would appear in grottoes in the late sixteenth and early seventeenth centuries, and may have been their inspiration."[55] Rosheim differentiates Leonardo's water clock from its predecessors in its foreshadowing of modern digital computers:

> Even in Leonardo's day, a water clock was an anachronism. But no doubt pressed by a demanding patron, Leonardo rose to the challenge and produced a fascinating, modular, compact, and reliable clock that anticipates modern, computer-like functionality. The key difference between Leonardo's clock and those of his predecessors is the modern conception of modularity and compactness…. Leonardo's accomplishment marks the

---

[51] Mark E. Rosheim, *Robot Evolution: The Development of Anthrobotics* (New York: John Wiley and Sons, 1994), 12-13.; Rosheim offers possible explanations for why there is no evidence, physical or textual, of this: "Leonardo, on the other hand, may have decided to keep such a book obscure for the same reasons he wrote in mirror handwriting: to elude competitors who might steal his intellectual property. Likewise, any hardware might have been disassembled. Such creations could have also been perceived by religious zealots of the time as blasphemous mockeries of God's handiwork." (Ibid..)
[52] Rosheim, *Leonardo's Lost Robots* (Berlin: Springer, 2006), ix.
[53] Ibid..
[54] Ibid., 21. In Rosheim's reconstruction of this automaton he notes the similarity of the carriage mechanism to that of Hero's automatic theatre, but with the addition of the recent improvements of the spring drive and escapement mechanisms.
[55] Ibid., 120. Rosheim's claim here is based on comments made in 1600 by Michelangelo, who may well have been unfamiliar with the automata in the Garden of Hesdin discussed in chapter one.

> first occurrence of a sequence controller in the modern sense of a simple, dedicated computer.[56]

According to Rosheim's research, the Jacquemart did not merely strike the bell on the hour; rather, he struck the corresponding number of hours (a "program" representing a revolution in hydraulics); in addition, the Jacquemart as designed by Leonardo was portable (if rather unwieldy)—based on drawings and notes, it was self-contained and would have fit on a horse cart. Several of Leonardo's other designs may have been built only after his death, such as his "Cittern Player," designed in 1497 but not built until the mid-1600s.[57] This automaton was at one time attributed to Juanelo Turriano, and may have served as inspiration for the clockwork Monk described in the previous chapter.

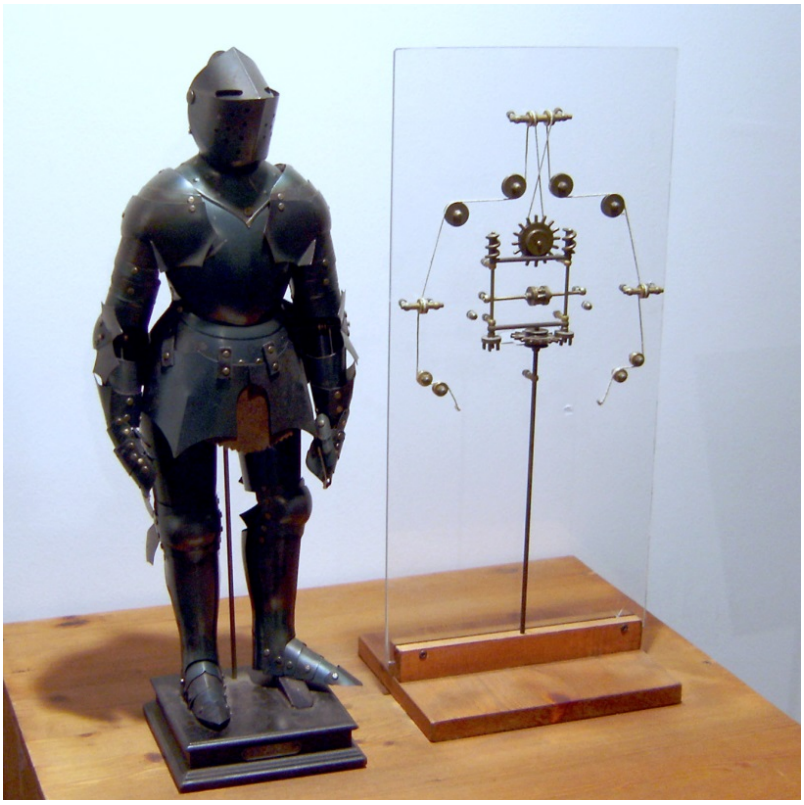


Figure 14. Reconstruction of Leonardo's animated suit of armor, showing the mechanism as it would have operated. From the 2005 exhibit "Leonardo da Vinci: Mensch—Erfinder—Genie" in Berlin. Photo by Erik Möller.

[56] Ibid., 160. A *sequence controller* determines the order in which operations are carried out.
[57] Ibid., 21.

While historian Derek J. de Solla Price lauds Leonardo da Vinci as "a link between the Middle Ages and the Renaissance… in every aspect of the history of technology,"[58] the truth of the matter is that Leonardo's work had greater impact on his successors in the realm of anatomy than mechanics. Leonardo da Vinci was a true polymath, versed not only in mechanics, architecture, and sculpture but medicine and art as well. His illustrations represent the application of this broad knowledge to his understanding of the human body: his explorations into the details of the human form reveal a highly mechanistic regard for the engineering involved in skeletomuscular systems, human and otherwise. Indeed, Glaser and Rossbach credit Leonardo with the earliest mechanistic impulse: "The early drawings show conclusively that the mechanical hypothesis dominated Leonardo's anatomical investigations from the start. In this light, it was not Descartes but Leonardo who first formulated the equation 'man=machine'."[59] Leonardo's notes and illustrations depict not only the systems he revealed in the course of his own dissections,[60] but describe how these systems work in conjunction with one another: "His commentaries demonstrate the movement of the hand in terms of the mechanical interplay of the groups of muscles and sinews that operate on the various bone structures."[61] He was often sought out for his knowledge and expertise in the illustration of the human body; however, he was also a notorious procrastinator, and left many works incomplete. For example, Leonardo had planned a *Treatise on Anatomy* as early as 1489;[62] although he worked on it periodically over the next several decades,

---

[58] Price, "Automata," 21.
[59] Horst Albert Glaser and Sabine Rossbach, *The Artificial Human: A Tragical History*, trans. Joseph Swan (Frankfurt: Peter Lang, Inc, 2011), 76.
[60] Ibid., 74. Glaser and Rossbach identify at least seven separate dissections performed by Leonardo over a twenty-year period based on evidence taken from his notebooks. Dissection was relatively rare during Leonardo's time; see my later section on anatomy ("A Living Machine") for a discussion of the rise of the practice during the early modern period.
[61] Ibid., 75.
[62] Vasari speaks of this work as a collaboration with the anatomist Marcantonio della Torre (1481–1511), anatomy professor at the University of Padua (Vasari, *Lives*, 2, 380-81); however, this collaboration did not begin until 1510, after Leonardo's project had already begun.

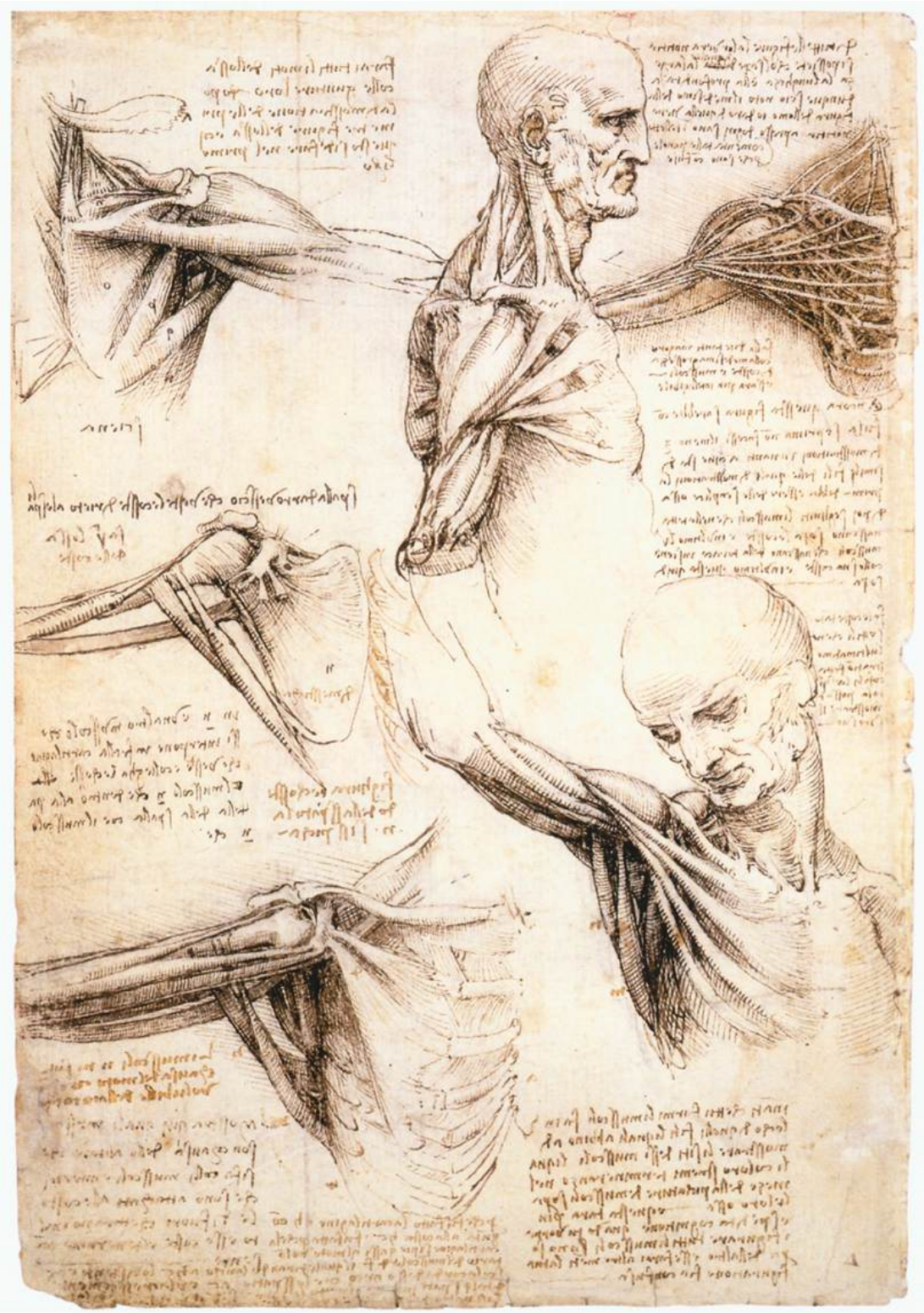


Figure 15. Anatomical studies of the shoulder, by Leonardo da Vinci (1510–1511), showing his incredible attention to mechanistic detail. From the collection of the Royal Library, Windsor, UK.

amassing numerous notebooks of illustrations and notations, as time passed his plan for the work grew to encyclopedic proportions, and it was never completed during his lifetime.[63]

Francesco Melzi (1491–1570), Leonardo's favorite pupil, inherited these notebooks upon Leonardo's death in 1519, and undertook to complete his master's *Treatise on Anatomy*. Unfortunately, this proved to be a formidable task. Leonardo's notes, when present, were written in a mirrored script; even when they could be deciphered, their content was largely incomprehensible, if not outright heretical, to his successors:

> During the sixteenth century, many of Leonardo's repetitive tentative gropings towards mechanical solutions of cosmic and human problems must have appeared not only incomprehensible, but if penetrated, outrageous. It cannot have been far from the thoughts of good citizens like Melzi and his friends that such speculations were of so blasphemous a nature as to raise the literal truth of the cry, 'publish and be damned'. Vasari's reference to Leonardo as heretical, and more a philosopher than a Christian shows how real this danger was in I550.[64]

The task of completing even a partial, less-ambitious version of Leonardo's original treatise proved intractable to Melzi and his assistants. According to medical historian Kenneth Keele, Melzi spent the rest of his life trying to "compile a worthy *Treatise on Painting* out of the mass of scattered notes"[65] but was largely hindered by the incomprehensible nature of the manuscripts. Artist-writer Giorgio Vasari (1511–1574), whose "short *Life of Leonardo* did more for the preservation of his name as an anatomist than any,"[66]

> could not appreciate the scientific value of the sheets of anatomical drawings which he saw; he had neither the time nor the knowledge to do so. But he was intelligent enough

[63] O'Malley, *Leonardo on the Human Body*, 31-33. Editors O'Malley and J.B. Saunders claim their edition to be the first attempt to "arrange Leonardo's note-books so as to indicate systematically what the extent of his anatomical studies was" (Ibid., 35).
[64] Kenneth D. Keele, "Leonardo Da Vinci's Influence on Renaissance Anatomy," *Medical History* 8 (1964): 362-63.
[65] Ibid., 362.
[66] Ibid., 365.

> to sense their importance: 'Whoever succeeds in reading these notes of Leonardo,' he wrote, 'will be amazed to find how well that divine spirit has reasoned of the arts, the muscles, the nerves and veins, with the greatest diligence in all things.'[67]

Melzi himself died in 1568 after devoting nearly forty years to the compilation of his master's work. Despite the employ of two assistants, Melzi was unable to complete the task, organizing and annotating only a fraction of the opus. In 1651 this partial compilation was published by Raffaelo du Fresne as *A Treatise on Painting by Leonardo da Vinci* (*Trattato della pittura*), consisting of numerous anatomical illustrations and focusing on the link between the artistic and the scientific. Leonardo's manuscripts were slowly dispersed in the generations to follow, and many of the pages remain irretrievably lost.[68]

While the scientific principles underlying Leonardo's work may have been ahead of their time, the quality and detail of his anatomical illustrations were clearly recognized by his contemporaries. Leonardo's illustrations offered "artistic exploration of the muscles, bones and joints, and their movements" which "successfully spread the roots of his anatomy into the studios of his fellows… due to the science it brought into their art."[69] According to Keele, by bringing a scientific rigor to art, Leonardo

> had broken the hard ground of bigotry and prejudice which had buried anatomy for so many centuries;… had stimulated the fusion of art and science in anatomical representation; and… had prepared the tilth to receive the masterpiece of Vesalius and Calcar. In I543, when this was published, it was neither lost nor damned.[70]

Although Keele's depiction of Leonardo da Vinci as the "spearhead of the new creative anatomy"[71] waxes hyperbolic at points, the impact of Leonardo's work is undeniable. According to physician

---

[67] Ibid..
[68] O'Malley, *Leonardo on the Human Body*, 33-35. See also Rosheim, *Leonardo's Lost Robots*, "Beginnings" 1-20, for a description of Rosheim's quest to recover Leonardo's manuscripts and illustrations in an attempt to recreate several of the automata depicted.
[69] Keele, "Leonardo Da Vinci's Influence," 369.
[70] Ibid., 370. A discussion of Vesalius' groundbreaking anatomical work follows in the next section.
[71] Ibid., 369.

Girolamo Cardano (1501–1576), Leonardo's explorations into the human body presaged that of later anatomists, such as Cardano's contemporary Vesalius (1514–1564), by nearly a century:

> In [Cardano's] *De subtilitate*, after remarking that a painter is at once philosopher, architect, and dissector, he continues, "for proof there is that remarkable imitation of the whole human body which [I saw] many years ago, by Leonardo of Vinci and of Florence, which was almost complete; but the task required a great master and investigator of nature such as Vesalius." Apparently Cardano the physician immediately recognized the non-systematic nature of [Leonardo's] work.[72]

Vasari describes the detail evident in Leonardo's anatomical illustrations, over 750 of which were drawn as part of his collaboration with the anatomist Marcantonio della Torre from 1510 to 1511: according to Vasari, Leonardo "filled a book with drawings in red crayons, outlined with the pen, all made with the utmost care from bodies dissected by his own hand."[73] Unfortunately, della Torre's budding career as a professor of anatomy at the University of Padua[74] was cut short by his death from the plague in 1511. Leonardo was again distracted from the ambitious project, ultimately leaving the revolution in medical science to later anatomists.

A LIVING MACHINE

The Scientific Revolution that swept Europe during the early modern period brought with it a more particular interest in the form and function of the human body. Until the early seventeenth century, medicine (including anatomy) as taught in European universities was founded in the writings of the Roman physician Galen (Aelius Galenus, 129–c. 200 CE), whose authority was previously accepted without question. Galen's texts, like so many texts from antiquity, were first transmitted to Europe

---

[72] O'Malley, *Leonardo on the Human Body*, 33; first and third insertions mine, second insertion original to O'Malley's text. See also Vasari, *Lives*, 2, 367, which speaks to the "instability of his character [which] caused him to undertake many things which having commenced he afterwards abandoned."

[73] Vasari, *Lives*, 2, 380-381.

[74] The University of Padua was a major center of Renaissance humanist thought and the Scientific Revolution, particularly in the advances of its faculty in the areas of medicine and anatomy.

through scholars of the Islamic Golden Age; in this case, primarily Avicenna (Abū ʿAlī al-Ḥusayn ibn ʿAbd Allāh ibn Sīnā, 980–1037), whose five-volume *Canon of Medicine* (1025) represented the most complete and authoritative summary of medical knowledge at the time. Translated into Latin in the mid-twelfth century, the *Canon* remained an influential medical textbook until the early seventeenth century.

Due to the Roman prohibition against human dissection,[75] Galen's original texts and illustrations had been based instead on dissections of Barbary apes, under the assumption that human anatomy would be similar. It was only as inaccuracies in Galen's work were revealed through close empirical exploration by scholars such as Belgian anatomist Andreas Vesalius (Andries van Wezel, 1514–1564) that the *Canon* was preempted by more modern texts. Vesalius, often referred to as the founder of modern human anatomy, published his *De humani corporis fabrica libri septem* (*On the Fabric of the Human Body in Seven Books*, 1543) based on lectures given during his tenure at the University of Padua, in which he illustrated his discussion points to his students through human dissection. Vesalius' work included numerous illustrations (see Figure 16 on next page for an example) of his dissections by the "little-known Flemish artist, Jan Stefan von Kalkar … [whose] woodcuts depict the opened and partitioned bodies of his subjects in shockingly realistic detail."[76] This attention to minute detail reflects the increasingly mechanistic view of the human body, and is itself reflected in the art of the time: in Rembrandt's 1632 painting *The Anatomy Lesson of Dr Nicolaes Tulp*, for example,

> the muscles and sinews of the forearm have been carefully laid bare, and the colleagues eagerly study their movement and function, guided by the explanations of the master who lifts (and thus activates) individual muscles or sinews with a medical instrument in

[75]For a discussion of this point and an overview of Galen's work, see Plinio Prioreschi, *A History of Medicine, Volume III: Roman Medicine*, 2nd ed. (Omaha: Horatius Press, 1996), 315-500.
[76] Glaser and Rossbach, *The Artificial Human*, 82.

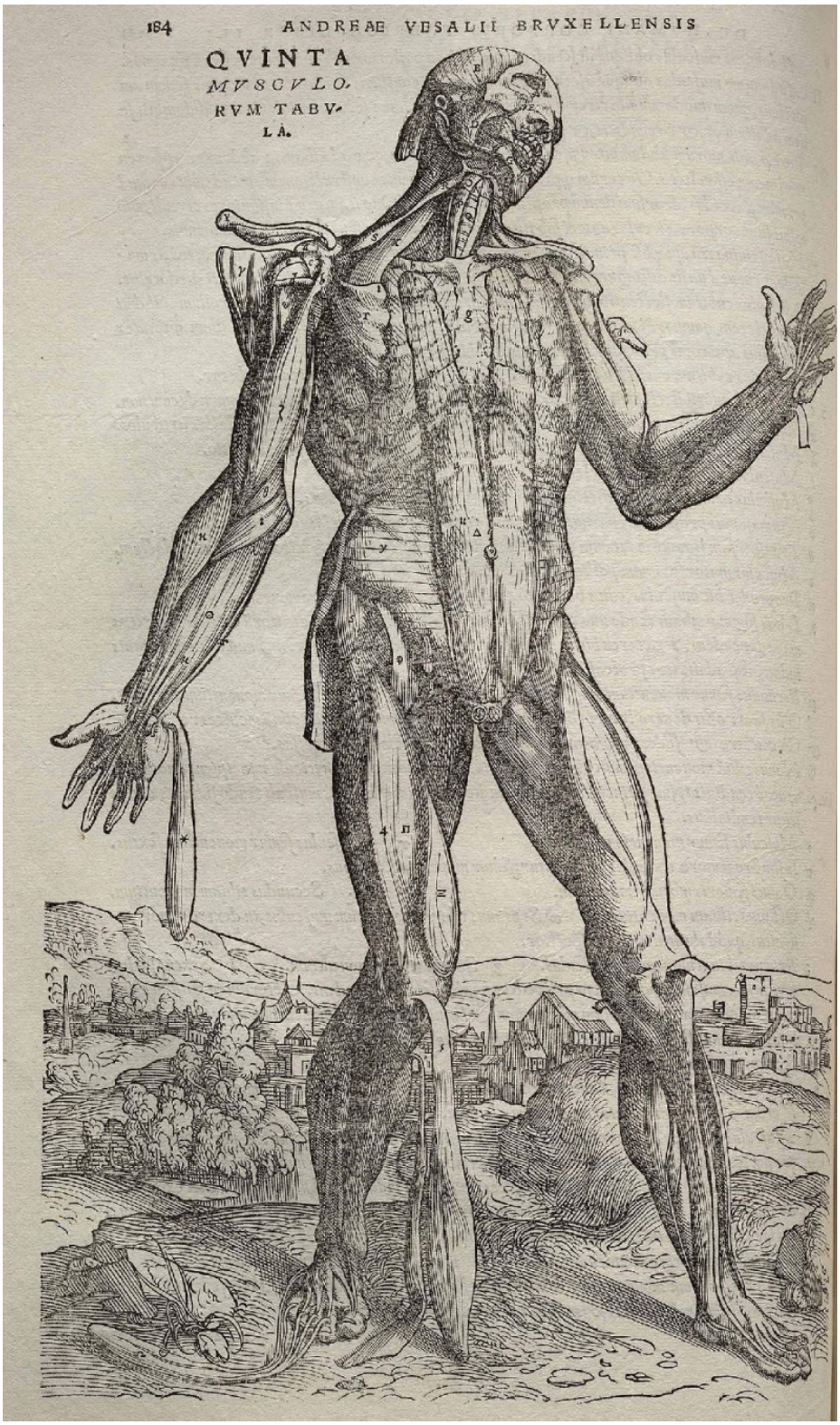


Figure 16. Engraving from Andreas Vesalius, De humani corporis fabrica *(1543), page 184. Woodcut by Jan Stefan von Kalkar.*

> order to demonstrate the structures and systems governing the movements of fingers and hand.[77]

According to Kenneth Keele, the detail of these later works owes much to Leonardo da Vinci's own explorations and illustrations of anatomical structures nearly a century and a half prior.[78] Leonardo's great anatomical treatise, intractable during his own time, was finally accessible in the mid-sixteenth century. By Vesalius' time, the taboo against human dissection, whether a result of a distaste for the notion of desecration[79] or the outright prohibition of the practice, was gone,[80] and scholars were able to turn their instruments to discovering the inner workings of the human body.

According to medical historian Pilino Prioreschi, "the first unambiguous account of a pedagogical dissection appears in the anatomy textbook of Alderotti's student Mondino de Luzzi, who mentions one performed in January 1314,"[81] although the first dissection for forensic purposes had likely occurred a few decades prior. Better known as Mundinus, de Luzzi (ca. 1270–1326) is credited with the first example of a modern dissection manual, the *Anathomia corporis humani* (*Anatomy of the Human Body*, 1316), written partly as a result of his explorations.[82] Unfortunately, Mundinus had difficulty gaining

---

[77] Ibid., 86.

[78] See Keele, "Leonardo da Vinci's Influence."

[79] While ancient anatomist Herophilus of Alexandria (335–280 BCE) first practiced human dissection in his work, he did so despite "concerns about the polluting power of the human body and the taboos about its 'desecration' that prevailed in the Greek world" (Pilino Prioreschi, "Determinants of the Revival of Dissection of the Human Body in the Middle Ages," *Medical Hypotheses* 56, no. 2 [2001]: 229); there is no further evidence of the practice of human dissection until the thirteenth century (Ibid., 230-231); see also Glaser and Rossbach, *The Artificial Human*, 74-86.

[80] Pilino Prioreschi accounts for the medieval adaptation of human dissection as a viable method of scientific exploration through three distinct vectors of change:

> a) the belief in the dichotomy between the body and the soul with the consequent sharp decline in status of the body; b) the slow shifting of medicine toward a scientific mode; c) the correspondence between life deeds and the importance of the cadaver. Of these three elements, the first eliminated the taboos surrounding the dead body and made possible dissection in general; the second stimulated the rebirth of anatomical dissection; and the third made cadavers available for anatomical dissection. Without the first element no dissection would have been possible, without the second there would have been no stimulus for systematic anatomical studies of the human body, without the third the material for such studies would not have been available in the social climate of the Middle Ages. (Prioreschi, "Determinants," 229).

In the history of the artificial human and its impact during the early modern period, Prioreschi's compelling analysis of each of these factors is secondary to their impact on the advancement of medical knowledge of the time.

[81] Ibid., 231. De Luzzi's mentor, Taddeo Alderotti (c.1210–1295), was a Florentine alchemist and professor of medicine at the University of Bologna.

[82] Luke Wilson, "William Harvey's 'Prelectiones:' The Performance of the Body in the Renaissance Theatre of Anatomy," *Representations*, no. 17 (1987): 64. Wilson further describes de Luzzi's work as "the most commonly used anatomy text on the

access to suitable cadavers, and his text relies heavily on Arabic translations of Hippocrates, Aristotle, and Galen, repeating their inaccuracies despite personal experience. To address some of these inaccuracies, University of Padua anatomist Marcantonio della Torre also attempted to create his own anatomy text in collaboration with Leonardo da Vinci, for which della Torre and da Vinci are each credited with performing a number of dissections. Were it not for his death of the plague, della Torre might have been successful in his efforts.

While dissections in Mundinus' time were purely pedagogical affairs, typically performed by "base" surgeons (demonstrators) under the instruction of a respected physician for the benefit of his students,[83] by the early sixteenth century attitudes toward the practice had evolved to the point that anatomy theatres were packed with spectators, not only from the university, but from the general public as well[84]—Vesalius being one of the pioneers of this change. With Vesalius, public dissections evolved into entertainments of their own, accompanied by various festivities and sanctioned by the Church itself.[85] *Why* was the general public so interested in attending dissections? While it was partly for the spectacle (Vesalius' dissections are described as performative events[86]), this interest also reflects the increasingly mechanistic worldview of the period. Inspired by the clockwork automata they saw in the church towers, the town clocks, and the windows of the clockmakers' shops, a curious public wanted to

---

continent for the next 250 years" (Ibid..), although it did not dislodge Avicenna's *Canon* as the primary medical text. In his discussion of Renaissance anatomy theaters as performative spaces, Wilson describes this new approach to dissection as seen in a woodcutting of Vesalius at work:

> Vesalius began a trend in which the anatomist seems to have become a central performer on the stage of the [anatomy] theater…. In the dissection Vesalius did away with the roles of demonstrator [the surgeon employed to actually dissect] and ostensor [employed to point out the structures to which the physician referred] and performed their functions himself. It is also believed that for the most part he did not use a written text during the dissection, and that his lectures were sometimes spontaneous and sometimes by rote. The anatomist's relation to the body here is no longer ostensive but rather presentative or performative: the one who shows is himself caught up in the act of showing; what he shows is both the body and himself doing the showing…. Most obviously new here is Vesalius' relation to the body—he is touching it, using his hands to offer it to the attention of the spectator, rather than simply pointing at it. (Ibid..)

Like his predecessors della Torre and Vesalius, William Harvey (1578–1657) was also associated with the School of Anatomy at the University of Padua, having received his MD from the university in 1602. Harvey is considered, along with Vesalius, as one of the founders of modern anatomical practices.

[83] Ibid., 68.

[84] Ibid., 69. See also Glaser and Rossbach, *The Artificial Human*, 76-77.

[85] Wilson, "William Harvey's *Prelectiones*," 68-69.

[86] See Ibid..

see for themselves whether the human body actually worked in such a mechanical way. Vesalius' direct, hands-on approach to the study (and teaching) of anatomy became the standard in the field; in a rejection of prior tradition, he exhorted his students "not simply to accept at face value all they heard in the lectures: they should themselves perform dissections and learn with their own hands and eyes how the human body was constructed."[87] This advocacy of empiricism was one of the hallmarks of the Scientific Revolution, evident not only among scholars but among general public as well.

Through their minute investigations and detailed illustrations, physician-scholars of the Scientific Revolution systematically stripped the human body of its mystery as they literally stripped away the layers of its systems. With increasingly larger fields of inquiry laid bare to the scholar's instruments, and exposed as explainable through purely physical means, universal mechanism seemed in many ways inevitable. Not only were scholars learning about the innermost workings of the universe, but they were also reproducing them in ever more clever and realistic ways. The rapid expansion in all the sciences represented a "celebration of man's mastery over nature"[88] as speculation gave way to empirical exploration; it seemed that, given enough time, nature's every secret would be exposed, catalogued, and explained through scientific principles. Indeed, this appeared to be the goal of medical scholars of the time: according to Kang,

> what these figures had in common was their commitment to achieving a comprehensive understanding of the natural body through a mechanistic methodology that precluded the role of any forces or entities whose existence could not be quantified or proved

[87] Glaser and Rossbach, *The Artificial Human*, 78. Vesalius' methods were met with a great deal of resistance from his contemporaries, most notably colleague Mattheus Curtius, who was in charge of the discipline at Padua. According to Glaser and Rossbach, Curtius used the Mundinus text in his dissections, but noted its numerous errors—namely, those points where it diverged from Galen. This privileging of the earlier, theoretical text over the later, practical one was a source of contention between Curtius and Vesalius, who preferred empirical observation (Ibid., 77).
[88] Mayr, *Authority*, 26.

Figure 17. Title Page of Giovanni Alfonso Borelli's De motu animalium *(1685). © Wellcome Trust, UK. Copyrighted work available under Creative Commons Attribution only licence CC BY 4.0.*

> through observation or experimentation. This agenda is revealed in the frequent appearance of the machine-body analogy and the automaton in their works.[89]

A central text in this mechanistic approach to medicine is *De Motu Animalium* (*On the Movement of Animals*), published in 1680 in two volumes, representing the work of Italian physiologist Giovanni

[89] Kang, *Sublime Dreams*, 125.

Alfonso Borelli (1608–1679). Borelli is considered by many to be the father of *biomechanics*—the application of mechanistic methodology to the study of the structure and function of biological systems. Biomechanics is seen in modern medical science in such subfields as ergonomics, kinesiology, orthotics and prostheses, and sports biomechanics and rehabilitation. Borelli expands upon the automaton analogy first expressed by Descartes in illustrating his ideas on the function of biological systems. "An automaton seems to present some resemblance with animals since both are mobile bodies that comply with the laws of mechanics,"[90] declares Borelli, comparing the heart to a clock and suggesting the contractile movement of muscles. After Borelli, numerous physicians carried the movement of mechanistic medicine forward across Europe well through the eighteenth century.[91]

In nearly every case, mainstream mechanistic medicine depicted the analogy between biological functions and automata in a positive way, celebrating it "in a veritable refrain of praising God for having created such a beautiful, intricate, and well-functioning machine as the natural body."[92] During this period, the difference between animal and automaton was one of degree rather than of kind: God, the Great Engineer, had designed a natural body-machine that "infinitely surpasses all artificial automata."[93] The *man=machine* analogy became so thoroughly a part of the worldview that

> even in colonial America, the famed Puritan minister Cotton Mather in his 1721 work of natural theology *The Christian Philosopher* marvels at the human body: "a *Machine* of a most astonishing Workmanship and Contrivance! *My God, I will praise Thee, for I am strangely and wonderfully made!*" … [T]he wonder and awe that Renaissance philosophers felt in response to the automaton was still in full effect.[94]

---

[90] Giovanni Alfonso Borelli, *On the Movement of Animals*, trans. Paul Maquet (Berlin: Springer-Verlag, 1989), 319.
[91] For a more thorough tracing of this evolution, see Kang, *Sublime Dreams*, 126-132; Joseph R. Roach, *The Player's Passion: Studies in the Science of Acting* (Ann Arbor: University of Michigan Press, 1993), 60-66.; and Mayr, *Authority*, chapter three.
[92] Kang, *Sublime Dreams*, 131.
[93] Gottfried Wilhelm Leibniz, *The Monadology and Other Philosophical Writings*, trans. Robert Latta (London: Oxford University Press, 1951), 254.
[94] Kang, *Sublime Dreams*, 131-132, emphasis in original.

The "aura of the marvelous around the self-moving machine"[95] carried forward from the Renaissance into the seventeenth and early eighteenth centuries as mechanistic philosophy came to dominate the Age of Reason.

ENLIGHTENMENT AUTOMATA: VAUCANSON AND JAQUET-DROZ

In his historical tracing of the history of the automaton as a conceptual tool in Western culture, Minsoo Kang identifies two separate periods of its dominance during the seventeenth and eighteenth centuries. The first, from 1637 to 1748, Kang calls "the intellectual golden age of automata,"[96] characterizing it as the height of mechanistic philosophy in the early modern period. Kang demarcates this golden age with the publication of two key treatises: Descartes' *Discourse on the Method* in 1637, which first articulated mechanistic philosophy, and Julien de La Mettrie's *L'homme machine* (*Machine Man*) in 1748, regarded by many as "turning [man] into an organic automaton and *nothing more*."[97] The second period Kang refers to as the *automaton craze* of the following century, during which "the public continued to be awed and amused by objects that moved, spoke, and played music even after the worldview they represented was challenged by vitalist thinkers and writers."[98] In the former period, automata served as a proof-of-concept for mechanistic principles; in the latter, as quaint entertainments of clever design that, by their very mechanical nature, were objects of amusement rather than serious consideration.

Of the hundreds of automata created during the early modern period, two collections stand out. The first is the trio of lifelike automata created by Jacques de Vaucanson (1709 –1782) —*The Flute Player* (1737), *The Tambourine Player* (1738), and *The Digesting Duck* (1738)—which "set off [the]

---

[95] Ibid., 132.
[96] Ibid., 112. For a more thorough discussion of mechanism and the extent to which machines were used conceptually during the early modern period, see Ibid., chapter three (regarding the automaton) and Mayr, *Authority*, chapter two (regarding the mechanical clock).
[97] Kang, *Sublime Dreams*, 130. We will return to La Mettrie's unique brand of mechanism in the next section.
[98] Ibid., 175.

veritable automaton craze that lasted through the rest of the century."[99] To this day Vaucanson is the most well-known automaton creator of the age; at the time his name was synonymous with the idea of the automaton itself. It is difficult to overestimate the impact of Vaucanson's work not only on the automaton industry but on the very concept itself; according to Kang, "the popular as well as intellectual success of his works represents the ultimate culmination and apotheosis of the intellectual golden age of automata"[100]—indeed, La Mettrie himself used Vaucanson's automata as his conceptual model, deeming him "a new Prometheus."[101] The second collection important to the period are the Jaquet-Droz automata—*The Writer*, *The Musician*, and *The Draftsman*—built by Pierre Jaquet-Droz (1721–1790) between 1768 and 1774 in collaboration with his son Henri-Louis and assistant Jean-Frédéric Leschot. Described by Kang as "the pinnacle of achievement in the construction of clockwork automata,"[102] and representative of Kang's *automaton craze*, the Jaquet-Droz trio resides to this day in *La Musée d'art et d'histoire* in Neuchâtel, Switzerland.[103]

The Vaucanson and Jaquet-Droz automata, in addition to their popularity and stature as cultural objects, represent a revolution in design and production as well: these automata, built around the time of La Mettrie's controversial treatise, were quite different from those experienced by Descartes in the grottoes of Saint Germain-en-Laye in the early 1600s. Historian Jessica Riskin explains this difference as one of *simulation* versus *representation*:

> Eighteenth-century designers tried to simulate life's textures and substances, and even physiology, making automata from this period look very different from those of the previous or subsequent periods. Mechanical animals of the previous period, the

---

[99] Ibid., 104.
[100] Ibid., 111-112. For a more complete account of Vaucanson's impact, see also Ibid., 107-111, 139, 174-175; Reilly, *Automata and Mimesis*, 81ff; Roach, *The Player's Passion*, 66; as well as numerous other sources cited elsewhere in this chapter.
[101] Julien Offray de La Mettrie, *Machine Man and Other Writings*, trans. Ann Thompson (Cambridge: Cambridge University Press, 1996), 34.
[102] Kang, *Sublime Dreams*, 105-106.
[103] Ville de Neuchâtel musée d'art et d'histoire, "Les Automates Jaquet-Droz," accessed 12 January 2014. http://www.mahn.ch/collections-arts-appliques-automates.

> seventeenth and early eighteenth centuries, present artistic renditions of animal movements but make no attempt to imitate physiological processes.[104]

Automata prior to the eighteenth century were *mimetic* in nature: performative representations of the animals and figures they were built to resemble, going through the motions of life in order to amaze or amuse. From Hero's "moveable theatre" of antiquity, with its small but static figures moving about on their stage, to the ingenious automaton created for Louis XIV in 1649 by an artisan known only as Camus, comprising "a miniature coach and horses complete with footmen, page and a lady within the coach; all [of which] exhibited a perfect movement,"[105] the performances of these automata became ever more intricate. These performances also certainly inspired the rise of mechanistic philosophy and contributed to its popularity. However, it was not until the 1730s that "mechanicians tried earnestly to collapse the gap between animate and artificial machinery"[106] in the form of automata that, rather than merely *imitating* life, sought to *simulate* it. According to Riskin, both the Vaucanson and the Jaquet-Droz automata manage to succeed in this attempt: both sets of musical automata *actually played* their instruments rather than simply imitating human movements while a hidden music-box played the tune. These figures offered concrete proof for the validity of mechanistic philosophy. Ironically, they also helped contribute to its decline, as we shall see in the next section. But first, let us look at what made the Vaucanson and Jaquet-Droz automata so remarkable.

When Jacques Vaucanson built his *Flute Player* in 1737, clockwork automata were already quite popular across Europe, although they were by and large considered toys. Vaucanson's automaton, on the other hand, was something quite different: a life-sized "faun" able to play eleven different tunes on an actual flute, using his own breath (provided by an elaborate bellows system) and moving his fingers, lips, and tongue (all covered in organic skin). According to Kara Reilly,

---

[104] Jessica Riskin, "Eighteenth-Century Wetware," *Representations* 83, no. 1 (2003): 100.
[105] John Cohen, *Human Robots in Myth and Science* (London: George Allen & Unwin, Ltd, 1966), 86.
[106] Riskin, "Eighteenth Century Wetware," 101.

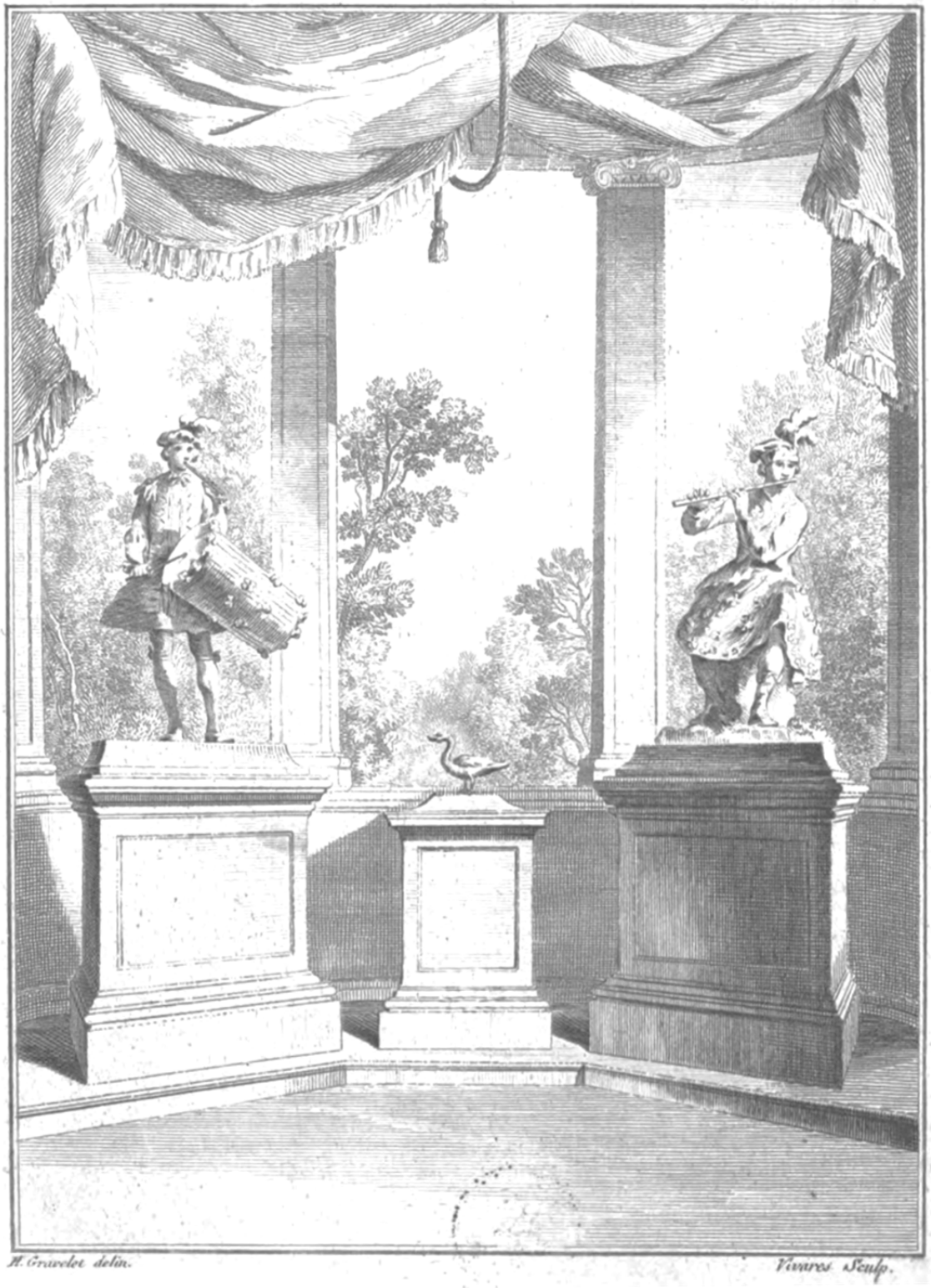


Figure 18. Engraving depicting Vaucanson's three automata. From Jacques de Vaucanson, Le mécanisme du fluteur automate: presenté a Messieurs de l'Académie royale *(Paris: Chez Jacques Guerin, 1738)*

> the wonder generated from the Flute Player derived less from its form as a realistic automaton, and more from its functional ability to play *as if it were a human musician*…. What made Vaucanson's automata so remarkable was their mimetic ability to actually perform their tasks. Up to this point, automata had merely copied external appearances, but Vaucanson's Flautist could actually perform by playing the flute…. [It] was the first automaton that actually performed the action that it represented.[107]

Reilly characterizes the Flute Player's ability to *actually play the flute*—or, more correctly, *a* flute, as one could substitute any similar German traverse flute and the automaton could play that one as well—as a form of mimesis. However, I would place this behavior in the realm of Riskin's *simulation*, as Vaucanson's *Flute Player* is not merely *imitating* a human flautist but has for the purposes of the exhibition *become* one. It is also possible that this automaton could be considered the first *cyborg*—a man/machine hybrid—due to the incorporation of organic matter into his construction. While most accounts relate that the lips, tongue, and fingertips are covered with a soft leather to make them supple, Vaucanson's original French description—"chaque bout de doigt est garni de *peau*, pour imites le mollesse du doigt naturel"[108] (each fingertip is lined with *skin*, to imitate the natural softness of the finger)—is ambiguous on this point. According to Gaby Wood, "the word *peau* in French makes no distinction between animal and human skin; but Vaucanson certainly never specified one or the other."[109] Regardless, the deliberate introduction of organic matter here serves to further facilitate the collapse described by Riskin by "simulating life's textures and substances"[110]—the construct's hybrid nature bringing the living and mechanical in even closer interplay.

---

[107] Reilly, *Automata and Mimesis*, 85; emphasis mine.
[108] Jacques de Vaucanson, *Le mécanisme du fluteur automate: présenté a Messieurs de l'académie Royale des sciences* (Paris: Chez Jacques Guerin, 1738), 11.
[109] Gaby Wood, *Edison's Eve: A Magical History of the Quest for Mechanical Life* (New York: Anchor Books, 2002), 26. Human or not, the use of skin is one of expediency rather than principle: the organic matter necessary to Vaucanson's Flute Player could easily be replaced by any number of synthetic materials today.
[110] Riskin, "Eighteenth Century Wetware," 100.

Vaucanson's *Tambourine Player* is similar in nature to the flautist: also life-sized (around 5 feet tall), the figure plays the pipe and drum, a combination quite familiar to audiences of the time. Unlike the flautist, however, which sought to imitate a human player as closely as possible, the pipe-player displayed superhuman talent:

> the pipe was played so fast that it exceeded the speed any living person could achieve (only a mechanical tongue could move that quickly), and it was also the instrument known to be the most exhausting to the human chest—the android, however, never got tired. So the Pipe Player embodied one idea of perfection, the idea that humans were messy, imperfect, fallible, and that a perfect machine would correct those flaws, improve on humanity.[111]

It is important to note that both Vaucanson's flautist and his piper play *wind* instruments. It is quite easy to design a clockwork mechanism to play an instrument that must be struck (such as a drum) or strummed (such as a lute). What was remarkable about these automata was that they *breathed*: "even though Vaucanson unveiled the fact that this breath was created by bellows, the very act of breathing, seen in an inanimate figure, continued to cause a stir well into the following century."[112]

*Le Canard Digérateur (*The *Digesting Duck,* 1938) in particular was a marvel of engineering and ingenuity: meticulously constructed (each wing contained over four hundred parts to simulate flapping), the creature appeared to eat grain, drink water, and defecate onto a small plate, much to the delight of the audience. According to historian Joseph Spence (1699–1768), who had the opportunity to witness this display in 1739,

> this duck eats, drinks, digests, and sh-ts. Its motions are extremely natural; you see it eager when they are going to give him his meat, he devours it with a good deal of

[111] Ibid., 25.
[112] Ibid., 24-25.

> appetite, drinks moderately after it, rejoices when he has done, then sets his plumes in order, is quiet for a time, and then does what makes him quite easy.[113]

While the duck did not *actually* digest the grain it ate—swallowed food was collected in one container while preformed "pellets" were released from a separate container—Vaucanson firmly believed that such a process *could* be mechanized given the proper materials.[114] The Digesting Duck, while admittedly not a *human* automaton, is nonetheless a lynchpin in the history of the artificial human. Here for the first time was an automaton whose explicit purpose was to demonstrate the most basic of physiological processes: Descartes' *bête machine* made real.[115]

Vaucanson's automata were more than mere amusements; rather, they offered physical proofs-of-concept for the mechanistic philosophy outlined by Descartes. The trio were commercial successes as well, although financial gain does not appear to be Vaucanson's primary motive. Vaucanson saw these automata as *proofs-of-concept*: he carefully studied how a physiological process worked in a living subject and then recreated this physiological action in a machine. In his presentation to the Academy of Science regarding the construction of his *Flute Player*, Vaucanson's detailed "objective description" of the German traverse flute and the process by which it is played is, according to David Fryer and John Marshall, "really a hypothesis concerning the mode of operation of man plus flute…. Vaucanson clearly regards his automaton as a test of the principles he has formulated: '*It will then follow ... according to the Principle* settled in my First Part, the flute will give a low Sound: and *this is confirmed by Experience*.'"[116] Vaucanson was well versed in the latest in anatomical research, having studied anatomy

---

[113] Joseph Spence, *Letters from the Grand Tour* (Montreal: McGill-Queen's University Press, 1975), 413-14.

[114] Indeed, such a machine was built in 2000 by Belgian concept artist Wim Delvoye. While admittedly looking nothing like a duck, nor any other animal, Delvoye's "Cloaca Machine" digests food and turns it into excrement, which is bagged and sold to art collectors and dealers worldwide. See Michael Amy, "The Body as Machine, Taken to Its Extreme," *New York Times* 151, no. 52004 (2002). See also Wim Delvoye, "Cloaca," accessed 23 January 2014. http://www.wimdelvoye.be/.

[115] See Reilly, *Automata and Mimesis*, 82.

[116] David M. Fryer and John C. Marshall, "The Motives of Jacques De Vaucanson," *Technology and Culture* 20, no. 2 (1979): 259-60, emphasis in original.

under famed surgeon Claude-Nicolaus Le Cat (1700–1768).[117] According to historian Joan Landes, Vaucanson's automata "borrowed directly from the materials, methods, and display techniques of the plastic and anatomical arts."[118] Vaucanson's automata were intended as *anatomie mouvante* (moving anatomies, a phrase coined by Vaucanson), designed to "reproduce as faithfully as possible the organs and functions of human beings or animals; their real role was not to divert and amuse but to teach, and so promote medical progress."[119] Vaucanson's descriptions to the Academy of Sciences bear "a strong resemblance to the point-by-point map of the human body offered by Descartes a century earlier."[120]

Like the Vaucanson trio, the Jaquet-Droz figures actually perform their actions rather than merely mimicking them. Built some thirty years after Vaucanson's automata, the members of this "mechanical family" are amazingly lifelike in their behaviors. The Draughtsman depicts a young boy, seated at a small desk with pencil in hand, who can draw any of four different scenes on command. The boy inspects his progress as he works, periodically blowing dust from his pencil. The Musician plays a melody on a keyboard designed specifically for her. According to John Cohen, this young lady is remarkably animated in her performance:

> A seated young woman is playing an organ-like instrument, her limbs being operated by levers, as in Vaucanson's flute player. She simulates human breathing by the rise and fall of her bosom, her head moves, and her eyes appear to be looking at her hands, the music or the audience. As if to read and hear better, she bends over the score, and at the end of her performance she bows politely.[121]

---

[117] André Doyon and Lucien Liaigre, *Jacques Vaucanson, Mécanicien De Génie Par André Doyon Et Lucien Liaigre* (Paris: Presses Universitaires de France, 1967), 18-19. See also Wood, *Edison's Eve*, 52-54 for a discussion of Vaucanson's relationship with Le Cat.

[118] Joan B. Landes, "The Anatomy of Artificial Life: An Eighteenth-Century Perspective," in Jessica Riskin, ed. *Genesis Redux: Essays in the History and Philosophy of Artificial Life* (Chicago: University of Chicago Press, 2007), 99.

[119] Catherine Cardinal, "Preface," in Vaucanson, *Le mécanisme*, xv-xvi; quoted in Landes, "The Anatomy of Artificial Life," 101. See also Kang, *Sublime Dreams*, 140-142.

[120] Wood, *Edison's Eve*, 24.

[121] Cohen, *Human Robots in Myth and Science*, 87-88. See also Adelheid Voskuhl, *Androids in the Enlightenment: Mechanics, Artisans, and Cultures of the Self* (Chicago: University of Chicago Press, 2013), Kindle edition, chapter two for a detailed account of *La musicienne* in action and her historiocultural place in the debate on the human-machine boundary.

Figure 19. Photo of the Jacquet-Droz trio of automata, housed in the musée d'art et d'histoire de Neuchâtel; from left to right, The Draughtsman, The Musician, and The Writer. Photo by Rama, licensed under the Creative Commons Attribution-Share Alike 2.0 (France) license.

The Writer, also a young boy seated at a desk, is the most complex of the three figures. Considered by many to be a precursor of the modern computer (and predating Babbage's Difference Engine by fifty years), this automaton is able to write any pre-programmed message of up to forty letters. Like his siblings he is rather animated in his performance: his eyes follow the goose-quill as he dips it in ink, shakes off the excess, and carefully forms each letter on the page.

The Jaquet-Droz "mechanical family," while strongly influenced by Vaucanson's earlier work, nevertheless exhibits some fundamental differences. First, the Jaquet-Droz trio are a great deal more complex in both their movement and their construction than the Vaucanson automata, building on the earlier technology in ways Vaucanson did not anticipate. According to Roland Carrera,

> when Vaucanson went to see the automatons and was unable to understand how the limbs worked before he had examined them, he said to Jaquet-Droz: "Young man, you're beginning where I've ended..."
>
> Vaucanson had used almost the same mechanical system as Jaquet-Droz; but he never went to the same lengths as his famous junior in exploiting all the different possibilities.[122]

Whereas the Defecating Duck, Vaucanson's most advanced automaton, appears to have consisted of a little over 1,000 movable parts, each Jaquet-Droz figure is a miracle of clockwork intricacy: the Draughtsman consists of over 2,000 individual mechanical pieces; the Musician, 2,500, and the Writer, over 6,000, including an individual cam for each of the forty characters of his custom message and a separate encoding wheel with which to program it. In addition, the Jaquet-Droz automata are much more highly anthropomorphized than the Vaucanson, both by their creators themselves and (possibly) by their later owners. The Jaquet-Droz figures have also been given personal names and backstories at some point in their histories: "Charles [the Writer] and Henri [the Draughtsman] are both three-year-old children, whereas Marianne, their older sister, is 14. All three children are prodigies."[123]

Most importantly, however, the Jaquet-Droz trio perform not only the behaviors required by their activities (drawing, writing, playing the organ) but also additional actions designed to further the illusion of life, such as eye-tracking behavior (the boys appearing to watch their hand movements) and breath. While Vaucanson's flute- and pipe-players rely on a complex bellows system to play their

---

[122] Roland Carrera, Dominique Loiseau, and Oliver Roux, *Androids: The Jaquet-Droz Automations* (Lausanne, Switzerland: Scriptar, 1979), 16..

[123] Reilly, *Automata and Mimesis*, 87. Reilly offers a detailed description of the three figures on pages 87-91, although she does not mention the origin of their names. Carrera et. al. also describe the three figures in great detail in their 1979 monograph; while the figures are highly personified, they are not actually called by name in the text. The illustration of the three from the frontispiece of the Carrera, et. al. work is inscribed with their given names, but the three are typically referred to as "The Draughstman," "The Writer" and "The Musician." A personal email dated 9 May 2014 from Dr. Christian Hörack of the Ville de Neuchâtel musée d'art et d'histoire indicates that "As far as we know, these names of the automata appear only at the very end of the 19th century. They refer to members of the Jacquet Droz family." The tourism website for the city of Neuchâtel further describes the three as its "most famous celebrities" (Ville de Neuchâtel, "Vivre La Ville! Inhabitants: Automata," accessed 12 February 2014, http://en.neuchatelville.ch/profils/residents.asp/1-11-160-21345-10001-1001-1-1-2-1/2-0-1345-10001-1000-2-0/).

instruments in a simulation of human performance, Marianne has no such requirement to play the keyboard of the organ. Yet she breathes nonetheless: "her chest heaves as she sighs in tune with the music, as though the music affects her emotionally."[124] Her gaze as she plays is similarly lifelike: "In three-quarters of an hour the movement of the eyes is never the same, making it appear perfectly like a human being."[125] While the Vaucanson automata were built as mechanical simulacra of functional physiological processes, the Jaquet-Droz family go beyond what is *required* of humans performing the same activities to what is *expected* of such a performance: rather than merely going through the motions of her organ performance, for example, Marianne appears to be *affected* by it, simulating the sentimentality her audience would expect of any musical performance of the time.

Ronald Carrera describes the trio's behavior in explicitly mimetic terms: "While it's certainly true that the Jaquet-Droz automatons do not exactly copy the processes of nature, the harmonious working of the bones and sockets and of the human body, they do however give a perfect imitation of the effects."[126] These additional performative actions by the Jaquet-Droz trio are again mimetic in nature, but they represent a mimesis of another order entirely. The Musician, the most complex of the trio in respect to this performance, is equipped with a dedicated "life mechanism" that operates independently of the action of her musical performance, assuring "in a totally independent manner the movements of breathing, the eyes, and the head…. Conceived with astonishing perfection, this mechanism makes the Musician the only automaton to possess a supreme gift, breathing: in other words, life itself."[127] The Musician exhibits two simultaneous performances: the first is the performance of "human" via the "life mechanism;" the second, the individual songs she performs in her role as musician. The Musician's performance as "Musician" comprises a repertoire of eight different melodies, each with a single, predictable program designed to last a few minutes. Her performance of "human,"

---

[124] Reilly, *Automata and Mimesis*, 90.
[125] Carrera, et. al, *Androids*, 78.
[126] Ibid., 18.
[127] Ibid., 76. Carrera is rather hyperbolic here in his description of the Musician as the only breathing automaton, although she is certainly the only one of the time still in operation.

Figure 20. Close-up of Jaquet-Droz' Musician, housed in the musée d'art et d'histoire de Neuchâtel. Unlike her brothers, the Musician's fingers are individually articulated so that she may actually play the custom-built instrument before her. Photo by Rama, licensed under the Creative Commons Attribution-Share Alike 2.0 (France) license.

on the other hand, lasts over an hour and a half, including randomized eye movement, which over a 45-minute period "is never the same, making it appear perfectly like a human being."[128] During the course of a single demonstration, the Musician will begin with a display of her humanity, giving "the impression that the Musician is waking, and preparing its audience to receive the musical message."[129] She will then play one or more melodies, as programmed, ending each with a curtsy during which she "seems to take on an amused air:"[130]

[128] Ibid., 78.
[129] Ibid., 76.
[130] Ibid..

> the Musician, when making her curtsy, is not content merely to incline the bust, but also turns her head to the left while lowering it, turns her eyes to the left and right, and up and down. Without this particular provision the masterpiece of Jaquet-Droz would lack a most important element: humour.[131]

While earlier automata offered a mimetic performance of "musician" by *appearing to play an instrument*, and the Vaucanson automata a simulation of "musician" by *actually playing the instrument*, the eye-rolling and breathing exhibited by the Jaquet-Droz trio functions as mimesis of "human" by appearing to perform behaviors that are recognized as markers of life: the imitation of physiological processes, the display of the visible indications of thought processes, of emotion, and the apparent randomness of free will.

The Vaucanson and the Jaquet-Droz automata represent a form of *functional equivalence*[132] to their living counterparts; namely, the programmed behaviors of the constructs perform the same function as the behaviors of the humans they simulate, often with either the same or very similar processes involved. Marianne Droz, and to a lesser extent her brothers Charles and Henri, additionally represent the first attempt at what I would call *social equivalence*, in that their programmed repertoires go beyond functionality into the realm of behaviors that are socially encoded and performed for the express purpose of facilitating social interaction. For the first time, actual automata (as opposed to those from literature or legend) deliberately perform social behaviors common to their human counterparts, a phenomenon that Adelheid Voskuhl explores in her recent work *Androids in the Enlightenment*: "The artisan automaton makers thus manifested mechanically in their android's body

---

[131] Ibid., 74.

[132] In his treatise *Psychological Explanation*, philosopher and cognitive scientist Jerry A. Fodor explains functional equivalency in machine behavior thusly:

> Let us say that a machine is *weakly equivalent* to an organism in some respect when the behavioral repertoire of the machine is identical with the behavioral repertoire of the organism in that same respect. We can then say that a machine is *strongly equivalent* to an organism in some respect when it is weakly equivalent in that same respect and the processes upon which the behavior of the machine is contingent are of the same type as the processes upon which the behavior of the organism are contingent. (*Psychological Explanation: An Introduction to the Philosophy of Psychology* (New York: Random House, 1968), 137-38, emphasis mine.)

practices that cultivated sociability, sentiments, and practices of subject- and society-formation."[133] These early attempts at behaviorally equivalent simulacra represent what Jessica Riskin argues are the first explorations into the field of artificial life[134]—and additionally foreshadow the behavioral definition of "human" that I develop in the next chapter.

As automaton-creators were collapsing the distance between the living and the mechanical in their creation of lifelike dolls, the aristocracy was busily undermining the difference between the two in its pursuit of the perfect doll-like human. This trend may be best illustrated in the figure of Marie Antoinette (1755–1793) who, according to Kara Reilly, "epitomizes the paradigm of the ideal as a perfectly constructed creature wound up and set into motion to rule her kingdom."[135] As a young woman Marie Antoinette was trained in courtly manners, dance and music, and held a fondness for theatre and dance throughout her lifetime. One of her dance instructors was the great ballet master Jean-George Noverre (1727–1810), who later benefitted from her royal patronage. In his discussion of "the systematization of the theatrical passions" during the Enlightenment, Joseph Roach credits Noverre with one of the earliest attempts to create a "linguistics of expression" in describing ballet movements in biomechanical terms:

> What is Ballet but a piece of more or less complicated machinery, which strikes or surprises the beholder by its various effects, only in proportion as those are diversified and sudden? That chain and connection of figures, those motions succeeding each other with rapidity, those various forms turning contrary ways, that mixture of different incidents, the ensemble and harmony which mark the steps and accompany the

---

[133] Voskuhl, *Androids in the Enlightenment*, Locations 164-65. Voskuhl focuses particularly on Marianne (the Jaquet-Droz organist described above) and the German dulcimer player built in 1784 by furniture-maker David Roentgen and clock-maker Peter Kinzing for Marie Antoinette, which appears in the following discussion (see Figure 21). She distinguishes these two from a group of about ten built during this period, including the Vaucanson and Jaquet-Droz automata and a few select others, because while all ten "allow us to draw several conclusions about mechanical arts, industrialism, and the human-machine boundary in the period on the cusp of the Industrial Revolution" (Ibid., Locations 130-132), the two female figures in particular "replicate in a more comprehensive and far-reaching way than the others, through their distinctive mechanical design, an entire setting of eighteenth-century cultural and political activity" (Ibid., Locations 143-145).

[134] See Riskin, "Eighteenth Century Wetware." See also Riskin, *Genesis Redux*.

[135] Reilly, *Automata and Mimesis*, 95.

> exertion of the dancers; do not all these give you the idea of a mechanism most ingeniously contrived?[136]

Kara Reilly further adds that "famous for her ability to move with delicacy and grace, Marie Antoinette had been trained by Noverre to move as 'a mechanism most ingeniously contrived.'"[137] At the pinnacle of French society, Marie Antoinette set the standard for style and courtly manners of the time. Courtly fashion "gave courtiers the appearance of being like mechanized dolls with gigantic powdered periwigs topping their heads, and large round marks of red rouge colouring their cheeks;"[138] "proper" courtly manners were modeled after the outward expressions of Cartesian passions as depicted in Charles Le Brun's *Méthode pour apprendre à dessiner les passions* (*Method to learn to draw the passions*, 1698):

> The king's face revealed the extreme limit of the court mask, designed to hide intentions by dissimulating the expressions that are their signs. This excess of calculation in the courtesans' expressions, this inscrutability, this prefiguration of expression—what Gracian spoke of as a "politics of the face"—was the sign of the perfect courtesan, and by extension the perfect politician, akin to the automaton of Descartes' mechanical model of a man.[139]

As automata were being designed to simulate people, the "best" people were being trained to emulate these automata; culminating in Marie Antoinette's decision to commission the construction of a tympanon-playing automaton in her likeness by the German pair David Roentgen and Peter Kinzing in 1784.[140] Donated by the Queen to the Royal Academy of Sciences, the automaton is "a celebration of the mechanized, perfect doll aristocrat of *ancien régime* France [that] demonstrates that courtiers

---

[136] Jean Georges Noverre, *The Works of Monsieur Noverre*, 3 vols. (London, 1782; New York: AMS Press, 1978); quoted in Roach, *The Player's Passion*, 76.
[137] Reilly, *Automata and Mimesis*, 93.
[138] Ibid., 94.
[139] Allen S. Weiss, *Mirrors of Infinity: The French Formal Garden and Seventeenth-Century Metaphysics*, (New York: Princeton Architectural Press, 1995), 25.
[140] See Voskuhl, *Androids in the Enlightenment* for a detailed analysis of this automaton.

Figure 21. Marie Antoinette, the automaton dulcimer player. Still from video "Demonstration of David Roentgen's Automaton of Queen Marie Antoinette, The Dulcimer Player," Metropolitan Museum of Art, Published on YouTube Nov 20, 2012, https://www.youtube.com/watch?v=nITEU4fsqCU, accessed 25 September 2014. For a longer demonstration of her mechanism and maintenance, see also http://www.cerimes.fr/articles/article_950/la-joueuse-de-tympanon.

aspired to be as perfectly mechanized as their celebrated automata."[141] Like her cousin Marianne Droz, the Marie Antoinette automaton is crafted to perform mimesis and simulation on two separate levels: not only to play her instrument with expertise but also to simultaneously demonstrate her emotional reaction to the music she plays.

The phenomenon of automata-as-simulacra faded in the 1790s as the commercial popularity of automata-as-amusements grew. While automata of the nineteenth century were "complicated displays, ultimately filled with the preoccupations of the Belle Époque, … [they] were markedly less ambitious than automata of the preceding century. Or, more precisely, their ambition was not to simulate"[142] but rather to offer clever imitations, designed to amuse and entertain. According to some scholars, the link between automata and the aristocracy led an increasingly disgruntled populace to equate mechanistic

[141] Reilly, *Automata and Mimesis*, 94.
[142] Riskin, "Eighteenth Century Wetware," 101.

ideas with the clockwork mindlessness of an increasingly totalitarian regime.[143] For others, the broad application of mechanistic explanations to all walks of life exposed flaws in its efficacy as an epistemological metaphor. Clocks broke, planets were not *literally* held in their spheres by gears and cables—and above all, the idea of man as an automaton flew in the face of the common-sense knowledge that man possesses free will. Moreover, universal mechanism presented an ontological crisis, manifested in the creation of automaton simulacra. Even as mechanistic philosophy reached its peak in popularity in the early eighteenth century, undercurrents of its vitalistic repudiation were evident.

IDIOT AUTOMATA AND THE VITALISTIC PRINCIPLE

Minsoo Kang claims that the 1748 publication of Julien Offray de La Mettrie's (1709–1751) *L'Homme Machine* (*Machine Man*) "marks the end point of the classical mechanist movement that commenced in the 1630s"[144] by offering a radical departure from Cartesian mechanism; according to Kang, "the exact type of machine La Mettrie pictures man to be is a different one from that envisioned by mechanists, namely an organic perpetual-motion machine that is better described as a dynamic living machine.… [La Mettrie] utilized the concepts and language of mechanism but went beyond some of its fundamental tenets in explicating the nature of life in materialist terms."[145] The change Kang criticizes here is the re-introduction of *vitalism*, a belief that living organisms possess some "vital energy" that renders them fundamentally different from non-living matter—a belief that is antithetical to the mechanistic philosophy of Descartes and his contemporaries of a century earlier.

This was not the first clash between the principles of mechanism and vitalism: since antiquity philosophers have debated on the very nature of matter and of being.[146] At the time of Descartes' *Meditations on First Philosophy*, the dominant metaphysical model was Aristotelian. In *De Anima* (*On*

---

[143] See Kang, *Sublime Dreams*, chapter four; Mayr, *Authority*, chapters four through six; and Reilly, *Automata and Mimesis*, chapter three.
[144] Kang, *Sublime Dreams,* 131.
[145] Ibid., 130.
[146] For a discussion of this debate and its relationship to artificial life, see Riskin, "Introduction," in *Genesis Redux,* 1-32.

the Soul, c. 330s BCE), Aristotle posits the existence of a *psyche*, or soul, as the foundation of his metaphysical inquiries, beginning with the premise that "the ensouled is distinguished from the unsouled by its being alive."[147]Aristotle goes on to identify five faculties of the soul and arranges them in a hierarchical manner in order to distinguish between plant, animal, and human:

> Now of the faculties of the soul, some living things have all those that we have talked of, as we said, some have some of them, and some only one. The faculties we spoke of were the *nutritive, perceptive, desiderative, locomotive and intellective* ….We must, then, seek out in each case what the soul of each thing is, what for instance is the soul of a plant and what of a man or a beast, and we must consider for what reason it is that they stand thus in series. For while there is never the perceptive faculty without the nutritive, in plants the nutritive is separated from the perceptive. And again without the sense of touch none of the other senses is present, but touch can be present without the others, many of the animals having neither sight nor hearing nor the sense of smell. And of those creatures that perceive, some have the locomotive faculty, others do not. Ultimately and most rarely, some have reasoning and thinking. For those of the perishable beings that have reason have all the other faculties but not all of those that have each of them have reason; indeed, some do not even have imagination, while others live by this alone. About the contemplative mind, however, another account will be needed.[148]

At its most basic level, an ensouled (living) being possesses the *nutritive* faculty, which for Aristotle includes not only nutrition itself, but also growth and decay as well as reproduction. Higher-level living things can also perceive (through various senses), desire (or possess *will*), move (by which he implies self-motion), and think (which includes reason, imagination, and contemplation). In her essay "The

---

[147] Aristotle, *De anima* [*On the Soul*], trans. Hugh Lawson-Tangred (London: Penguin Books, 1986), Kindle edition, 159.
[148] Ibid., 162-163, emphasis mine.

imitation of life in ancient Greek philosophy," Sylvia Berryman addresses the question of "whether ancient technology offered anything to challenge the distinction between living and nonliving as Aristotle forms it,"[149] concluding that, while there are "scattered indications" that some thinkers may have explored the machine analogy, there is little evidence that they applied this analogy in the same way as did the philosophers of the seventeenth century.[150] The Aristotelian definition of life has dominated Western philosophy for the past two millennia, and as we shall see in the next chapter, remains with us today, forming the basis of many of the "tests" by which we determine whether or not something should be deemed "alive."

Cartesian dualism represented an attempt by Descartes to reconcile the Aristotelian soul with his growing mechanistic understanding of physiological processes, and his *bête machine* doctrine[151] safely maintained the boundary between human and animal as Renaissance science offered increasingly mechanistic explanations for faculties Aristotle held as "ensouled," such as locomotion and perception. However, the vitalistic principle as outlined by Aristotle is a difficult position to argue against. It simply seems to make sense: according to Joseph Roach, "naïve vitalism never abdicated its reign over common sense; flesh just seemed to have a life of its own."[152] But as Renaissance mechanists pushed back the frontiers of the unexplained, material proofs-of-concept of their mechanistic theories were developed in the form of automaton simulacra. The secrets once hidden inside the living flesh were laid bare by anatomists and reified in metal and skin by automata-makers: the organist Marianne Droz, perfect in

---

[149] Sylvia Berryman, "The Imitation of Life in Ancient Greek Philosophy," in *Genesis Redux: Essays in the History and Philosophy of Artificial Life*, ed. Jessica Riskin (Chicago: University of Chicago Press, 2007). 36. According to Berryman, Aristotle distinguishes categorically between living and non-living, rejecting outright the idea that such an analogy is valid.
[150] Ibid., 43-44.
[151] This highly controversial doctrine holds that man is distinguished from, and superior to, "brute" animals in in that animals do not possess a conscious soul; they are "living machines" that have no faculty for reason. Reason is the exclusive domain of man, as evidenced by his command of language. The *bête machine* doctrine still holds sway in contemporary artificial intelligence debates as to whether machines can "think" (see Newman, "Unmasking Descartes's Case," 389-426), as well as in the battle over the legal status of some animals as "non-human persons" (see Steven M. Wise, *Drawing the Line: Science and the Case for Animal Rights* [Cambridge, MA: Perseus Books, 2002]).
[152] Roach, *The Player's Passion*, 93.

her poise and performance even 250 years after her construction, also seems to have "a life of her own."

By the end of the eighteenth century the comparison of an individual to an automaton was largely a derogatory one. Minsoo Kang identifies three trends during the eighteenth century that contributed to this change: first, the failure of the mechanistic metaphor in medical sciences; second, the rise of sentimentality; and third, the increasing application of mechanism to political philosophy.[153] To this I would also add the Church's rejection of universal mechanism's attempt to "explain away the soul." All of these trends represent the rising ontological collapse triggered by the metaphor of *man=machine*: as the boundaries between the living and the non-living collapsed, the liminal figure of the automaton again fell under suspicion. Ultimately, however, the rising wave of vitalism during the Enlightenment was not so much a rejection of mechanism as it was an amelioration of it. "Yes, we *are* all automata," this particular manifestation of the vitalistic principle seems to say, "but we are an altogether different sort of machine than the clock."

This shift from mechanism back to vitalism during the Enlightenment is an important one in the history of the artificial human. For the first time, the ontological status of *human* itself was called into question, assaulted by the mechanistic repudiation of Aristotelian humanocentrism. Vitalistic reactions to mechanism during this time period may be grouped into three basic forms: the argument from biology, the argument from sentimentality, and the argument from autonomy. The first, the argument from biology, simply states that mechanistic principles cannot account for *all* faculties of a living being; that some vital force (*élan vital*) is present in living matter that makes it fundamentally different from nonliving matter. The argument from sentimentality addresses the capacity for *sentiment*, or strong emotion, as something possessed by the ideal man or woman, without which a human being may only be considered "incomplete." The final argument, the argument from autonomy, addresses the concepts

[153] Kang, *Sublime Dreams*, 148 ff. For Kang's complete argument, see his chapter four, "From the Man-machine to the Automaton-man, 1748–1793."

of spontaneity and free will. *Autonomy* is the principle of self-governance; that is, having freedom in determining one's own actions. All of these arguments distinguish the living from the nonliving—and in addition, the human from the nonhuman—through the presence in the former of a *soul*. The soul here represents the motive force, the seat of consciousness, the "I" in Descartes' legendary "I think, therefore I am" formula. In the next chapter, I argue for a slightly different definition of *human* (or more precisely, of *personhood*) that is behavior-based rather than contingent upon an immeasurable and unprovable soul as a necessary and sufficient condition. First, however, I would like to briefly address each of these vitalistic arguments in order to paint a more comprehensive picture of the state of the ontological crisis brought about by the mechanistic philosophy of the seventeenth century, ending with a discussion of the first fictional account of an artificial human to directly address this ontological crisis: Mary Shelley's *Frankenstein* (1818).

A SHOCKING DISCOVERY: THE ARGUMENT FROM BIOLOGY

The developing mechanistic paradigm of the Renaissance offered a model of living flesh that was based in pneumatics and clockwork. The nervous system, for example, was envisioned in the Galenic tradition as a network of hollow "tubes" through which "animal spirits" moved the body. Yet as empirical evidence gathered through experimentation by anatomists such as Vesalius and William Harvey revealed inadequacies in this paradigm, physicians struggled to find an alternative explanation. In 1662 English physician Francis Glisson (1598–1677) "described a nonmechanical living force present in all substances that manifested itself as 'irritability' in animal fiber."[154] Dutch naturalist Jan Swammerdam (1637–1680) also found in the 1660s that the muscles of small animals were able to move even after having been removed from the body through vivisection, "completely disconnected from the supposedly

[154] Ibid., 149. "Irritability" was defined by Swiss naturalist Albrecht von Haller (1709–1777) as a specific property of the muscle which consisted in the tendency of a muscle to contract in response to any external (or internal) stimulus (or "irritation").

indispensable tubes"[155] described by earlier anatomists. While Swammerdam's findings were not published until much later, his experiments were replicated in the early 1700s by scientist Pierre Polinière (1671–1734), whom Joseph Roach labels a "popularizer of science,"[156] as he performed many public demonstrations of his work. Polinière's *Expériences de Physique* (*Physics Experiments*, 1709) was enjoying its fifth edition by the time Swammerdam's own work was published. Polinière's surprise at what he observed reflects that of Glisson, Swammerdam, and numerous others of the time:

> Pierre Polinière freely confessed his mystification. On one side of the dissection table he had placed the body of a frog that still showed signs of vitality. On the other side of the table he placed the frog's heart, which, even more perplexingly, continued to beat after it had been removed from the body. All the tubes connecting the heart to the brain had been severed. No animal spirits could exert pressure on the throbbing muscle. Nevertheless … the heart's spasms persisted "for about an hour" following vivisection. Obviously, he could not explain this phenomenon by means of the paradigm of mechanical transmission available to him.[157]

The failure of the mechanist principles of pneumatics—the tubes and animal spirits upon which early mechanistic medicine relied—to explain the functioning of the living body led to a discernible shift in medical research in the eighteenth century, as physiologists became "more concerned with understanding the operation of the vital force theorized by Glisson"[158] and others. What was the nature of this vital force, if not a fluid?

Simultaneous with the medical exploration into the animating force behind living tissue, seventeenth and eighteenth century physicists were investigating another natural phenomenon: electricity. Although people had been aware of the effects of electricity since ancient times—reports of

---

[155] Roach, *The Player's Passion*, 93.
[156] Ibid., 94.
[157] Ibid., 93.
[158] Kang, *Sublime Dreams*, 149.

electric fish are found as far back as ancient Egypt, and their use as a form of medical treatment can be traced to early Greece and Rome—there was no systematic exploration of the phenomenon until the late sixteenth century, when English scientist William Gilbert (1544–1603) studied the effect of rubbing amber in order to make it attract light objects such as feathers. Dubbed *electricity* after the Greek word for amber, this phenomenon was studied extensively over the next few centuries: German inventor Otto von Guericke (1602–1686) discovered the principle of electrostatic repulsion; English astronomer Stephen Gray (1666–1736) was one of the first to experiment with electrical conduction (as opposed to static electricity); French chemist Charles François du Fay (1698–1739) discovered two different types of electricity, namely *vitreous* and *resinous* (now known as positive and negative charge respectively), as well as the existence of conductors and insulators; and of course American polymath Benjamin Franklin's (1706–1790) famous experiments in the 1750s proved conclusively that lightning was a form of electricity.

Enter Luigi Galvani (1737–1798): Italian physicist, physician, and philosopher. After studying medicine at the University of Bologna, Galvani took a position at the university as a lecturer in surgery and theoretical anatomy, although he "showed a continuous interest in electricity, and kept updated with the progress in the discipline through the most important works published in the field, which enriched his personal library."[159] During the mid-eighteenth century there was a great deal of interest in the field of "medical electricity," or the therapeutic application of electricity to the human body (much as ancient Greek and Roman physicians did with their "electric" fish), a practice that "found in Bologna [where Galvani was located] a very fertile field."[160] Aware that the therapeutic application of electricity required a deeper understanding of the underlying physiological mechanisms, Galvani conducted an extensive series of experiments into the effect electricity had on living tissue, using the frog as his

[159] Marco Bresadola, "Medicine and Science in the Life of Luigi Galvani (1737–1798)," *Brain Research Bulletin* 46, no. 5 (1998): 368.
[160] Ibid., 371.

subject (a common practice at the time): "Galvani discovered to his surprise that the contractions in the long muscle of a prepared frog could be triggered at a distance by sparks from a static electricity machine or even by flashes of lightning in the atmosphere. He concluded that nerves possess an innate fluid or current that can excite or be excited at a remove."[161] This discovery led him to develop the theory of animal electricity, concluding in 1780 that the "vital force" that so confounded his mechanistic predecessors was none other than bioelectricity. While legend holds that Galvani was "surprised" by his results, historians such as Marco Bresadola and Marco Piccolino attribute this unfavorable characterization to the antagonistic relationship between Galvani and his contemporary Alessandro Volta (1745–1827) who, although confirming Galvani's experiments, disagreed with and ultimately disproved some of the details of Galvani's theories.[162]

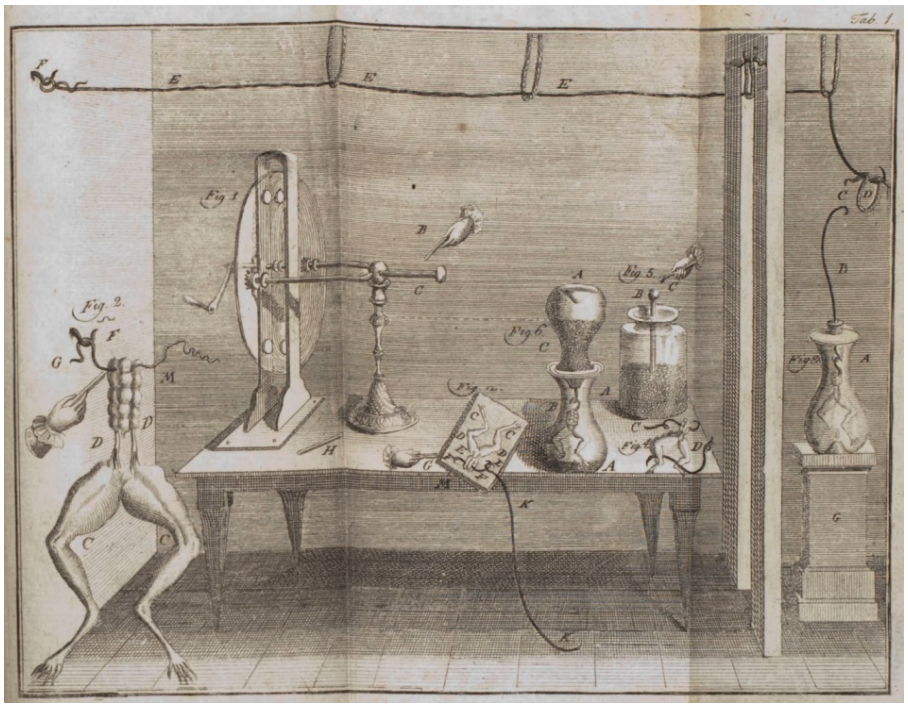


Figure 22. An illustration depicting some of Galvani's laboratory equipment, including his famous frog experiment. From Aloysii Galvani, De viribus electricitatis in motu musculari commentarius *(Bologna: Ex Typographia Instituti Scientiarium, 1791)*

[161] Roach, *The Player's Passion*, 95.
[162] See Marco Piccolino and Marco Bresadola, *Shocking Frogs: Galvani, Volta, and the Electric Origins of Neuroscience* (London: Oxford University Press, 2013).

Galvani's treatise *De viribus electricitatis in motu musculari commentaries* (*Commentary on the effect of electricity on muscular motion*, 1791) caused an immediate sensation among the scientific community, as his supporters hailed it as revolutionary while his detractors sought to refute his findings.[163] His memoirs were widely distributed and highly sought after, his theories the subject of much discussion among the *salons* in vogue at the time. While the controversy between he and Volta over the exact nature of "animal electricity" raged on (Galvani held the vitalistic position that the frog's legs were powered by a chemically-induced "animal electric fluid" intrinsic to living material, while Volta claimed that the reaction was a physical phenomenon caused by rubbing the frog's skin—a more mechanistic explanation), Galvani continued his investigations until 1797, when his refusal to swear loyalty to the newly-founded Cisalpine Republic resulted in his being stripped of all academic and public positions; he died in poverty the next year. After Galvani's death, Volta went on to prove that, while the electricity found in the frog's muscles was indeed chemically produced (a process that Volta named *galvanism* out of respect for his fellow scientist) , it was not "innately" animal, creating an early form of electric battery in the process (the *voltaic pile*, 1800). Though incorrect in some of the details, Galvani's groundbreaking work left its mark in the modern field of electrophysiology, which studies the electrical properties of biological cells and tissues and has far-reaching application in a broad range of medical disciplines. Thus the vitalistic quest for an "innate" animal force that could not be explained through the mechanistic principles of pneumatics and clockwork was instead satisfied through the mechanistic principle of electricity.

SENSE AND SENSIBILITY: THE ARGUMENT FROM SENTIMENTALITY

The idea of bioelectricity as adhering to mechanistic principles, however, is a modern one. During the Enlightenment, this "animal living force" that manifested itself in "irritability"—

[163] See Bresadola, "Medicine and Science," *Brain Research Bulletin* 46, no. 5 (1998): 376-377, as well as its companion article, Marco Piccolino, "Animal Electricity and the Birth of Electrophysiology: The Legacy of Luigi Galvani," *Brain Research Bulletin* 46, no. 5 (1998): 381-407.

responsiveness to external and internal stimuli—was seen as the hallmark of the emerging paradigm of *sensibility*: "an inherent bodily capacity, differing markedly between individuals, that registers and communicates feeling,"[164] as opposed to mechanism, which reduced individuals to mere machines, and human passions to bodily functions. Descartes' *Les passions de l'âme* (*Passions of the Soul*, 1649) represented a mechanistic attempt to explain passions as the "manifestations of a moving machine propelled by its own physical springs and guided by its ghostly soul"[165]—a position that Joseph Roach claims "objectified and exteriorized the passions, draining them of their subjective content."[166] Sensibility, on the other hand, returned to individuals the capacity for individually varied response based on both external stimuli (the world around oneself) and internal stimuli (one's internal thoughts and feelings, or *sentiments*). Sensibility allowed the machine-man to rise above his mechanistic nature, positioning vitalism not as an outright rejection of mechanism but as a further elaboration on it.

The rise of sentimentality was reflected in the literature of the time, and while a thorough examination of this phenomenon is outside the scope of this study, I believe a short illustration of the shift from reason to sentiment is in order. In the clockwork rationality of the mechanistic era, man was hailed for his unique ability to reason and to overcome his "animal" passions. A prime example of this is Shakespeare's *Othello*, whose struggle between passion and reason is depicted as a character flaw, and whose ultimate failure to overcome base jealousy leads to tragedy:

> Now, by heaven,
> My blood begins my safer guides to rule
> And passion, having my best judgment collied,
> Assays to lead the way.[167]

---

[164] Roach, *The Player's Passion*, 95.
[165] Ibid., 72.
[166] Ibid., 71.
[167] William Shakespeare, *Othello* (1603), from *The Oxford Shakespeare* (London: Oxford University Press, 1914), II.iii 163-166.

The ideal man (or woman) of the sentimental age, on the other hand, is one who allows himself (or herself) to be ruled by his passions—*not* to do so has become the character flaw. The protagonist of Jean-Jacques Rousseau's *Pygmalion: Scene Lyrique* (*A Lyrical Scene*, 1762) is an example of such a man of passion, and he brings to life the statue of Galatea by the strength of his passions alone:

> It seems as though fiery shafts were emanating from that object in order to enflame my senses, then to return, bearing my soul back with them to their source! Alas, it remains cold and immobile while my heart, enflamed by its charms, wants to leave its own body that it might warm the stone…. I think that I can give her my life, and animate her with my soul.[168]

Pygmalion's "hot" passions are contrasted here with the "cold" stone of which Galatea is formed. This is one of the hallmarks of the argument from sentimentality: the ideal, natural man is "hot" both in nature and in temperament, while the machine-man is as "cold" emotionally as his mechanical cousin of marble or iron is physically. A "cold" machine could never possess the sentimentality of a human being; indeed, only the best of humanity rise above this mechanical nature.

As a short scene of spoken word and pantomime set to the music of Horace Coignet, *Pygmalion* is considered by many to be one of the earliest melodramas: originally intended as text "intercut with passages where music and mime transcend the words into a higher realm of experience,"[169] the term has since come to mean a thing overly (and falsely) emotional. At the time, however, the unabashed display of passions was the mark of a noble sentiment. As Louis James explains,

> Rousseau changes the story, leaving out the plot of Venus and the depraved women, removing the erotic tone, and focuses on the creative power of the artist's passion and

---

[168] Jean-Jacques Rousseau, *Pygmalion, un Scene Lyrique* [*Pygmalion: A Lyrical Scene*] (Bibliothèque de Neuchâtel, 1762), 1227-1228, trans. 1983 by Paul De Man, from the collection "The Portable Rousseau" at University of California at Irvine, digitized April 2010, accessed 26 January 2014, http://ucispace.lib.uci.edu/bitstream/handle/10575/1094/08demanportablerousseaupygmalion.pdf?sequence=11.

[169] Louis James, "Frankenstein's Monster in Two Traditions," in *Frankenstein, Creation, and Monstrosity*, ed. by Stephen Bann, 82.

> imagination…. The effect is that of a profane Mass, with the inanimate transubstantiated into flesh and blood not by divine agency, but through the artist's emotion and creative imagination.[170]

Othello's passion was a critical flaw: acting on his emotion rather than submitting his jealous suspicions to the critical consideration of reason led to tragedy. Pygmalion's extreme passion, on the other hand, is rewarded by Galatea's "awakening"—the "flame" of his creative energy becoming the Promethean "spark" that brings her to life.

Minsoo Kang describes the difference between the rational and the passionate character type as viewed from a vitalist/sentimentalist perspective:

> The first is the ideal man of the mechanistic age, a rational individual who avoids enthusiasms and other forms of extreme passion, which can lead to everything from religious fanaticism to libertine debauchery, favoring a calm, organized, and decorous mode of living based on moderate, reasonable principles…. In direct contrast, the protagonist of the sentimental age is a passionate and sensitive nonconformist constantly struggling with the powerful emotions that overtake him, compelling him to commit actions of a grand but sometimes destructive nature…. But he is also a genius whose brilliance is fueled by his soul's undying fire…. In the literature of sentimentality, the person of the first type is described as an incomplete being and often derided as an automaton, in direct contrast to the man or woman of passion.[171]

As noted above, while man may be a machine, the analogy is incomplete, and according to the model of the period, the best of humanity is able to rise above the "limits" of pure rationality through sensibility: that is, through passions and emotions, and the individual capacity to respond to these sentiments in others. While later eras reacted negatively to the extreme display of emotion in the literature of this era

---

[170] Ibid., 82.
[171] Kang, *Sublime Dreams*, 153-154.

(both *sentimentality* and *melodrama* have since become pejorative terms), this innate capacity for emotion has been used as a test of "humanness" to this day, which I will further examine in the following chapter.

MASTER OF HIS OWN FATE: THE ARGUMENT FROM AUTONOMY

The final vitalist argument against mechanism in the eighteenth century was that regarding *autonomy*, or the ability of an individual to determine his own actions, and by extension his own fate, through the principle of self-regulation. While the philosophical intricacies of this debate fall outside the scope of this study, a brief overview of its direction during the Enlightenment is in order. The argument against mechanism from the principle of autonomy is two-pronged: on the one hand stands the role of God in the universe; on the other, the individual in the state. Both, however, may be reduced to the debate between determinism and free will. Central to our discussion of the artificial human, however, is the mechanist/vitalist debate regarding the autonomy of the individual, for herein lies the historical origin of the modern paradox of the automaton as both an object capable of self-movement (operating *autonomously*) and one lacking in self-direction in his activities (operating *automatically*).[172] Radical mechanism puts mankind on equal footing with the machine as one *acted upon* by external forces and rules—the logical conclusion of a strictly determinist explanation of human conduct, which denies that our decision making is free in any sense from causal determinants; that is, we act *automatically* based on external stimuli and natural laws, and any perception of free will is an illusion. This position is problematic on a number of levels: for example, an individual with no agency can be held responsible neither legally nor morally for his actions, rendering moot the concepts of crime and punishment, sin and salvation. The concept of free will has been hotly debated since the earliest formulation of mechanistic philosophy in the early modern period, although recent literature in the field almost

[172] Mark Seltzer, *Bodies and Machines* (New York: Routledge, 1992), 109. See also Kang, *Sublime Dreams*, 158-174, for a detailed discussion of this development.

without exception supports a form of free will or remains agnostic on the topic[173]—as do I for the purposes of this study. What follows is a brief overview of the debate between mechanism and vitalism as it played out during the period.

As discussed earlier in the chapter, Descartes stepped back from his original mechanistic philosophy (as outlined in his "Treatise on Man") to a more moderate position regarding the soul as the source of agency in humans. This shift gave rise to the "ghost in the machine" of Cartesian dualism that was not only heavily criticized by his contemporaries but has also been a source of contention in the mind-body debate ever since. Descartes' dualistic position, while offered as part of a mechanistic explanation of how the soul operates the body, seems instead to support vitalism in its allowance for a separate class of "material" comprising the soul. This position was dismissed by La Mettrie (and others) as the fallacy of *spiritualism*, a term used by La Mettrie to describe those vitalists who held the dualistic notion that the soul consisted of a different form of matter: "They have spiritualized matter rather than materializing the soul…. Descartes and all the Cartesians… made the same mistake, they admitted two distinct substances in man as if they had seen and counted them"[174] For La Mettrie and many other mechanistic philosophers, this nod toward vitalism seemed a betrayal of mechanistic principles. Others, including some scholars who began their careers as Cartesians, viewed dualism as an opening for a more vitalistic position. Certainly, in the nearly four centuries since Descartes first formulated his position, many scholars since have referred back to his philosophies, either to use his work as a basis for their own arguments or to refute his position.

One of the earliest proponents of universal mechanism in the early modern period, and a harsh critic of Cartesian dualism, was Thomas Hobbes (1588–1679), who in 1651 began his *Leviathan* with a starkly unsentimental description of man, and by extension society ("the Commonwealth") as operating

---

[173] For a concise overview of the ongoing debate regarding free will, see Robert Kane, *The Oxford Handbook of Free Will* (Oxford: Oxford University Press, 2002). See also Kang, *Sublime Dreams*, chapter four; and Mayr, *Authority*, chapter six for more thorough discussions of the various philosophical positions during this time.

[174] Julien Offray de La Mettrie, *Machine Man and Other Writings*, trans. Ann Thompson (Cambridge: Cambridge University Press, 1996), 3.

under purely mechanistic principles—similar to that outlined in Descartes' "Treatise on Man" and "Description of the Human Body," but without Descartes' allowance for "what depends on the soul."[175] La Mettrie similarly held that man was nothing more than a living machine: his formulation in *Machine Man* rejects both Cartesian dualism and the *bête machine* doctrine, describing the soul as "a matter of physical organization"[176] rather than existing as a separate unextended substance; La Mettrie further implies that this was Descartes' own intention, had he been free to publish his true beliefs on the matter.[177] Other mechanists held more complex positions. Denis Diderot (1713–1784), for example, opposed the most extreme form of determinism, but ultimately took a stand against the existence of free will, despite the implications: "Study it closely and you will see that the word liberty is meaningless; there is none and there cannot be a free agent."[178]

On the spiritualist side of the debate were such figures as Immanuel Kant (1724–1804) whose *Critique of Pure Reason* (1781) rejects the *man=machine* metaphor as unconscionable, according to Otto Mayr:

> Kant's preferred metaphor for determinism was generally the automaton. In discussing a proposition with deterministic implications, Kant concluded that in it 'freedom could not be saved. Man would be a marionette or an automaton like Vaucanson's, fabricated and wound up by the Supreme Artist; self-consciousness would indeed make him a thinking automaton, but the consciousness of his spontaneity, if this is held to be freedom, would be a mere illusion.'

---

[175] Descartes, *The World and Other Writings*, 172. See Thomas Hobbes, *Leviathan, or, the Matter, Forme, & Power of a Common-Wealth Ecclesiasticall and Civill* (London: Andrew Crooke, 1651); as well as Hobbes, *The Third Set of Objections to Descartes' Meditations* (1641); and Pierre Simon Laplace, *A Philosophical Essay on Probabilities*, trans. Frederick Wilson Truscott and Frederick Lincoln Emory, (New York: John Wiley & Sons, 1902). See also E. J. Dijksterhuis, *The Mechanization of the World Picture: Pythagoras to Newton*, trans. C. Dikshoorn (Princeton, N.J.: Princeton University Press, 1986).
[176] Roach, *The Player's Passion*, 66.
[177] Ann Thompson, "Introduction" to La Mettrie, *Machine Man*, xi.
[178] Denis Diderot, "Letter to Landois," 29 June 1756, in John Robert Loy, *Diderot's Determined Fatalist* (New York: King's Crown Press, 1950), 134; see also Lester G. Crocker, *Diderot's Chaotic Order: Approach to Synthesis* (Princeton, NJ: Princeton University Press, 1974), 35-44.

> There is no need here to follow the complex argument by which Kant eventually reconciled the causality of physical nature with the principle of human freedom. But it is noteworthy that he rejected the notion of animal automatism in general.[179]

Otto Mayr claims that not only did the majority of philosophers during this time reject animal automatism, "the farther reaching proposition of the man-machine was never taken seriously."[180] Mayr discusses extensively the position held by Robert Boyle (1627–1691), stating that "at once an enthusiastic mechanist and a devout Christian, perhaps he spoke for a majority."[181] Boyle subscribed to the position mentioned earlier that the *man=machine* metaphor was valid to a point, but ultimately failed due to its limitations. This could be viewed as a hybrid position between mechanism and spiritualism. According to Otto Mayr,

> the difference [for Boyle] between the mechanisms of animal bodies and mechanical clocks was not one of degree but of kind; living organisms were "engines of God's own framing." The main distinction was this: humans and, presumably to a lesser degree, animals had a capacity of discernment, judgment, and free choice that distinguished them from automata and that was conventionally called "the soul."[182]

On its surface, Boyle's concept of the soul does not appear appreciably different from that of Aristotle or Descartes: it is comprised of Aristotle's five faculties, to a greater or lesser degree depending on whether an animal or human soul—the human soul naturally being the more complex, as fitting Descartes' *bête machine* doctrine. But according to Mayr, the word *soul* held yet another meaning for Boyle:

> Boyle realized that the word soul, apart from its theological meanings, disguised phenomena that were insufficiently understood.

---

[179] Mayr, *Authority*, 133.
[180] Ibid., 91.
[181] Ibid.
[182] Ibid.

> Soul, then, was a label for the unexplained interactions between body and brain, but it had yet another meaning. Traditionally the concept, soul, had served as symbol and criterion for that complex dynamic condition of the animal mechanism known as *life*.[183]

This position may be seen as an early attempt to articulate what is known in modern philosophy of mind as *emergence*: namely, that the soul is neither Descartes' "fountaineer" residing somewhere inside the body and controlling its function, nor a fiction designed to soften the sting of hard mechanism's lack of free will. Rather, the soul is seen as an *emergent property* of the living body, not contained in any particular part but emerging from the whole and its organizational structure. This hybrid position contains elements of both mechanism and vitalism, and offers the possibility of variation between the souls of humans and animals (different structures result in different emergent properties) as well as between different humans, depending on how well they "rise above" their basic mechanical nature.

According to Minsoo Kang, "The eighteenth-century notion of the mechanical being as a representation of humanity lacking in freedom, self-will, and originality is one of the most powerful and enduring legacies of the late Enlightenment to the modern imagination, something that is very much present today."[184] Here, the image of the automaton as an incomplete human was something to be pitied. However, starting in the late eighteenth century, an increasingly industrialized society saw the rise of machines able to produce more goods, more rapidly, and at a cheaper cost than human labor. In this cultural milieu, the machine came to represent something to be feared as well. As the debate raged on regarding the existence and exact nature of the soul, a new image of the artificial human emerged: an "uncanny entity of mysterious nature that unsettled and frightened people."[185] It is just such an entity, one "that is created as a product of reason and science but turns into the murderous monstrosity

---

[183] Ibid.
[184] Kang, *Sublime Dreams*, 167.
[185] Ibid., 187.

of the Terror,"[186] that Mary Shelley presents us with in her 1818 novel of reanimation and mankind's power over death.

FRANKENSTEIN'S CREATURE AS THE FIRST ARTIFICIAL *PERSON*

Mary Shelley's *Frankenstein: Or, the Modern Prometheus*, is regarded by many as the first work of science fiction due to the manner in which it deals with the period's prevailing scientific theories and discoveries. Inspired by the late eighteenth century exploration into the principle of *galvanism*, Frankenstein's eponymous central character is a fictionalized, slightly mad version of Galvani, for whom the phenomenon is named. Galvani's exciting new scientific discovery was quite the topic of conversation in the intellectual *salons* in which Mary Shelley spent her leisure time, as were stories of Galvani's contemporary Erasmus Darwin (1731–1802), who was said to have explored the possibility of reanimating dead tissue. Shelley herself explicitly attributes to just such a *salon* the genesis of her novel:

> During one of these, various philosophical doctrines were discussed, and among others the nature of the principle of life, and whether there was any probability of its ever being discovered and communicated. They talked of the experiments of Dr. Darwin (I speak not of what the Doctor really did, or said that he did, but, as more to my purpose, of what was then spoken of as having been done by him), who preserved a piece of vermicelli in a glass case, till by some extraordinary means it began to move with voluntary motion. Not thus, after all, would life be given. Perhaps a corpse would be re-animated; galvanism had given token of such things: perhaps the component parts of a creature might be manufactured, brought together, and endued with vital warmth.[187]

Head filled with these notions, Shelley claims she went to bed and suffered a horrible nightmare in which she witnessed a "pale student of unhallowed arts kneeling beside the thing he had put together. I

---

[186] Ibid., 195.

[187] Mary Wollstonecraft Shelley, *Frankenstein; or, the Modern Prometheus*, Revised Classic Books edition (1831). (Reprint Amazon Digital Services, Inc., 2014), Kindle edition, Locations 119-24. Further references to this source will be marked within the text by year and Kindle location numbers.

saw the hideous phantasm of a man stretched out, and then, on the working of some powerful engine, show signs of life, and stir with an uneasy, half vital motion." (1831: 128-129) It is this image, as well as one in which her pale student wakes to find the "horrid thing [standing] at his bedside, opening his curtains, and looking on him with yellow, watery, but speculative eyes," (1831: 134-135) that becomes the kernel around which Shelley builds her story of an artificial human, built upon mechanistic principles but riddled with vitalistic questions regarding the his origins and purpose. The first image, which hails back to Prometheus and creation mythology, underscores the hubris of mankind's attempt to usurp the role of the Creator God. The second image, on the other hand, reflects rather the underlying question prevalent at the time regarding the nature of the human soul.

Frankenstein's Creature, while never given a name in the novel itself, compares himself several times to Adam, the first human created by God. The Creature represents a return to the biological

Figure 23. Frankenstein at work in his laboratory. From Mary Wollstonecraft Shelley, Frankenstein: or, The Modern Prometheus (Boston: Cornhill Publishing Company, 1922)

artificial human, the medieval homunculus, now for the first time explained according to the science of biomechanics rather than alchemy: Frankenstein's early education consisted of his studying the work of Albertus Magnus, Cornelius Agrippa, and Paracelsus, all medieval figures associated with the creation of artificial humans. Once at the university in Ingolstadt, however, Frankenstein is introduced to the "modern sciences" and disparaged for his previous studies of these alchemists by M. Krempe, one of the principal professors of natural philosophy. He eventually reconciles the "old" science with the "new" through contact with his mentor M. Waldman, who encourages Frankenstein to study "every branch of natural philosophy, including mathematics,"[188] while describing the alchemists of the medieval era as men to whose indefatigable zeal modern philosophers were indebted for most of the foundations of their knowledge." (1818: 433) Thus is the stage set for Frankenstein's own discovery, which synthesizes

Figure 24. Frankenstein attempts to bring his creation to life. Here, in the 1994 version directed by and starring Kenneth Branagh (shown), Frankenstein uses several forms of electricity in his process: the scene shows images of flywheels, voltaic pile batteries, acid batteries, and electric eels in a viscous liquid, which Frankenstein releases into the copper chamber holding his creation. Still from Mary Shelley's Frankenstein *(1994).*

[188] Mary Shelley, *Frankenstein, Or, the Modern Prometheus*, 3 vols (London: Lackington, Hughes, Harding, Mavor, & Jones, 1818; Reprint e-artnow, 2013), Kindle edition, Location 442. Further references to this text will be indicated within the text by year and Kindle location number.

"medieval alchemy, in which the creation of an artificial human was the mark of a great alchemist, with the combined knowledge of the Scientific Revolution, which values discovery and experimentation.

While it is never expressly stated, it is strongly implied that Frankenstein brings his Creature to life through the use of electricity: as Frankenstein obliquely comments, "I collected the instruments of life around me, that I might infuse *a spark of being* into the lifeless thing that lay at my feet" (1818: 548-549).[189] This *spark of being* is reminiscent of the phenomenon described by vitalists as "animal electricity," the subject of much debate an experimentation at the time, as alluded to in the Preface: writing as "Marlow," Shelley claims that "the event on which this fiction is founded has been supposed, by Dr. Darwin, and some of the physiological writers of Germany, as not of impossible occurrence" (1818: 8-9). The author here refers to Darwin's alleged hypothesis regarding the reanimation of dead tissue through the principle of *galvanism*, the phenomenon in which muscles contract upon stimulation by an electric current—as Galvani and others discovered through their experimentation with frogs. Such speculation can be seen as an extension of the position outlined in La Mettrie's *Machine Man*, for if man is truly no more than a living machine (powered by "animal electricity"), then in theory it should be possible to re-start a non-working machine after simply replacing the defective parts. Frankenstein reveals that after much detailed study of anatomy and "the natural decay and corruption of the human body" (1818: 472) he is struck by the surprisingly simple solution—which he naturally does not share with us, given its disastrous result. But his aspirations are Promethean:

> A new species would bless me as its creator and source, many happy and excellent natures would owe their being to me. No father could claim the gratitude of his child so completely as I should deserve theirs. Pursuing these reflections, I thought, that if I could bestow animation upon lifeless matter, I might in process of time (although I now

[189] Italics mine.

> found it impossible) renew life where death had apparently devoted the body to corruption. (1818: 508-511)

Imagining himself in the role of the Creator God, Frankenstein sets about collecting the materials necessary for his work—the best "parts" from the various man-machines buried in the local cemetery—in order to create this new life.

In a rather disastrous turn of events, Frankenstein is horrified by his creation, and abandons the Creature, failing even to give the being a name. The description of this would-be Creator God running in horror from his Adam (see Figure 25), while intended to invoke pity for Frankenstein, instead does the opposite:

> I beheld the wretch—the miserable monster whom I had created. He held up the curtain of the bed; and his eyes, if eyes they may be called, were fixed on me. His jaws opened, and he muttered some inarticulate sounds, while a grin wrinkled his cheeks. He might have spoken, but I did not hear; one hand was stretched out, seemingly to detain me, but I escaped, and rushed down stairs. I took refuge in the court-yard belonging to the house which I inhabited; where I remained during the rest of the night, walking up and down in the greatest agitation, listening attentively, catching and fearing each sound as if it were to announce the approach of the demoniacal corpse to which I had so miserably given life. (1818: 565-570)

It is as if the God of Michelangelo, rather than reaching out to Adam on the ceiling of the Sistine Chapel, recoiled in horror at the life he had created. Left for dead, the Creature nonetheless survives.

It is at this point that most stage and film adaptations of Frankenstein diverge from the original text.[190] Influenced by Boris Karloff's portrayal in the early films *Frankenstein* (1931), *Bride of*

---

[190] A comprehensive analysis of the various film and stage adaptations of *Frankenstein* is outside of the scope of this dissertation; I am speaking here in generalities regarding the two archetypical, and diametrically opposed, depictions of Frankenstein's creature.

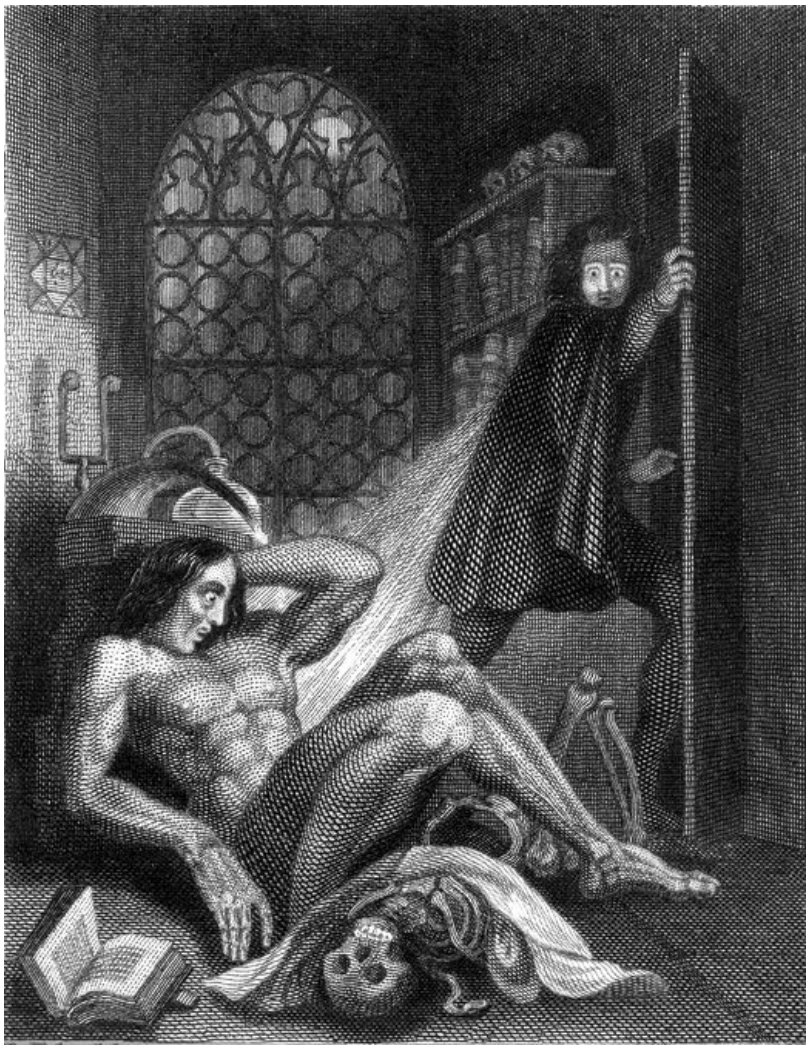


Figure 25. A horrified Frankenstein flees before the spectre of his creation. Frontispiece from the 1831 (revised) edition, Mary Wollstonecraft Shelley, Frankenstein: or, The Modern Prometheus *(London: Colburn and Bentley, 1831).*

Frankenstein (1935), and *Son of Frankenstein* (1939), most versions of the story generally depict the Creature as a hulking brute, violent and unable to communicate, capable only of inflicting pain in his animalistic rampage. Popular culture is filled with this iconic image of a monster with protruding forehead and bolts in his neck, typically with green/greenish skin tone, clumsy but incredibly strong. This is the biological version of Kang's Enlightenment automaton: an incomplete human, a "person with something essential missing in him that causes him to be stupid, bumbling, callous, [or] cruel."[191] The

[191] Kang, *Sublime Dreams*, 153.

Creature of Shelley's imagination, however, is of an altogether different sort. Intelligent and articulate, this Creature was able not only survive after his abandonment, but had wisdom enough to secrete himself in a peasant family's pigsty and learn through observation how to speak, how to read, and how to interact with others in a socially acceptable way.[192] Rather than being hounded and destroyed by angry villagers, the Creature of the novel ultimately weeps over the death of his creator, filled with remorse at the result of his campaign of revenge against Frankenstein, as our narrator describes:

> [The Creature] paused, looking on me with wonder; and, again turning towards the lifeless form of his creator, he seemed to forget my presence, and every feature and gesture seemed instigated by the wildest rage of some uncontrollable passion. "That is

Figure 26. Promotional photo of Boris Karloff from The Bride of Frankenstein *(1935) as Frankenstein's monster. This is the iconic image of Frankenstein's monster, a dumb brute with no higher mental functions, able only to wreak destruction. Photo is public domain, via Wikimedia Commons.*

[192] See the creature's account of this learning process in Shelley, *Frankenstein* (1818), Volume II, chapters three and four.

> also my victim!" he exclaimed; "in his murder my crimes are consummated; the miserable series of my being is wound to its close! Oh, Frankenstein! generous and self-devoted being! what does it avail that I now ask thee to pardon me? I, who irretrievably destroyed thee by destroying all thou lovedst. Alas! he is cold; he may not answer me." (1818: 2857-2861)

Far from an automaton—a being stupid, bumbling, callous, or cruel—this description better fits the traditional image of the sentimental protagonist, a "passionate and sensitive nonconformist constantly struggling with the powerful emotions that overtake him, compelling him to commit actions of a grand but sometimes destructive nature…. a genius whose brilliance is fueled by his soul's undying fire."[193] This Creature desires nothing more than to be a part of the society that shuns him. Filled with the twin passions of love and rage, the Creature proclaims that if he cannot have the former, he must then indulge the latter, though it pains him to do so. The sympathy of a single soul would be enough to satisfy him: "If any being felt emotions of benevolence towards me, I should return them an hundred and an hundred fold; for that one creature's sake, I would make peace with the whole kind!" (1818: 1810-1812). But his request of Frankenstein to create another of his kind as a companion, an Eve for his Adam, ultimately remains unfulfilled. Faced with the death of his creator and the enormity of his actions, the Creature seeks redemption through death at his own hand:

> "But soon," he cried, with sad and solemn enthusiasm, "I shall die, and what I now feel be no longer felt. Soon these burning miseries will be extinct. I shall ascend my funeral pile triumphantly, and exult in the agony of the torturing flames. The light of that conflagration will fade away; my ashes will be swept into the sea by the winds. My spirit will sleep in peace; or if it thinks, it will not surely think thus. Farewell." (1818: 2920-2923)

---

[193] Kang, *Sublime Dreams*, 153-154.

No longer able to face judgment at the hands of his Creator, the Creature judges himself according to human moral standards, finds himself guilty, and accepts his fate. The language is steeped in Christian sentiment: "my spirit will sleep in peace," he claims. But what is the nature of the spirit the Creature possesses?

In the novel, both Frankenstein and the Creature speak definitively in terms of the Creature *possessing a soul*—although whether this should be considered an Aristotelian soul (as in, that which separates the living from the non-living) or the Cartesian soul (the "fountaineer" that is the exclusive domain of the human) remains unclear—Shelley uses the term uncritically. Similarly, in the 2011 adaptation by the National Theatre, directed by Danny Boyle, Victor Frankenstein speaks of his creation as human, albeit beastly, proclaiming to Elizabeth that he built "a functioning man—a brute of a man!"[194]—positioning himself as the *deus faber* of a new species: as Victor exclaims to his murdered brother Will in a dream sequence, "I breathe the breath of God!"[195] Other adaptations, however, have chosen to focus specifically on whether or not Frankenstein's creation is indeed possessed of a soul: for example, in the 1994 film version, directed by Kenneth Branagh (who also portrays the eponymous doctor), the question of his soul is of particular importance to the Creature (portrayed by Robert DeNiro). Two particular exchanges toward the end of the film address the Creature's soul: the first in vitalistic terms as a distinct entity; the second, in the more contemporary language of emergent properties. In the first exchange, the Creature confronts Frankenstein regarding the circumstances of creation and its tragic consequences:

THE CREATURE: You gave me these emotions, but you didn't tell me how to use them.

Now two people are dead because of us. Why?

---

[194] Nick Dear, *Frankenstein, Based on the Novel by Mary Shelley*, (London: Faber & Faber, 2011), Kindle edition, Locations 1119-20. It is interesting to note here a related identity issue created by the National Theatre adaptation: during the performance run, actors Benedict Cumberbatch and Jonny Lee Miller alternated the roles of Frankenstein and the Creature, and in the NT Live Cinema broadcast for the NT's fiftieth anniversary, the two performances were intercut and shown to telepresent audiences as a single dramatic event.

[195] Ibid., Location 957.

VICTOR FRANKENSTEIN: There was something at work in my soul which I do not understand.

THE CREATURE: And what of my soul? Do I have one? Or was that a part you left out?[196]

Frankenstein is unable to answer. Moments later, the Creature again presses Frankenstein regarding his genesis, asking about the nature of the individuals who contributed his various body parts:

THE CREATURE: What kind of people is it in which I am comprised? Good people? Bad people?

VICTOR FRANKENSTEIN: Materials. Nothing more.

THE CREATURE: You're wrong. [*Picks up recorder*] Did you know I knew how to play this? From which part of me did this knowledge reside? From this mind? From these hands? From this heart? And reading and speaking. Not so much things learned as things remembered.

VICTOR FRANKENSTEIN: Slight trace waves in the brain perhaps.

THE CREATURE: Did you ever consider the consequences of your actions? You made me, and you left me to die. Who am I?

VICTOR FRANKENSTEIN: You? I don't know.

THE CREATURE: And *you* think that *I* am evil.[197]

Frankenstein describes his unholy creation at various times in both the novel and in this film version as evil, as does M. Krempe (actor Robert Hardy) as a voiceover in the film: "You fool, Victor Frankenstein of Geneva, how could you know what you had unleashed? How was it pieced together? Bits of thieves? Bits of murderers? Evil stitched to evil stitched to evil."[198] However, it is obvious from the context and

---

[196] *Mary Shelley's Frankenstein*, directed by Kenneth Branagh, performed by Kenneth Branagh, Robert De Niro, Helena Bonham Carter (1994; Culver City, CA: TriStar Pictures, 1995), iTunes MP4.
[197] Ibid.
[198] Ibid.

Figure 27. The Creature questions his creator about his origins, revealing Frankenstein's own notebook as the source of his learning. Movie still from Mary Shelly's Frankenstein *(1994).*

from the Creature's response here that the viewer should consider Frankenstein the more evil of the two. Both these scenes reflect a Cartesian self-knowledge; in questioning the nature of his existence, the Creature exhibits clear evidence that he *thinks*, therefore he *is*.

Frankenstein's Creature is the ideal artificial human to represent its particular moment in time: borne of the ambitions of medieval alchemy, built through the scientific rigor of the Renaissance, and possessing the sentimentality of the Enlightenment, his story is told through the framework of Romanticism as a fantastic story of human hubris and inhuman redemption. But how can we categorize this artificial being? The Creature is neither golem, created of clay and brought to life through divine magic, nor homunculus, grown from human sperm into a miniature humanoid, both the subject of medieval accounts of artificial humans. Neither is this Creature an automaton in the literal sense, as it is comprised of bones and organs rather than clockwork. But can he be considered an automaton as the term was used during the Enlightenment? I would argue no, for as Minsoo Kang defines this use,

> In an intellectual culture dominated by the ideas of vitality, sensibility, and sentimentality, the machine-man idea did not disappear but was made much fruitful use of as an image of an incomplete human being (in the way the mechanist model of the body came to be seen as an incomplete picture)—a person with something essential missing in him that causes him to be stupid, bumbling, callous, cruel, vapid, or just plain tedious.[199]

The "something missing" here is precisely that characteristic the automaton is criticized for lacking: no *vital force*, no *sentiment*, no *free will*. Yet in the figure of Frankenstein's creation is a self-admittedly artificial human who fulfills all three of the vitalistic criteria outlined above. Far from stupid or bumbling, the Creature is highly intelligent and possesses superhuman strength and agility. Formed of reconstructed human remains, and brought to life through the "animal force" of electricity, the Creature is not an undead zombie but rather a truly living Creature, newly born, who experiences on an accelerated scale the learning processes of childhood and youth. The ability of Frankenstein's creation to question his ontological status marks him as one of a new breed of artificial human. No longer homunculus, automaton, or golem, Frankenstein's monster is more properly classified as the first of a new category—the *Homo artificialis* —the artificial human who questions his ontological status as Other eloquently and compellingly. In the next chapter I will tease out exactly what it means to be a *Homo artificialis* and provide a taxonomy of how these beings are perceived in contemporary society, both in the real world and the worlds of our imagination.

---

[199] Kang, *Sublime Dreams*, 153.

CHAPTER THREE: THE PATH TO PERSONHOOD

And this proves not only that the brutes have less reason than man, but that they have none at all: for we see that very little is required to enable a person to speak; and since a certain inequality of capacity is observable among animals of the same species, as well as among men, and since some are more capable of being instructed than others, it is incredible that the most perfect ape or parrot of its species, should not in this be equal to the most stupid infant of its kind or at least to one that was crack-brained, unless the soul of brutes were of a nature wholly different from ours.

Rene Descartes, Discourse on the Method[1]

What makes a person? On December 2, 2013, attorneys for the Nonhuman Rights Project (NhRP) filed a landmark lawsuit in the New York Supreme Court on behalf of their client Tommy, a chimpanzee held in captivity in Gloversville, NY, in an effort to have the state "recognize Tommy as a legal person, with a right to liberty."[2] Though the suit failed, it reflects a changing attitude toward the ontological crisis brought about by the mechanistic philosophy of the early modern period, offering a glimpse into a broader, more inclusive definition of "person" as determined not by biology but by behavior—a definition that rejects Descartes' *Bête Machine* Doctrine as outlined in chapter two. In a 2012 Dyson Lecture at Pace University, Steven M. Wise, law professor and president of the NhRP, outlined the basic principles he claims entitle a nonhuman entity to the legal status of "person" and thus the same protection under the law as is afforded all humans in regards to fundamental rights such as bodily liberty and bodily integrity. Wise began with a simple question: "What is a *sufficient* condition for having fundamental rights? Not a necessary condition, but a *sufficient* condition?"[3] In formulating an

---

[1] René Descartes, *Discourse on the Method and the Meditations*, trans. by John Veitch (New York: Cosimo, Incorporated, 2008).
[2] As of this writing, this and similar suits filed by the NhRP are working their way through the appeals process. According to Steven M. Wise, president of the NhRP, the goal of the Project is to grant basic legal protections to certain nonhuman animals through recognition as "legal persons" in the eyes of the law—a status already granted to other nonhuman entities such as corporations. See http://www.nonhumanrightsproject.org for more information. James Gorman, "Rights Group Is Seeking Status of 'Legal Person' for Captive Chimpanzee," *The New York Times*, December 3 2013, accessed 21 March 2014, http://www.nytimes.com/2013/12/03/science/rights-group-sues-to-have-chimp-recognized-as-legal-person.html?pagewanted=all&_r=0.
[3] Steven M. Wise, "Nonhuman Rights to Personhood," *Pace Environmental Law Review* 30, no. 3 (2013): 5.

answer to this question, Wise claims, he developed the concept of *practical autonomy*: "Practical autonomy has three elements. First, one must be cognitively complex enough to want something. Second, one must be able to act intentionally to achieve one's desires. Third, one must have a sense of self complex enough so that it matters to whether one achieves one's own goals."[4] The NhRP began their legal tests with chimpanzees because, according to Wise, "from an evolutionary point of view, the closer the common ancestry is between any two beings, the more likely it is that their similar behaviors have similar mental causes."[5] Wise's definition of *practical autonomy* is based on decades of scientific research into nonhuman cognition, and while his approach begs a number of key questions that are central to this chapter (such as how to determine whether an entity has desires or a sense of self), it reflects a more nuanced understanding of "reason" than Descartes' definition. In fact, in the nearly four centuries since Descartes first formulated his "proof" confirming the "unique and superior nature of man over beast and machine," numerous figures, including scientists, philosophers, politicians, lawyers, and religious leaders, have debated the validity of both Descartes' conclusions and the criteria upon which they are based.

As evidenced by the NhRP legal campaign, this debate has begun to have implications outside the realms of theoretical discussion and to become a matter with practical applications. But the criteria outlined by Wise and others need not be applied to nonhuman animals alone. As I will illustrate in the later case studies, numerous fictional accounts of artificial humans, from *Frankenstein* onward, have also addressed the ontological crisis of "personhood," offering examples of beings that problematize not only humankind's unique and superior status but the very necessity of biology to the definition at all. Our science fiction is rapidly becoming scientific reality. As our technology catches up with our imaginations and aspirations, the actual creation of an artificial human that meets this standard of

---

[4] Ibid., 6. I will return to Wise's arguments later in this chapter, including his "scale of practical autonomy" as outlined in *Drawing the Line: Science and the Case for Animal Rights* (Cambridge, MA: Perseus Books, 2002). While my study concerns artificial nonhuman beings, not animals, I feel that the arguments used in many of these cases (such as the NhRP lawsuit) could theoretically be applied to *Homo artificialis* as well.

[5] Wise, "Nonhuman Rights," 6.

practical autonomy has moved out of the realm of fantasy and into that of theoretical possibility. The symbiotic relationship between mythology and technology we have explored throughout the last two chapters has only accelerated over the course of the twentieth century, which is as heavily riddled with iconic images of *Homo artificialis* in the world of technology as it is in the world of our imagination. Rising from the mechanization of the worldview of the early modern period, and brought into sharp focus during the Industrial Revolution of the nineteenth century, the similarities between the natural and artificial human have led to inevitable clashes in ideology and identity politics which seem to have no viable resolution in sight.

In this chapter, I trace the history of the artificial human over the last century, paying particular attention to the problematic relationship between *Homo sapiens* and *Homo artificialis* in both our reality and our imaginations in terms of the various "tests" devised to distinguish between the two. Beginning with the early twentieth century shift from the automaton to the robot as the paradigmatic image of the artificial human, I will chart the relationship between the treatment this paradigm receives in the century's fiction and the technological advances that bring *Homo artificialis* closer to reality. As these advances in technology inexorably shrink the realm of "what machines can't do," our concept of what it means to be *human* likewise evolves. How this definition evolves holds great repercussions for upcoming generations of *Homo sapiens* and *Homo artificialis* alike, as we learn as a society to navigate these fluid meanings. To that end, I also offer a new definition of my own that, not unlike the efforts of the NhRP to redefine the concept of *person*, situates the *human* identity as a performative construct apart from the biological substrate in which it is realized.

THE FIRST ROBOT

While Victor Frankenstein's creation in Shelley's 1818 novel represents the birth of the *Homo artificialis*—a new breed of artificial human who questions his Otherness—the kind of being Frankenstein builds still has strong roots in medieval magic and Renaissance craftsmanship. Like the

automata of the eighteenth century, Frankenstein's Creature is a singular work of fine craftsmanship, created by a skilled, though misguided, master, whose despair at his failure to create a *beautiful* Creature overshadows the genius involved in *actually creating* a living being *at all*. His unique nature is the source of the Creature's greatest anguish: had he but one other of his kind with whom to share his fate, claims the Creature, he would happily exile himself, never again to impose his presence upon his creator or society in general. Frankenstein's refusal to provide the Creature with a mate brings about the tragedy witnessed in the novel. As the only one of his kind, Frankenstein's Creature has only limited impact, affecting mainly Frankenstein and his family, plus a few other folk between Ingolstadt and Geneva, and ends his rampage upon the death of his creator.

The fiction of the early twentieth century, however, gave rise to a new type of artificial human: the *robot*. Born out of the factories of the Industrial Revolution, this mass-produced artificial human posed a threat on a global scale. The robot was introduced to the world by Karel Čapek in 1921 through his play *R.U.R.* (*Rossum's Universal Robots*), which clearly articulates the existential crisis of the period, epitomizing "the traumatic transformation of modern society by the First World War and the Fordist assembly line."[6] Derived from the Czech word *robota*, meaning compulsory labor, Čapek's Robots[7] are envisioned as an enslaved underclass, manufactured to free humans from work and fill factories with cheap labor, thus reducing the costs of consumer products and making the owners of the firm R.U.R. Ltd. quite rich in the process. While automata were individually crafted representatives of mechanistic philosophy, epitomizing the intellectual philosophy of their time, Čapek's mass-produced Robots exemplify the anxiety of "The Machine Age," a period during which society struggled in the midst of a love-hate relationship with the machine. According to Harold Segel,

---

[6] John Rieder, "Karl Čapek," in *Fifty Key Figures in Science Fiction*, edited by Mark Bould (London: Routledge, 2010), 49.
[7] Although the word *robot* has entered our common vocabulary in the decades since the play's premiere, most translations of *R.U.R.* tend to capitalize the word Robot, as does most scholarship surrounding the play. Therefore, when referring to the characters in the play, I shall follow traditional scholarship in the use of the capital.

> The Futurists extolled the machine as the foundation of a postbourgeois Europe. The Expressionists viewed the new technological era with profound anxiety, fearing the domination of humans by machines…. *R.U.R.* embraces the common Expressionist vision of humans so overwhelmed by machine culture that they become subservient to and are ultimately swept away by the very robots they have created to enlighten their own burden.[8]

As a "landmark play" of the "'golden age' of science fiction,"[9] *R.U.R.* helped to insure that "the robot or android would claim a central place in the developing science fiction genre"[10] and inspired a host of other stories run through with "the sort of double-vision that attends the image of human artifice—and that points up our culture's deeper, unsettled view of the technological."[11]

The influence of *R.U.R.* on the Western worldview can hardly be overestimated. The play was quite famous in its own time, having been translated into thirty languages by 1923, and productions stirred a Robot fever across the globe. According to Kara Reilly, one London reviewer "began to see everyday people in the streets as Robots. Another reviewer postulated that perhaps Čapek himself was a Robot."[12] In the near-century since its debut, *R.U.R.* has become an iconic part of popular culture: the play has been produced innumerable times throughout the world, in original form, in translation, and in adaptation (including recently as a rock-musical entitled *Save the Robots*[13]); there have also been numerous adaptations for radio, television, and film, as well as homages, tributes, and references in both science fiction and other genres. As Čapek's Robots have played such a formidable role in the

---

[8] Harold B. Segel, *Pinocchio's Progeny: Puppets, Marionettes, Automatons and Robots in Modernist and Avant-Garde Drama*, PAJ Books (Baltimore: Johns Hopkins University Press, 1995), 311.
[9] J. P. Telotte, *Replications: A Robotic History of the Science Fiction Film* (Urbana: University of Illinois Press, 1995), 40.
[10] Ibid.
[11] Ibid.
[12] Kara Reilly, *Automata and Mimesis on the Stage of Theatre History* (New York: Palgrave Macmillan, 2011), 152.
[13] Christopher Arnott, "Save the Robots: A Westport playwright/producer rocks with metal people," *Orlando Sentinel,* 7 May 2012, accessed 6 June 2014, http://www.orlandosentinel.com/topic/wtxx-save-the-robots-a-westport-playwrightproducer-rocks-with-metal-people-20120507,0,4974811.story.

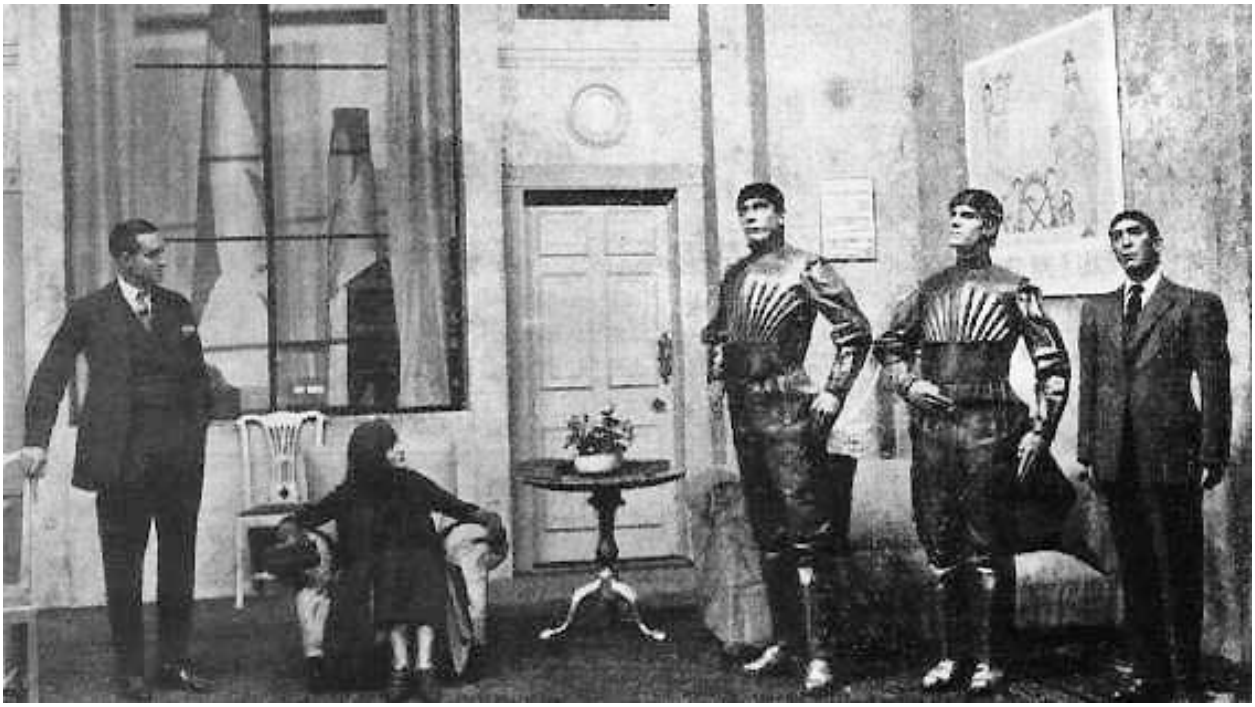


Figure 28. Scene from the 1923 London production of Rossum's Universal Robots*. From Wikimedia Commons.*

shaping of our contemporary understanding of the artificial human, a detailed examination of their nature is in order.

While our modern image of the robot is that of a mechanical artificial human similar to an automaton, Čapek's original Robots were biological in nature, closer to the medieval homunculi or Frankenstein's Creature, born out of science rather than magic. As General Manager Harry Domin explains in the play's opening scene,

> It was in the year 1920 that old Rossum, the great physiologist, who was then quite a young scientist, took himself to this distant island for the purpose of studying the ocean fauna, full stop. On this occasion he attempted by chemical synthesis to imitate the living matter known as protoplasm until he suddenly discovered a substance which

> behaved exactly like living matter although its chemical composition was different. That was in the year of 1932.[14]

This protoplasm-like substance formed a paste which, unlike natural tissue, "didn't mind being sewn or mixed together"(4). Domin describes to a horrified Helena Glory the island's factory, from the preparation of the synthesized paste to the formation of various organs:

> HELENA: Kneading trough?
>
> DOMIN: The pestle for beating up the paste. In each one we mix the ingredients for a thousand Robots at one operation. Then there are the vats for the preparation of liver, brains, and so on. Then you will see the bone factory. After that I'll show you the spinning mill.
>
> HELENA: Spinning mill?
>
> DOMIN: Yes. For weaving nerves and veins. Miles and miles of digestive tubes pass through it at a time. (10)

These interchangeable body parts are assembled ("stitched together") in various configurations to create several grades of Robot, from the Manual Labor Robot "as powerful as a small tractor [and] guaranteed to have average intelligence" (7) to the highest grade, who act as "officials" (accountants, presumably, and other mid-level business professionals) and can last up to twenty years. These assembly-line products are then stored in a warehouse to be shipped out to clients worldwide.

The production method Domin describes differs from that of earlier biological artificial humans in one very important way: while a medieval homunculus is grown individually from male sperm, and Frankenstein crafted his singular Creature from the already-extant parts of various corpses, the entire

---

[14] Karel Čapek, *R.U.R. (Rossum's Universal Robots)*, trans. Paul Selver and Nigel Playfair, Dover Thrift Editions (Mineola, NY: Dover Publications, 2001), Kindle edition, 3-4. While a new, more complete translation of *R.U.R.* was published in the volume *Toward the Radical Center: A Karel ČApek Reader* (Highland Park, NJ: Catbird Press, 1990), I have chosen to reference the 1923 Selver translation as an artifact of the document that had such an influence on the later English-language literature I reference in this chapter. The two versions do not differ appreciably in their treatment of the themes I address in this chapter; instances in which the 1990 translation offers a marked difference from the 1923 are addressed in the footnotes. Future references to the text through inline citations are from the 1923 Selver edition.

process of creating a Robot is based upon the assembly line of the Industrial Revolution. Kara Reilly cites this change as the end of an era:

> It is as if the transfer of agency from human to machine occurs on a grand scale, and the more mechanical society becomes, the more machine-like people become. With the advent of Robots, automata become a thing of the past. Automata are unique, hand-created entertainers, whereas Robots are mass-produced workers…. [This change] marks the transition from automata to automation or that cultural moment that signifies the birth of the Robot and the death of the automaton.[15]

The change from individual to mass production of the artificial human reflects the widespread automation of Industrial Age society in general, as individual genius and creativity are devalued in favor of productivity, efficiency and utility.

Rossum's Universal Robots are imagined as the industrialized "ideal workforce," stripped of all nonessential features in order to maximize their efficiency. Man is a terribly inefficient machine: he has desires and preferences, tires easily, and is beset by emotions which interfere with his productivity. He must eat and sleep on occasion, and so needs a home in which to live when he is not working. And above all he expects to be paid a living wage for his work in order to buy those things he desires. Robots, on the other hand, are not distracted by such "ornaments":

> DOMIN: … a working machine must not play the piano, must not feel happy, must not do a whole lot of other things. A gasoline motor must not have tassels or ornaments, Miss Glory. And to manufacture artificial workers is the same thing as to manufacture gasoline motors. The process must be of the simplest, and the product of the best from a practical point of view. What sort of worker do you think is the best from a practical point of view?

[15] Reilly, *Automata and Mimesis,* 149-50.

HELENA: What?

DOMIN: What sort of worker do you think is the best from a practical point of view?

HELENA: Perhaps the one who is most honest and hardworking.

DOMIN: No; the one that is the cheapest. The one whose requirements are the smallest.

> Young Rossum invented a worker with the minimum amount of requirements. He had to simplify him. He rejected everything that did not contribute directly to the progress of work— everything that makes man more expensive. In fact, he rejected man and made the Robot. My dear Miss Glory, the Robots are not people. Mechanically they are more perfect than we are, they have an enormously developed intelligence, but they have no soul. (6)

According to Domin, all these human characteristics drive up the cost of production for most consumer goods. Once human laborers are replaced by Robots—which due to Robotic efficiency also means reducing the total workforce by roughly sixty percent—the cost of consumer goods plummets to less than a third of what they cost using human labor. It also has a snowball effect: after so many years of Robot-based production, factories wishing to compete in the open market "will go pop like chestnuts if they don't at once buy Robots to lower the cost of production"(15). In much the same way as the cotton mill (actually, the culmination of a number of inventions over the eighteenth and nineteenth centuries) had displaced workers in the textile industries of the Industrial Revolution, the Robots have displaced factory workers in the twentieth century world of the play.

In outlining how Rossum's Universal Robots are *not* human, Domin and the other executives of R.U.R. offer a definition of what *is* human that comprises a number of criteria offered by history. First, humans possess the *desiderative* faculty as outlined by Aristotle: they have desires and interests, and act upon these in a display of what some would call free will. As Domin outlines in the passage above, humans play the piano—a talent he deems superfluous to the function of the average factory worker.

Further, when Helena suggests that the Robots should receive a salary in order that "They would buy—what they need … what pleases them"(13), Hallemeier replies that "there's nothing that does please the Robots"(13)—they have no interests, no appetite, and no will of their own. In other words, having no desires, Robots have no soul. Humans further possess the *sentimentality* so prized by the Enlightenment, an attribute that is also lacking in the Robot. Hallemeier goes on in the same passage to explain to Helena that robots have no concept of emotions: they do not love, do not feel defiance, or even pain—which has become a problem for the firm, as Robots are constantly injuring themselves for lack of this protective feedback mechanism. Robots also do not fear death, as demonstrated by Robots Sulla and Marius in the following passage:

> HELENA: … Sulla, you wouldn't let yourself be cut to pieces?
>
> SULLA: Yes.
>
> HELENA: Oh, you're not afraid of death, then?
>
> SULLA: I cannot tell, Miss Glory.
>
> HELENA: Do you know what would happen to you in there?
>
> SULLA: Yes, I should cease to move.
>
> HELENA: How dreadful!
>
> DOMIN: Marius, tell Miss Glory what you are.
>
> MARIUS: Marius, the Robot.
>
> DOMIN: Would you take Sulla into the dissecting room?
>
> MARIUS: Yes.
>
> DOMIN: Would you be sorry for her?
>
> MARIUS: I cannot tell.
>
> DOMIN: What would happen to her?
>
> MARIUS: She would cease to move. They would put her into the stamping-mill.

DOMIN: That is death, Marius. Aren't you afraid of death?

MARIUS: No.

DOMIN: You see, Miss Glory, the Robots have no interest in life. They have no enjoyments. They are less than so much grass. (9)

When instructed to report to the dissecting room to be taken apart, Sulla shows no reticence; neither does Marius, who escorts her there. Both Robots comprehend on an intellectual level what is to happen to Sulla, yet exhibit none of the instinct for self-preservation that is inherent in all living creatures. The final "human" faculty lacking in Robots is one which is again common to all living creatures: reproduction. Robots are designed without reproductive organs—another "expensive" extra superfluous to factory work, not to mention detrimental to the company's bottom line. Robots have a limited life span; according to Domin, the longest-lasting models must be replaced every twenty years. Were Robots able to reproduce, R.U.R. would have little repeat business, as clients could simply "grow their own" after their first shipment. Like death, sexuality is an abstract concept to the Robots. Indeed, the only reason gendered Robots are constructed at all is to adhere to societal expectations, as is the case with the female Robot Sulla, who works as a secretary. So while these artificial beings are unquestionably biological, they are considered merely "living machines" by the humans who created them.

As it turns out, the fact that the Robots are lacking in these "human" traits is the only thing that has maintained the status quo in the decades since R.U.R started manufacturing them.[16] As mentioned above, Robots feel no pain, which results in frequent damage to the Robots as there is no feedback mechanism in place to prevent it. Therefore, Dr. Gall informs Helena in Act I that he is working on "pain-nerves" to supplement the Robots' limited nervous system, making them "more perfect from a technical

[16] Although the events of the play take place in an unspecific future time relative to the original production date, historian Kara Reilly references a poster from the 1921 Prague production listing the year as 2000—a date not mentioned in any version of the script (*Automata and Mimesis*, 148). Barring other evidence to the contrary, the play could conceivably be set anywhere from thirty to perhaps eighty years after Old Rossum's initial 1932 discovery.

point of view"(14), even though this "suffering" will inevitably reduce their quality of life (such as it is). Unfortunately, Dr. Gall's tinkering does not stop here: in Act III, ten years later, he reveals that he has been further manipulating the Robot fabrication process, to disastrous effect:

> DR. GALL: No! I'm to blame for this, for everything that's happened.
>
> FABRY: You, Gall?
>
> DR. GALL: I changed the Robots.
>
> BUSMAN: What's that?
>
> DR. GALL: I changed the character of the Robots. I changed the way of making them. Just a few details about their bodies. Chiefly—chiefly, their—their irritability.
>
> HALLEMEIER: Damn it, why?
>
> BUSMAN: What did you do it for?
>
> FABRY: Why didn't you say anything ?
>
> DR. GALL: I did it in secret. I was transforming them into human beings. In certain respects they're already above us. They're stronger than we are.
>
> FABRY: And what's that got to do with the revolt of the Robots?
>
> DR. GALL: Everything, in my opinion. They've ceased to be machines. They're already aware of their superiority, and they hate us. They hate all that is human. (40)

Dr. Gall's use of the term *irritability* for the quality he has modified hearkens back to the vitalist terminology of the early modern period, yet re-frames the term to include not only *sensation* (necessary to the biological functioning and feedback of living machines such as Robots are) but also the *sensibility* of the Enlightenment. At the behest of Helena, Dr. Gall has imbued a small number of Robots with this "inherent bodily capacity, differing markedly between individuals, that registers and communicates

feeling"[17]—not a soul, per se, although this was Helena's original request, but a "physiological correlate"(41) that was intended to make them more human.

As it turns out, making the Robots more human proves a fatal mistake, for as Domin points out, "Nobody can hate man more than man"(42). In its production of Robots, R.U.R. has created what Minsoo Kang terms a "mechanical subaltern,"[18] an automated servant class designed to replace an already marginalized human worker. Domin's goals were lofty: his intention was to create nothing less than a Paradise on Earth for all mankind:

> DOMIN: There will be no poverty. All work will be done by living machines. Everybody will be free from worry and liberated from the degradation of labor. Everybody will live only to perfect himself.... The servitude of man to man and the enslavement of man to matter will cease…. *The Robots will wash the feet of the beggar and prepare a bed for him in his house*…. Man shall be free and supreme; he shall have no other aim, no other labor, no other care than to perfect himself. He shall serve neither matter nor man. He will not be a machine and a device for production. He will be Lord of creation. (15-16)[19]

This desire for an ideal world is a common theme of utopian literature, and as is also common, such a society can only be built through the labor of a subaltern class that is *not* part of the society; in Domin's speech above, Robots have become this subaltern. As Kang points out, this impulse has been tied up in the concept of the automaton since ancient times: "The stories of Daedalus's moving statues, for instance, gave Aristotle visions of the self-weaving shuttle and the self-playing lyre in a world in which 'chief workmen would not want servants, nor masters slaves.'"[20]

---

[17] Joseph R. Roach, *The Player's Passion: Studies in the Science of Acting* (Ann Arbor: University of Michigan Press, 1993), 95.
[18] This mechanical subaltern is a common theme in science fiction. Minsoo Kang, *Sublime Dreams of Living Machines: The Automaton in the European Imagination* (Cambridge, MA: Harvard University Press, 2011), 264ff.
[19] Čapek, *R.U.R,* 15-16, emphasis mine.
[20] Kang, *Sublime Dreams*, 266.

Throughout history this subaltern has been of necessity human—variously termed slaves, servants, or workers—but treated as something less than human in order to justify their social status, transforming them into "useful but unnoticed things existing only to serve."[21] How much better to create a class of workers that are demonstrably *not* human, and therefore exempt from the baggage of exploitation—for something that isn't *human* cannot by definition be *dehumanized*! Having no emotion, no *irritability*, Robots also lack class consciousness: while aware of their subaltern status, they lack the self-reflection necessary to consider this status unfair or degrading—it simply *is*. Only the occasional spell of "Robot's cramp," in which a Robot will suddenly stop work and gnash his teeth, suggests otherwise. While Domin and the other managers dismiss this occurrence as "some breakdown in the mechanism… a flaw in the works that has to be removed" (14) Helena sees it as evidence of the Robot's emerging soul. This validates her original mission—as a member of the Humanity League, Helena initially came to the island in order to "liberate the Robots"(13)—to ensure they are treated like human beings. In order to improve their lot, Helena convinces Dr. Gall to make the Robots "more human" so that they may be "equal" to man (40). But humans and Robots are *not* equal, and as the Robot Radius (the first of this new breed of Robot) demonstrates, awareness of this can only end poorly for the humans:

> HELENA: Do you hate us? Why?
>
> RADIUS: You are not as strong as the Robots. You are not as skillful as the Robots. The Robots can do everything. You only give orders. You do nothing but talk.
>
> HELENA: But someone must give orders.
>
> RADIUS: I don't want any master. I know everything for myself.
>
> HELENA: Radius, Dr. Gall gave you a better brain than the rest, better than ours. You are the only one of the Robots that understands perfectly. That's why I had you put into the library, so that you could read everything, understand everything, and

[21] Ibid.

then— oh, Radius, I wanted you to show the whole world that the Robots are our equals. That's what I wanted of you.

RADIUS: I don't want a master. I want to be master. I want to be master over others.

HELENA: I'm sure they'd put you in charge of many Robots, Radius. You would be a teacher of the Robots.

RADIUS: I want to be master over people. (26-27)

Possessing a self-reflexivity not present in earlier Robots, Radius rejects his position as subaltern and, lacking the compassion that motivated Helen's actions, decides that the most logical course is to reverse the roles, casting himself as master over a human subaltern.

At the end of *R.U.R.,* Rossum's creations have truly become examples of *Homo artificialis*. Like Frankenstein's Creature of a century before, Radius—as one of Dr. Gall's "improved" Robots—possesses the vital force, sentiment, and free will which mark him as one of these new beings. While Robots were always considered living by their creators, and thus possessed of a vital force, it was not until Dr. Gall introduced the vitalist faculty of *irritability* that Robots developed sentiment. Upon threat of being sent to the stamping-mill, Radius reacts violently, performing an exhibition of self-preservation heretofore unknown among Robots:

DR. GALL: Reaction of the pupils; increase of sensitiveness. It wasn't an attack characteristic of the Robots.

HELENA: What was it, then?

DR. GALL: Heavens knows. Stubbornness, anger or revolt— I don't know. And his heart, too!

HELENA: What?

> DR. GALL: It was fluttering with nervousness like a human heart. He was all in a sweat with fear, and—do you know, I don't believe the rascal is a Robot at all any longer. (28)

Now that Radius is performing physical reactions that would be expected of an average human in his situation, Dr. Gall recognizes that the Robot has become something more: no longer a Robot, yet not quite human.

While Radius' sentimentality is mainly of the negative variety—he only manifests pain, fear, stubbornness, and hate—two other Robots, Primus and Helena, exhibit more positive sentiments in the Epilogue—wonder, an appreciation of beauty, and the racing heart and impulse toward self-sacrifice in the face of loss that Alquist (and presumably the audience) reads as emerging love between the pair. Finally, all of these "improved" Robots exhibit free will. Earlier Robots were easily controlled, obeying any command given them. With the exception of the occasional bout of "Robot's cramp," they were entirely tractable, to the point that they had to be taught to defend themselves: Helena reminds Domin of "when the working men in America revolted against the Robots and smashed them up, and when the people gave the Robots firearms against the rebels"(23). In the "improved" Robots, however, an emerging self-reflexivity gives rise to the ability to exercise free will. In the scene quoted above, Radius refuses to work, but it is not the "Robot's cramp" previously seen; instead, it is a conscious objection to his status as subaltern. Similarly, Primus refuses Alquist's direct order to take the Robotess Helena to the dissecting room, offering instead to take her place (57). This combination of human-like behaviors leads Alquist to dub Primus and Helena the new Adam and Eve, bidding them go forth into the world in a bittersweet reenactment of the Abrahamic creation myth.

Like Frankenstein's Creature, Čapek's Robots—or, at least Dr. Gall's "improved" Robots—are able to clearly articulate their own consciousness. Both clearly represent a new take on the concept of the artificial human, representing something more than the golem, homunculus, or automaton of

former times. However, both these examples of *Homo artificialis* differ from the ones we will examine in the next chapter in one very important way: they are both without question *biological* constructs. Frankenstein's Creature is crafted from dead human remains reanimated through some arcane scientific secret; Čapek's Robots are manufactured from a new form of protoplasm discovered on a remote island, grown in vats and "stitched" into living organs. While both the Creature and the Robot problematize the ontological status of human in their separate quests for identity, they still adhere to the basic vitalistic principle of the Enlightenment: namely, that biology defines ontology, that life is a characteristic exclusively inherent to living matter rather than the result of particular processes or functions. Only living matter is capable of being *ensouled*. In the fictional universe in which these creations exist, mankind is reduced to a machine, but it is a machine still made of living matter. In contrast, the *Homo artificialis* of recent imagination is one with mechanical origins: not man becoming a machine, but rather machine replacing man.[22] Production has always valued efficiency and cost-effectiveness, and as *R.U.R.* attests, machine labor (even of the biological sort) is cheaper than human labor; therefore, as machines become more efficient, more "intelligent," and less expensive, human workers are seen as a liability.

The machines threatening to replace humans during the nineteenth century were not humanoid ersatz workers but rather industrial machines: the steam engine, the automated loom, and the assembly line. The earlier versions of these machines "functioned as a kind of prosthesis, or extension of the artisan's body,"[23] representing no more than high-powered tools that still required the supervision of human workers, albeit fewer of them. As the Industrial Revolution progressed, however, the machines also became more complex, requiring less skilled intervention for an ever increasing output. Rather than a few skilled craftsmen multiplying their output through the use of mechanized assistance, the new model of production required only the knowledge of how to operate the machine itself (as the

[22] In this formulation, the characteristics that make us human are separate from the biological substrate in which they are instantiated—an important position within the study of theory of mind, and one to which we will return later in this chapter.
[23] Reilly, *Automata and Mimesis* 163.

managers in *R.U.R.* exemplify, as they are able to operate the machines that manufacture the Robots but cannot replicate their formula), fully justifying workers' fear of being replaced by machines. For the working class, this was a disaster, as the stories of the factory conditions during the Industrial Revolution show. But a machine of another type was on the horizon, one whose strength lay not in the brute force and physical prowess of the industrial machine, but rather in the ability to quickly and accurately perform logical operations and calculations—the activities of mental labor. Even at the height of the Machine Age, the roots of a new paradigm were evident: the computing machine, which would prove as revolutionary to the twentieth century as the automated loom was to the nineteenth.

THE RISE OF THE COMPUTING MACHINE

Although the trio of automata created by Jacques Vaucanson between 1737 and 1738 brought their inventor great fame and even greater success, in 1741 he sold them to a group of businessmen from Lyons[24] and turned his attention to other pursuits. Louis XV of France, who had greatly admired Vaucanson's Duck, appointed Vaucanson to the position of Inspector of Silk Manufacture, hoping that the Duck's creator could bring some of his mechanical genius to bear in rescuing the struggling French silk industry.[25] His efforts in this vein revolutionized the industrial process,[26] and in 1747 he introduced an automated silk loom—a central part of which, the mechanism that regulated the speed of the weaving process, was "manufactured from a steel drum identical to that of the automaton Flute Player."[27] Vaucanson's silk loom proved highly unpopular among silk workers at the time. Kara Reilly recounts the story of the workers' uprising:

> Silk workers in Lyons quickly realized that Vaucanson's loom would make their jobs obsolete, and they rioted at the sheer outrage of a machine replacing a human being. Vaucanson was forced to flee Lyons, dressed as a monk, in order to avoid being

[24] Gaby Wood, *Edison's Eve: A Magical History of the Quest for Mechanical Life* (New York: Anchor Books, 2002), 31.
[25] Ibid., 40-41.
[26] Ibid., 41.
[27] Reilly, *Automata and Mimesis* 162.

> murdered. In response to this presumably harrowing experience, Vaucanson designed and built a silk loom manned by a donkey, a loom from which an elegant baroque floral fabric was produced, in order to prove that an ox, a horse, or even an ass could make cloth more beautiful and flawless than even the most reliable workers.[28]

This "improved" loom, a reconstruction of which resides in the museum of the Conservatoire des Arts et Métiers in Paris, was built upon an ingenious system of punched-paper cards to determine the pattern woven into the cloth,[29] thus removing the human artisan from the production process. Vaucanson developed a working model of his automated loom, but due to the unrest among their workers, the silk industry in Lyons never adopted it for production purposes.

The ingenuity of this "improved" loom did, however, serve as a source of inspiration for Joseph Marie Jacquard (1752–1834), who gained hands-on experience with Vaucanson's invention through being tasked with its repair for the Conservatoire.[30] Approached by the Lyonese silk manufacturers some decades after the Vaucanson debacle, Jacquard began development on the loom which today bears his name; by 1804 Jacquard had perfected his design, and despite resistance from the silk workers,[31] the Jacquard loom was rapidly adopted by the silk industry, with over ten thousand looms in use in France by 1812 and spreading to England by the 1820s.[32] According to historian Abbott Usher, Jacquard and his

---

[28] Ibid.

[29] Frenchman Basile Bouchon is generally credited with devising (in 1725) an early version of this system, in which perforated paper passed around a cylinder in order to push back needles that were not part of the pattern; Jean-Baptiste Falcon replaced the continuous paper in 1728 with individual "punchcards," each of which represented a single row of the pattern. Both of these still required manual operation by a human weaver, however; Vaucanson's version was completely automated and needed human intervention only to change out the pattern or empty spools of thread (E. Broudy, *The Book of Looms: A History of the Handloom from Ancient Times to the Present* [University Press of New England, 1979], 134). See also Abbott Payson Usher, *A History of Mechanical Inventions*, Rev. ed. (New York: Dover, 1988), 289-93.

[30] Reilly, *Automata and Mimesis,* 164. Usher, however, claims that it was Lyonese silk manufacturer Dutilleu who urged Jacquard to find Vaucanson's model in order to finish Vaucanson's work and automate the silk industry in Lyons. (Usher, *A History of Mechanical Inventions*, 292).

[31] Like Vaucanson's before it, Jacquard's loom met with hostility from the silk weavers, who sabotaged and burned the looms as well as attacking Jacquard himself. However, the French authorities intervened to put down the rebellions. See Reilly, *Automata and Mimesis,* 164.

[32] Ibid.

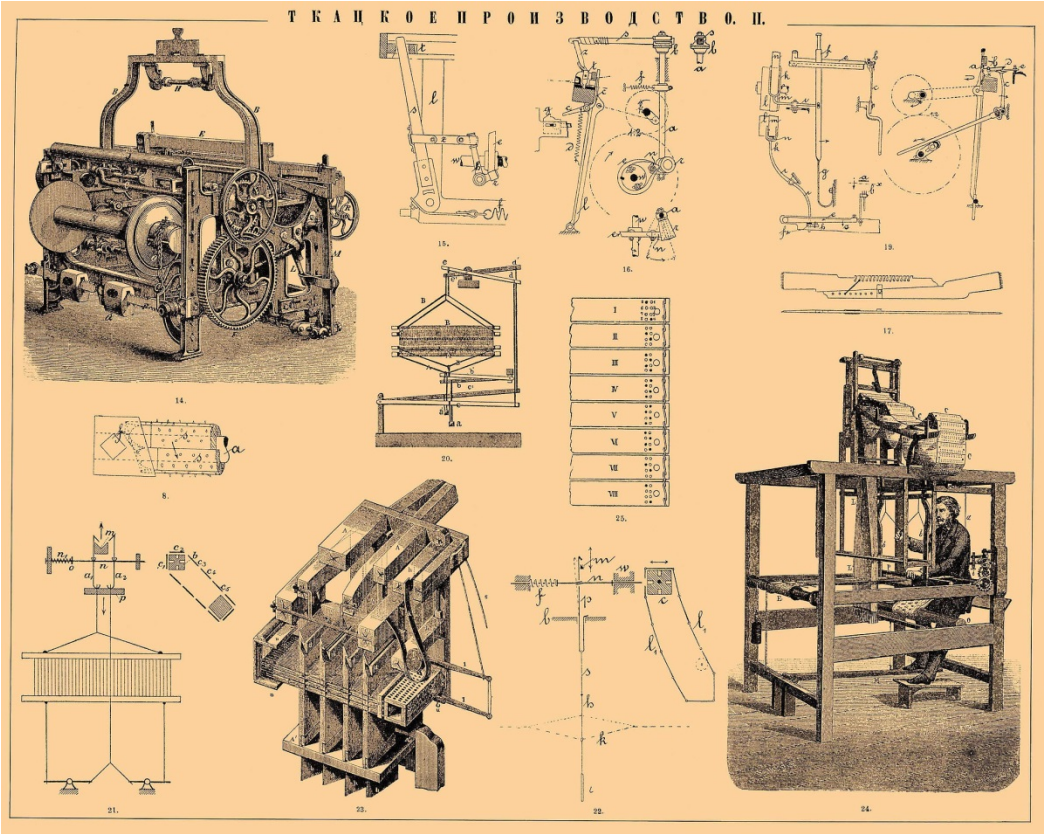


Figure 29. Illustration of the Jacquard automatic loom, showing (bottom right) the chain of punched card patterns that inspired Babbage. From the Brockhaus and Efron Encyclopedic Dictionary *(*Энциклопедический словарь Брокгауза и Ефрона*; St. Petersburg: Brockhaus and Efron Publishing House, 1907).*

team "started with the fragments of a highly refined piece of mechanism; they worked with such thoroughness that no primary feature of the action has since been changed."[33]

While the automated loom proved central to the industrialization of cottage industries across Europe during the nineteenth century, it was the punched-paper control mechanism (see Figure 29) that drew the attention of English philosopher and engineer Charles Babbage (1791–1871). In his memoirs, Babbage outlines his proposal for an "Analytical Engine"—a general purpose mechanical computing machine—that operates along similar principles to the Jacquard loom:

> To those who are acquainted with the principles of the Jacquard loom, and who are also familiar with analytical formula, a general idea of the means by which the Engine

[33] Usher, *A History of Mechanical Inventions*, 293.

> executes its operations may be obtained without much difficulty.… It is known as a fact that the Jacquard loom is capable of weaving any design which the imagination of man may conceive. It is also the constant practice for skilled artists to be employed by manufacturers in designing patterns. These patterns are then sent to a peculiar artist, who, by means of a certain machine, punches holes in a set of pasteboard cards in such a manner that when those cards are placed in a Jacquard loom, it will then weave upon its produce the exact pattern designed by the artist…. But the manufacturer might use the same cards, and put into the warp threads of any other colour. Every thread might even be of a different colour, or of a different shade of colour; but in all these cases the form of the pattern will be precisely the same—the colours only will differ. The analogy of the Analytical Engine with this well-known process is nearly perfect.[34]

Babbage envisioned his Analytical Engine as a machine capable of performing any number of complex algebraic calculations quickly and accurately, using one set of punched cards as the formulae (the weaving pattern) and another set as the variables (the threads). The end product would be the result of the calculation, performed automatically and guaranteed to be accurate within the tolerances of the machine itself (which Babbage designed for fifty decimal places)—a huge improvement over manual calculations, which were done incorrectly often enough that it was common practice at the time for several mathematicians to simultaneously perform the same work in order to cross-check their results.

Babbage first conceived of the Analytical Engine in 1837 as an improvement upon an earlier design for a computing machine, the Difference Engine, which had grown out of Babbage's frustration with the manual calculation of logarithmic tables:

> One evening [in 1812 or 1813] I was sitting in the rooms of the Analytical Society at Cambridge, my head leaning forward on the Table in a kind of dreamy mood, with a

[34] Charles Babbage, *Passages from the Life of a Philosopher* (London: Longman, Green, Longman, Roberts, & Green, 1864), 116.

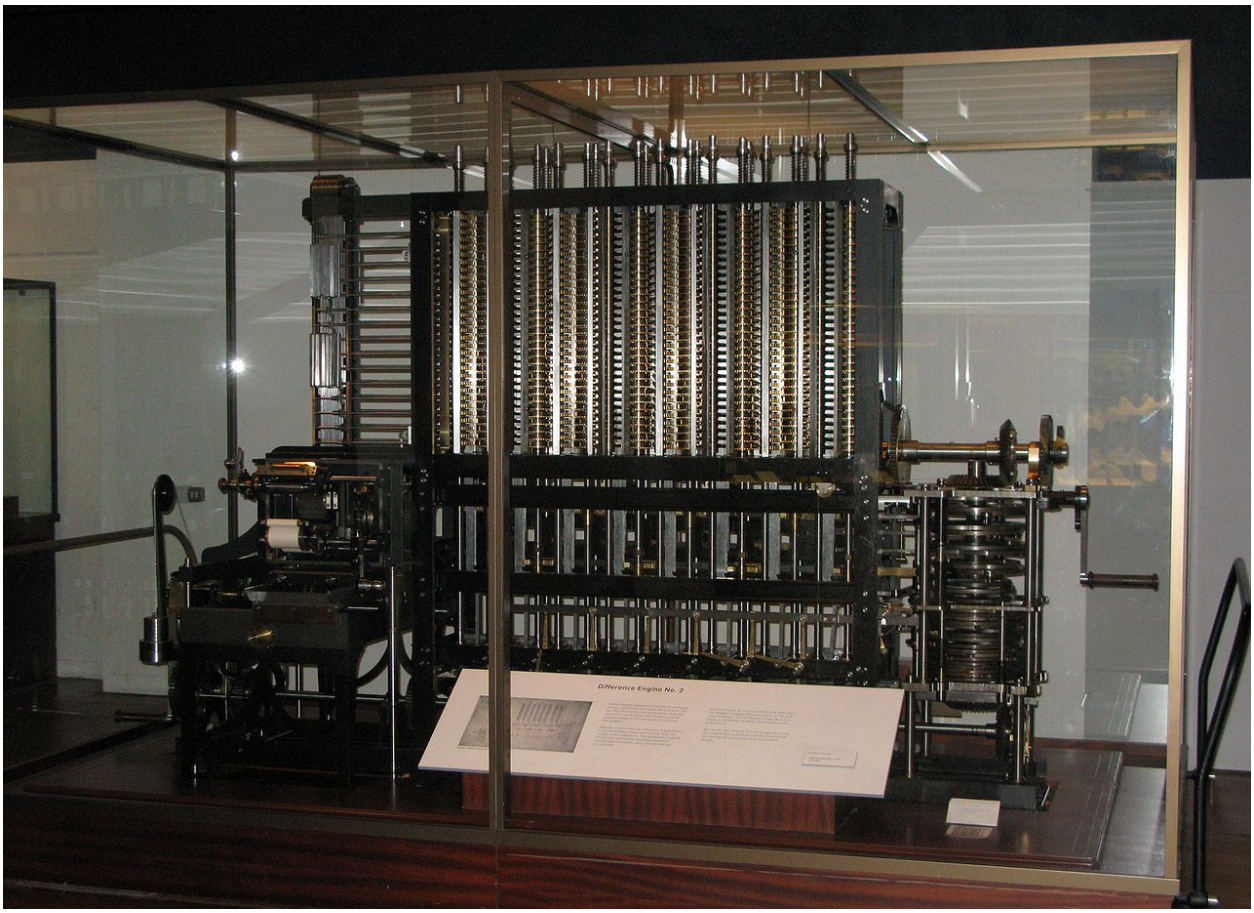


Figure 30. A photo of the Difference Engine constructed by the Science Museum based on the plans for Charles Babbage's Difference Engine No. 2. Photo taken September 2008 by "Geni" and is available via GNU Free Documentation License.

> Table of logarithms lying open before me. Another member, coming into the room, and seeing me half asleep, called out, "Well, Babbage, what are you dreaming about?" to which I replied, "I am thinking that all these Tables (pointing to the logarithms) might be calculated by machinery."[35]

By 1819 Babbage had begun drawing up concrete plans for such a machine, and in 1823 he received a grant from the British government to begin construction on his Difference Engine, designed to calculate logarithmic tables using the method of *finite differences* and thus avoiding multiplication and division.

[35] Ibid., 42.

While the Difference Engine was conceived as a limited-purpose machine, designed only to calculate logarithmic tables, the Analytical Engine was of a more general nature, incorporating arithmetical logic, conditional loops, and integrated memory in the form of Jacquard's punch cards, creating a machine that "embodies almost all the important functions of the modern digital computer."[36]

The majority of histories of computing begin with Babbage; many also trace his influence back to the Jacquard loom. For this study it is also important to highlight the connection between Babbage (and thus the history of computing) and automata: first, in the link from Vaucanson's Flute Player through the silk loom to Babbage's Engines, as we have already seen; and second, in Babbage's interest in automata from a young age. In his memoirs, Babbage recounts his first encounter with automata as a child:

> During my boyhood my mother took me to several exhibitions of machinery. I well remember one of them in Hanover Square by a man who called himself Merlin. I was so greatly interested in it, that the Exhibitor remarked the circumstance, and after explaining some of the objects to which the public had access, proposed to my mother to take me up to his workshop, where I should see still more wonderful automata. We accordingly ascended to the attic. There were two uncovered female figures of silver, about twelve inches high.[37]

Babbage was so taken by these figures that he acquired one of them later in life (which he dubbed the Silver Lady), restored her, clad her in Chinese crepe, and displayed her under glass for the amusement of his friends.[38] During his early work on the Difference Engine, Babbage also had occasion to observe the infamous Chess-Playing Turk[39] in action. While this "automaton" ultimately proved to be a hoax, it

---

[36] Martin Campbell-Kelly, William Asprey, Nathan Ensmenger, and Jeffery R. Yost. *Computer: A History of the Information Machine. The Sloan Technology Series*, (Boulder, CO: Westview Press, 2013), Kindle edition, 42.
[37] Babbage, *Passages from the Life of a Philosopher*, 17.
[38] Ibid., 364-66.
[39] I have deliberately chosen not to include the Turk in my history of automata, for while the Turk was toured internationally as a chess-playing machine, it was revealed decades later to be actually controlled by a human operator hidden within the base. See Tom Standage, *The Turk: The Life and Times of the Famous Eighteenth-Century Chess-Playing Machine* (New York: Walker & Co, 2002).

piqued Babbage's curiosity as to whether an machine could actually be designed to play games of skill, heretofore considered the exclusive realm of human reason; he ultimately concluded that his Analytical Engine would be up to the task: "Allowing one hundred moves on each side for the longest game at chess, I found that the combinations involved in the Analytical Engine enormously surpassed any required, even by the game of chess."[40] Babbage even sketched out a plan for a Tic-Tac-Toe-playing automaton, considering briefly the idea of touring such a "wonder" in order to raise money for his Analytical Engine project.[41] Many historians considered Babbage before his time. His Analytical Engine was never completed during his lifetime,[42] although Babbage tinkered with the concept and design from its first description in 1837 until his death in 1871, even building a trial model of a section of the machine near the end of his life.[43]

While the automatic loom that was a source of inspiration for Babbage's Engines was rapidly industrializing the textile industry at the turn of the twentieth century in Europe, punch card technology of another sort was revolutionizing the American census. Frustrated by the time it took to process the data generated by a rapidly-increasing population, the US Bureau of the Census held a competition in 1888 to find a faster alternative. The winner was American engineer and entrepreneur Herman Hollerith, who devised an automated system of data storage and manipulation using what he called a "punch photograph"[44] of each person's census information. In trial runs, Hollerith's system handily beat those of the competitors, proving up to ten times faster than their manual systems. The Hollerith Electric Tabulating system completed the 1890 US census in two and a half years (as opposed to seven

---

[40] Babbage, *Passages from the Life of a Philosopher*, 467.

[41] See ibid., 465-71 for Babbage's account of his development of this automaton. Babbage ultimately decided against this, as at this point most touring automata "were entire failures in a pecuniary point of view" (ibid., 470).

[42] For a discussion of the myriad setbacks and failures Babbage encountered in his quest to build his Engines, see Campbell-Kelly, et al, *Computer*, 41-46.

[43] This trial model, built by Babbage from parts of the uncompleted Difference Engine, is currently on display at the Science Museum in London. See Anthony Hyman, *Charles Babbage: Pioneer of the Computer* (Princeton: Princeton University Press, 1985), 246. See also the Science Museum's website display on the Analytical Engine: Science Museum, "Babbage's Analytical Engine, 1834–1871 (Trial model)," accessed 20 March 2014, http://www.sciencemuseum.org.uk/objects/computing_and_data_processing/1878-3.aspx?keywords=babbage+analytical+engine.

[44] Hollerith was inspired by train conductors' technique of using a series of holes punched in tickets to identify passengers. Hollerith extrapolated this into a single punch card for each individual's census return. Campbell-Kelly, et al, *Computer*, 14.

Figure 31. Replica of early Hollerith punched card tabulator and sorting box (right) at the Computer History Museum in Mountain View, CA. Photo © Adam Schuster and is licensed under the Creative Commons Attribution 2.0 Generic license.

for the previous census) at an estimated two-thirds the cost of the previous system; the company Hollerith founded to oversee construction of the machines, the Tabulating Machine Company, was one of the larger firms folded into the Computing-Tabulating-Recording Company (CTR) in 1911, earning Hollerith $1.2 million and a place on the board until his retirement. In 1924 CTR was renamed the International Business Machine Corporation (IBM), today the largest technology company in the world.[45]

The twentieth century saw an explosion in automated computation, and a century after Hollerith's electric tabulation system, which required hundreds of human operators, computing power orders of magnitude greater can be held in the palm of your hand. While a detailed account of this development is best left one of the many extant histories of computing, a brief outline of major

[45] Ibid.

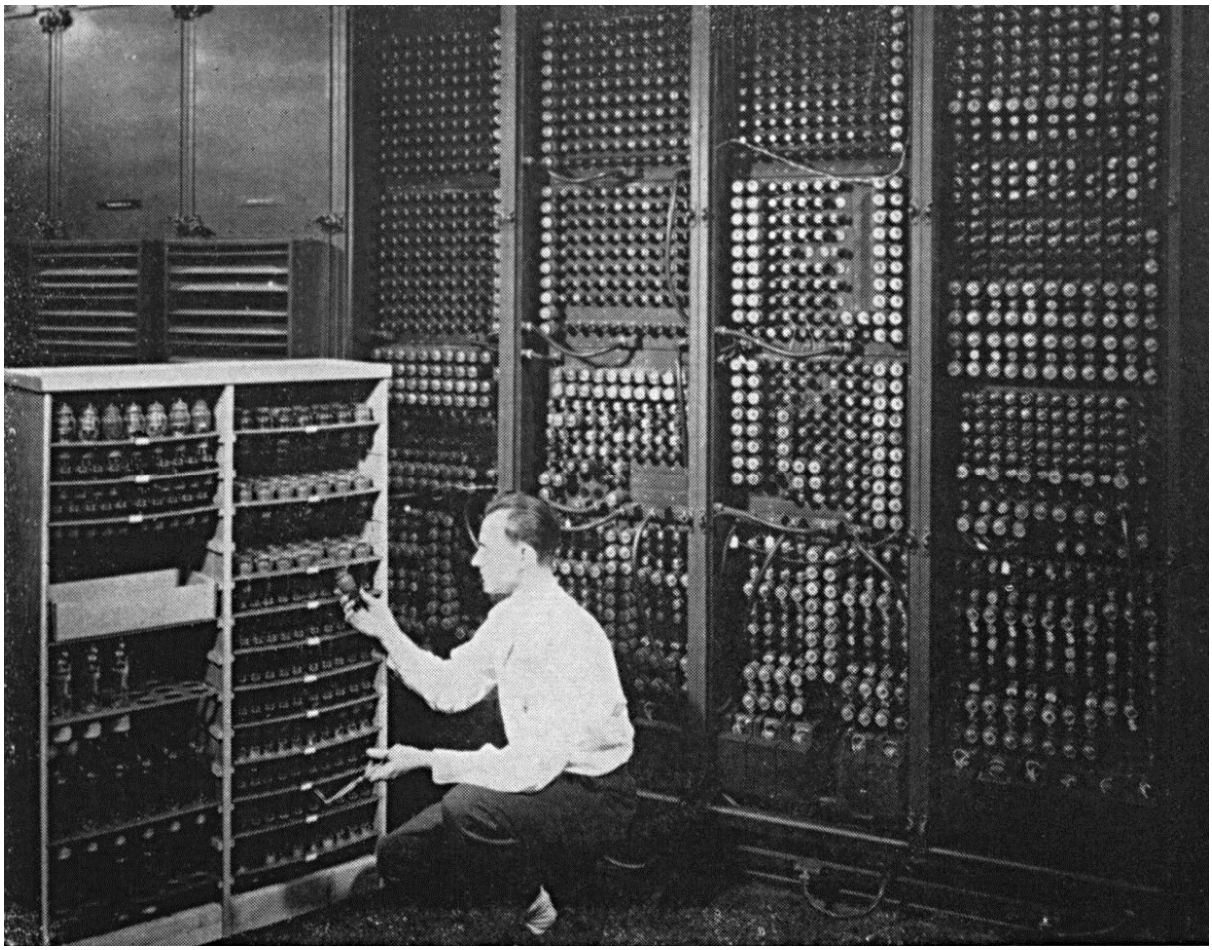


Figure 32. A technician replacing a burnt-out tube on ENIAC, a process of elimination among the thousands of possibilities. U.S. Army photo from "The ENIAC Story," Martin H. Weik, 1961.

milestones is in order. By the 1930s, the punched-card system invented by Hollerith was ubiquitous in business and government accounting and recordkeeping, led by IBM; at the same time, desktop mechanical calculators able to perform basic mathematical operations became commonplace in offices and shops throughout the US and Europe. As the threat of a second World War increased, governments in the US and UK focused on efforts to develop faster and more powerful computing machines to aid in the war effort, and in 1943 *Colossus* was born: the first electronic, programmable, digital computer, developed by the British, whose codebreaking capabilities proved decisive to the outcome of WWII. The United States followed suit in 1946 with the *Electronic Numerical Integrator and Computer* (ENIAC). ENIAC made use of similar vacuum tube technology as Colossus, but was faster and more versatile:

Colossus, the first system to use vacuum tubes on such a large scale, consisted of around 1,500 tubes; ENIAC boasted over 17,000.

Colossus and ENIAC were "programmable" only in the most rudimentary sense. Due to the speed at which they operated, a program on one of these machines was defined through the physical configuration of patch cables and switches; thus the computer could only be configured to run one program at a time:

> But the ENIAC would perform five thousand operations per second, and it was infeasible to use a card or tape reader at that kind of speed. Instead, Eckert and Mauchly decided that the machine would have to be specially wired-up for a specific problem. At first glance, when programmed, the ENIAC would look rather like a telephone exchange, with hundreds of patch cords connecting up the different units of the machine to route electrical signals from one point to another. It would take hours, perhaps days, to change the program.[46]

This hardwiring proved a serious shortcoming, leading computer pioneers to work toward a more streamlined architecture in which both instructions and data were stored electronically: the stored-program computer.[47] While several different machines have been granted the title of "the world's first practical stored-program computer,"[48] by the early 1950s such machines were commercially available, heralding the dawn of the computer age.

---

[46] Campbell-Kelly, et al., *Computer*, 73.

[47] The credit for this architecture, often referred to as *Von Neumann* architecture after mathematician and physicist John von Neumann, is a source of some controversy. Von Neumann and Herman Goldstein released a first draft of a report on the EDVAC in 1945, describing this architecture, but failed to credit other collaborators; this release was considered a legal publication, which prevented patenting of the EDVAC and resulted in other members of the design team not receiving credit for their work. See ibid., 78-79.

[48] Campbell-Kelly, et al. confer this title on the Electronic Delay Storage Automatic Calculator (EDSAC), built at the University of Cambridge, which ran its first program on 6 May 1949 (ibid., 86); Jack Copeland refers instead to the prototype Manchester Small-Scale Experimental Machine (SSEM, nicknamed "Baby") designed as proof-of-concept for the idea of stored-program computing, which ran its first program on 21 June 1948 (B. Jack Copeland and Alan Mathison Turing, *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, Plus the Secrets of Enigma* [Oxford: Oxford University Press, 2004], 16).

ALAN TURING AND THE QUEST FOR ARTIFICIAL INTELLIGENCE

The concept for stored-program computing stems from the work of Alan Turing (1912–1954), British mathematician, logician, cryptanalyst, and computer scientist. After graduating with honors in mathematics from King's College in Cambridge in 1934, Turing published the theoretical essay "On Computable Numbers, with an Application to the *Entscheidungsproblem* [Decision Problem]" (1936), which is widely regarded as the work which founded the field of modern computer science. "On Computable Numbers" outlines Turing's pioneering theory of computation, including a description of an abstract "universal machine" (now known as a Turing machine) capable of performing any conceivable mathematical computation representable as an algorithm. Unlike the punched-card systems that were common at the time, in which one set of cards provided the instructions and another the data, or the hardwired computing machines that were developed soon after, Turing's universal machine is envisioned as a simple computing system in which a scanner head reads both instructions and data from a limitless memory-tape—a theoretical stored-program computer. Working through the tape one symbol at a time, the scanner head is also able to erase the symbol on the tape, write a new symbol, and move the tape back and forth, thus enabling it to constantly refer back to previous instructions and store new information on the same memory-tape. This makes Turing's machine capable of imitating any other simply by reading its description from the memory-tape. Turing himself described the connection between his universal machine and the stored-program computer in a lecture to the London Mathematical Society in 1947, as the first stored-program computers were being developed:

> It can be shown that a single special machine of that type can be made to do the work of all. It could in fact be made to work as a model of any other machine. The special machine may be called the universal machine; it works in the following quite simple manner. When we have decided what machine we wish to imitate we punch a description of it on the tape of the universal machine.

> This description explains what the machine would do in every configuration in which it might find itself. The universal machine has only to keep looking at this description in order to find out what it should do at each stage. Thus the complexity of the machine to be imitated is concentrated in the tape and does not appear in the universal machine proper in any way.
>
> If we take the properties of the universal machine in combination with the fact that machine processes and rule of thumb processes are synonymous we may say that the universal machine is one which, when supplied with the appropriate instructions, can be made to do any rule of thumb process. This feature is paralleled in digital computing machines such as the ACE. They are in fact practical versions of the universal machine.[49]

The ACE, or Automatic Computing Engine, was a project of the National Physical Laboratory (NPL) in London, headed by Turing, to develop a concrete form of Turing's universal machine. Today, any computing machine capable of performing these functions is referred to as "Turing complete."

Unlike other early stored-program systems under development at the time, the ACE reflected Turing's philosophy to achieve complex behavior through complex programming rather than complex equipment—basically making the computer system more intelligent rather than more powerful.[50] This philosophy drove Turing's pioneering theoretical work on the idea of Artificial Intelligence (AI): in 1948 he completed a report entitled "Intelligent Machinery" for the NPL; although it was never published, having been rejected by the director of the NPL as unsuitable, B Jack Copeland hails the report as "the first manifesto of AI."[51] This report first outlines many of the concepts which later became central to the study of artificial intelligence and artificial life,[52] including the concept of a genetic algorithm (central to

---

[49] Copeland and Turing, *The Essential Turing*, 383.
[50] Ibid., 366.
[51] Ibid., 401.
[52] Artificial life, or *A-life*, is a field of study founded in the late 1980s that examines the processes and evolution of living systems through their recreation or simulation through computer modelling, robotics, and biochemistry. AI is considered a form of *soft*

research in A-life), an early description of the "imitation game" for evaluating machine intelligence (which I will return to in a moment), and a discussion of machine learning in which Turing compares machine and human intelligences.

By far Turing's best-known contribution to the budding field of artificial intelligence is his essay "Computing Machinery and Intelligence," first published in the philosophy journal *Mind* in 1950. This essay was dubbed more propaganda than philosophy by fellow logician Robin Gandy, who claimed that "Turing thought the time had come for philosophers and mathematicians and scientists to take seriously the fact that computers were not merely calculating engines but were capable of behaviour which must be accounted as intelligent; he sought to persuade people that this was so."[53] For Turing, the question "Can machines think?"—made popular with the rise of computing machines—was "too meaningless to deserve discussion."[54] Turing's essay offered a more practical alternative: a mind experiment in the form of *the imitation game*, a parlor game in which an interrogator (C) must determine through question and answer which of two respondents is male (A) and which is female (B), as seen in Figure 29 below. Turing's proposed imitation game places a machine in the role of one of the human participants:

> We now ask the question, "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?" … The new problem has the advantage of

---

(from software) A-life as it seeks to replicate the neurological processes and capabilities of the human mind. For more information see Christopher G. Langton, *Artificial Life: An Overview* (Cambridge, MA: MIT Press, 1997).

[53] Robin Gandy, "Human versus Mechanical Intelligence," in Peter Millican and Andy Clark, eds, *Machines and Thought: The Legacy of Alan Turing*, 2 vols. (New York: Clarendon Press, 1996), 125.

[54] Copeland and Turing, *The Essential Turing*, 449.

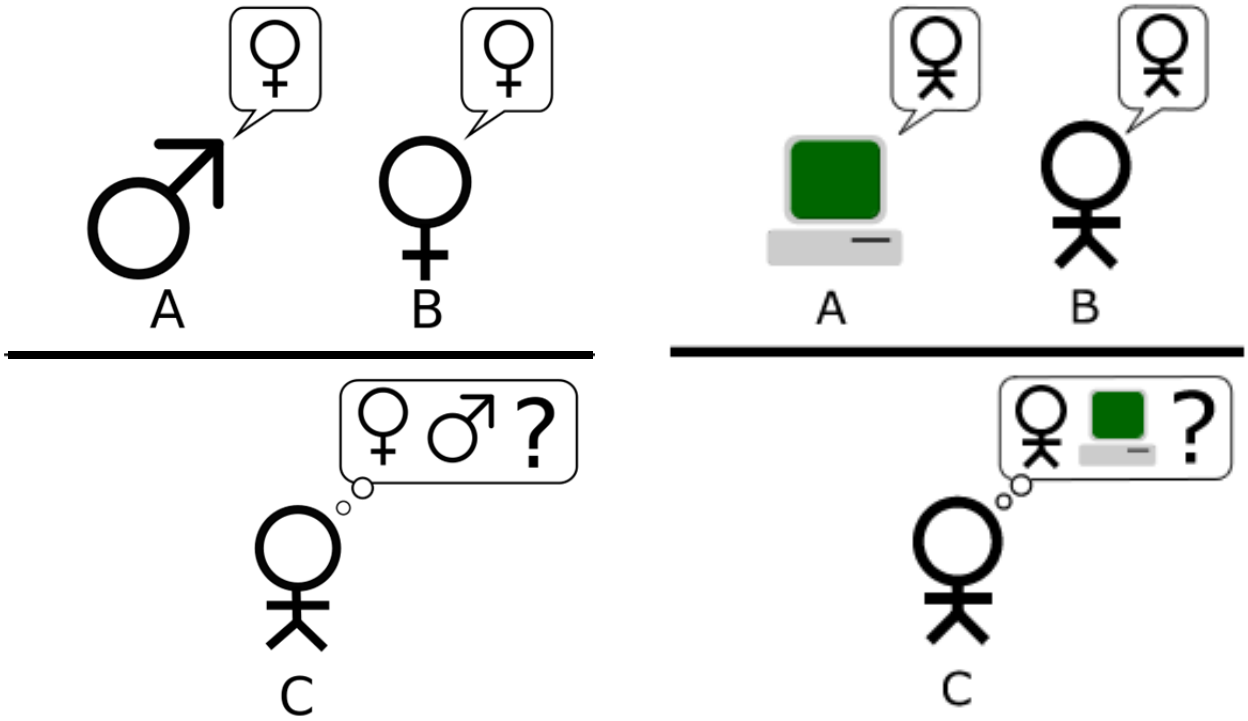


Figure 33. The imitation game, as described by Alan Turing in "Computing Machinery and Intelligence." Player C, through a series of written questions, attempts to determine which of the two players is the man and which of the two is the woman. Player A—the man—tries to trick player C into making the wrong decision, while player B tries to help player C. Turing uses this game as the basis for his test for intelligence, substituting a machine for Player A as seen in the diagram on the right. Illustration © 2011 by Hugo Férée and is licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license.

> drawing a fairly sharp line between the physical and the intellectual capabilities of a man.[55]

In a move reminiscent of Descartes' *bête machine* concept as seen at the opening of this chapter, the now-famous "Turing Test" establishes *verbal behavior* as a criterion for determining whether an entity is intelligent; while Turing never intended success in the imitation game as a *definition* of intelligence, he nevertheless viewed it as a *sufficient condition* for its existence.

The validity and ramifications of the Turing Test have been hotly debated in the half-century since its first proposal. Turing himself predicted success: "I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the

[55] Ibid., 441-42.

imitation game so well that an average interrogator will not have more than seventy percent chance of making the right identification after five minutes of questioning."[56] Further, Turing predicted that "at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted."[57]

While Turing's predictions have yet to come true, it isn't for lack of trying: in 1990 American inventor Hugh Loebner (b. 1942), in conjunction with the Cambridge Center for Behavioral Studies in Massachusetts, instituted an annual competition in search of a machine capable of passing the Turing Test. This annual competition, based on a restricted version of Turing's imitation game, [58] is known as the Loebner Prize, and awards a Bronze Medal and small cash prize ($4000 in 2013) to the most human-like chatbot[59] participating in each year's event. Additionally, the committee offers two as-of-yet-unclaimed prizes: the first is the Silver Medal, which includes a $25,000 cash award, for the first program to "fool" two or more judges when compared to two or more humans. This would open a "multi-modal" stage for future competitions, incorporating text, visual, and auditory input. The first program that judges cannot distinguish from a real human in a multi-modal Turing Test is to receive a solid eighteen-karat Gold Medal, engraved with the busts of Loebner and Turing, and an additional $100,000 cash award; once this is achieved, the annual competition will end. To date, the Silver and Gold Medals remain unclaimed. However, in June 2014 a computer program named "Eugene Goostman" made

[56] Ibid., 449.
[57] Ibid.
[58] Both the scope (range of possible topics) and tenor (to preclude the use of trickery or guile) of the conversation is limited by the prize committee. See Hugh Gene Loebner, "Home Page of the Loebner Prize in Artificial Intelligence: The First Turing Test," last updated 1 April 2013, accessed 27 March 2014, http://www.loebner.net/Prizef/loebner-prize.html.
[59] A *chatbot* (from *chatting robot*, alternatively *chatterbot* or simply *bot*) is a text-based conversation program designed to mimic human conversation through natural language processing; the most well-known example is the early chatbot ELIZA (MIT 1964–1966), designed to mimic the responses of a Rogerian psychotherapist. Joseph Weizenbaum, who authored ELIZA, noted how much human participants invest in their conversations with ELIZA (Joseph Weizenbaum, "ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine," *Communications of the ACM* 9, no. 1 [1966]: 42).

headlines when it successfully "passed" a Turing Test competition at the Royal Society of London, convincing a number of interrogators that it was a 13-year-old boy from St. Petersburg, Russia.[60]

Since its inception in the 1950s, the field of artificial intelligence has grown exponentially, with implications in areas as far-flung as cognitive science and banking. The world of the twenty-first century could hardly function without availing itself of the myriad applications of modern AI; however, for the purposes of this dissertation, most of these applications are of only tangential interest, as they represent the area of *weak* or *narrow AI*: "artificial intelligence that performs a useful and specific function that once required human intelligence to perform, and does so at human levels or better."[61] A prime example of narrow AI is an internet search engine, able to scour vast depositories of information quickly and deliver the results practically instantly; as search engine algorithms improve, results are more useful, and the best engines offer personalized results based on your location, personal preferences, and past search history.[62] While a search engine is able to perform this function much faster and more efficiently than a human, I would not argue for Google as an example of *Homo artificialis*. Even Deep Blue, IBM's chess-playing computer that beat Grandmaster Gary Kasparov in 1997, is only an example of narrow AI, albeit a very capable one.

The key to a real-life example of *Homo* artificialis similar to those seen in science fiction lies in the realm of *strong AI* or *Artificial General Intelligence (AGI)*: a machine intelligence that is capable of performing *any* intellectual task that a human can. It is this form of AI that is associated with ideas of sentience and consciousness, and is the subject of much debate among AI researchers as to whether it is achievable or even a desirable focus of research. Even the form AGI would take is highly contested; but

---

[60] Eugene managed to convince 33% of the judges (out of 30) that it was human. See University of Reading, "Turing Test Success Marks Milestone in Computing History," press release dated 8 June 2014, accessed 31 October 2014, http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx.

[61] Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (New York: Viking, 2005), 264.

[62] Whether this is a blessing or a curse is a topic for a separate discussion.

while there is no universally accepted definition (the Turing Test is often cited as one possible definition), AI researchers generally agree upon a number of capabilities a successful AGI must display:[63]

Reason—the ability to strategize and to solve problems, even under uncertain conditions or with incomplete information;

Knowledge Representation—the ability to represent information about the world in a form that can be utilized to problem solve (such as in terms of rules, hierarchies, frames, or taxonomies);

Planning—the ability to determine the best path toward a given goal;

Learning—for example, the ability to devise rules from existing data with can then be applied to new data, beyond initial programming (such as in voice recognition, spam filters, and optical character recognition); and

Natural language communication—the ability to derive contextual meaning from natural language input, as well as generate natural-language responses (basically, successful participation in a Turing Test).

While many applications of narrow AI are able to successfully exhibit one or more of these abilities to some extent, a successful display of AGI would integrate all these abilities simultaneously.

For some researchers, AGI is the holy grail of artificial intelligence research; for others, it is an impossibility and a waste of time. Yet while philosophers and computer scientists struggle over predictions of what may be in store for artificial intelligence in the future, another group has taken the prospect of the intelligent machine and given it ample space for development unfettered by technological realities. This group is the science fiction community, who since the early twentieth

[63] See John Haugeland, *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, Rev. and enl. ed. (Cambridge, MA: MIT Press, 1997); Morris W. Firebaugh, *Artificial Intelligence: A Knowledge-Based Approach* (Boston: Boyd & Fraser, 1988);Stefano Franchi and Gèuven Gèuzeldere, *Mechanical Bodies, Computational Minds: Artificial Intelligence from Automata to Cyborgs* (Cambridge, MA: MIT Press, 2005); among others cited elsewhere in this chapter.

century has made use of the age-old automaton—now in the form of a sentient robot—to explore a host of very human issues in print and on screen.[64]

THE GOLDEN AGE OF SCIENCE FICTION

When *R.U.R.* debuted in 1920, its Robots were reminiscent of the homunculi of medieval alchemy, biological machines pitied by Helena (in her role as representative of the Humanity League) as existing in a broken state, needing only a small push to realize their true potential and take their place as the equals of their human masters. Neither Helena nor the managers of the Rossum factory claimed that these beings were *not alive*, only that they were *not human*. The artificial human played a central role in the emerging genre of science fiction; in fact, J.P. Telotte claims it is our technological double, our Baudrillardian *Doppelgänger*, that comprises "the core of the science fiction genre… [in] its abiding concern with artifice and the human, especially with the intersection between the two."[65] While Adam Roberts indicates that the origins of science fiction are "fiercely contested,"[66] by the mid-twentieth century SF had become an immensely popular genre, and the image of the artificial human filled the pages of pulp SF magazines. Initially (i.e., in the 1920s and 1930s) the terms *robot* and *android*[67] were seen as largely interchangeable, applicable to any artificial humanoid creature. Ultimately, however, the

---

[64] I have made a conscious decision here to focus the following section regarding the "Golden Age of Science Fiction" on literary depictions for two reasons: (1) the revolution in the image of the robot discussed in this section was primarily the result of the literature, particularly the short stories collected in periodicals of the time; and (2) I plan to address television and film depictions of robots in the following chapter.

[65] Telotte, *Replications*, 2. Telotte himself uses the Baudrillardian term *simulacrum* in his discussion; for Baudrillard, the *simulacrum* is not a copy of the real, but becomes truth in its own right: the hyperreal. I prefer the German *Doppelgänger*, the physical/behavioral double that is often seen as a bad omen.

[66] Roberts himself offers three separate origin stories. The first claims a long history, dating back to Johannes Kepler's *Somnium, sive astronomia lunaris* (*A Dream, or Lunar Astronomy*) in 1634. The second argues for more Gothic origins, claiming *Frankenstein* as the first science fiction novel and including nineteenth century authors such as Jules Verne and H. G. Wells. The final history is more recent, beginning with *Amazing Stories* creator Hugo Gernsback's coining of the term "science fiction" in the 1920s. See Adam Roberts, *Science Fiction*, 2nd ed, The New Critical Idiom (London: Routledge, 2006), 37ff.

[67] From the Greek root *ἀνδρ-* (man) and the suffix *-oid* (having the form or likeness of), *android* was first used in medieval alchemy to refer to Albertus Magnus' legendary automaton creation. It was popularized in its more modern sense by the French author Auguste Villiers de l'Isle-Adam in his work *Tomorrow's Eve* (1886), in which a fictional version of Thomas Edison creates the artificial human (android) Hadaly as the image of the perfect woman. Hadaly becomes possessed by the spirit Sowana and comes to life, to Edison's delight, but she is lost at sea when Edison's passenger ship sinks on their voyage to their new home. See Wood, *Edison's Eve*, chapter three.

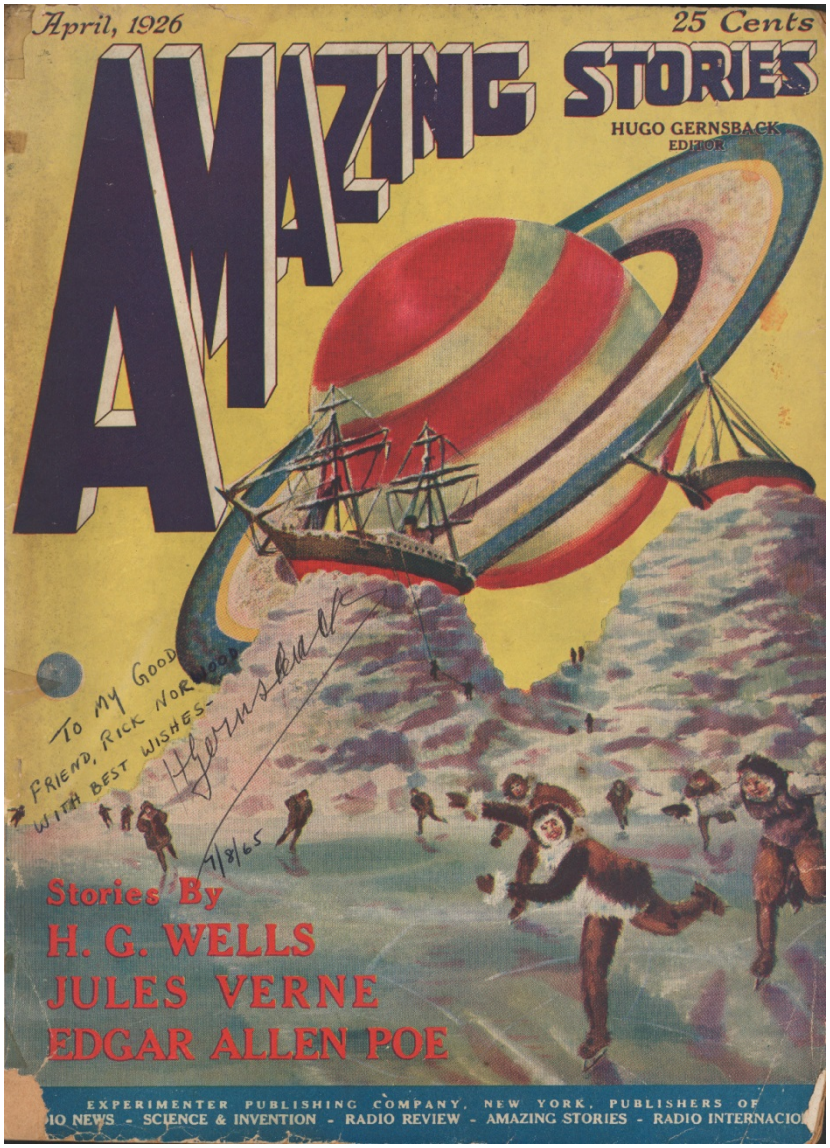


Figure 34. Amazing Stories*, April 1926. Volume 1, Number 1. Founded by Hugo Gernsback. Copyright not renewed, now in public domain. Some consider Gernsback's coining of the term "science fiction" in the 1920s the beginning of SF as a literary genre; Amazing Stories was the first commercially successful pulp magazine dedicated to science fiction (See Roberts,* Science Fiction*, 51).*

term *robot* came to mean a *mechanical* artificial human in the science fiction universe; terms such as *android* remain more fluid in their definition, with the lines blurring between different authors and different subgenres.

Science fiction as a literary genre underwent a radical transformation at the end of the 1930s, however. In mid-1938, editor and publisher John W. Campbell took over the periodical *Astounding*

Stories, renaming it *Astounding Science Fiction* (*ASF*). Within two years of his reworking of the publication, Campbell had published stories by many of the writers who would become central figures in science fiction: not only existing writers, such as L. Ron Hubbard, Jack Williamson, and L. Sprague de Camp, who became regulars in either *ASF* or its sister magazine, *Unknown*; but a new generation of writers as well, who published some of their first stories in *ASF*, such as Lester del Rey, Isaac Asimov, A.E. van Vogt and Robert Heinlein. Historian Adam Roberts credits *ASF* with ushering in what is widely regarded as a "Golden Age of Science Fiction" in the late 1930s and 1940s, bringing the science fiction genre out of its "pulp" origins and granting it popular and critical respectability.[68] This was an era of increased realism and psychological depth of character, along with a shift in focus to so-called *hard SF*, with its emphasis on scientific and technical detail and accuracy. The robot plays a central role in this golden age, as Telotte argues in his discussion of three major authors of the period: Edgar Rice Burroughs, Jack Williamson, and Isaac Asimov. According to Telotte,

> [these authors] seem to sum up the various attitudes toward human artifice that came to dominate the literature. In their focus on the possibilities of organic and mechanical creation, on the menace and the utility of these figures, and on their potential for extending humanity's power and for replacing us, they suggest the complexity of that fantasy of robotism and sketch the sort of dynamic discussion that has eventually come to inform our films.[69]

Although each of the three took very different approaches to the prospect of the artificial human, they served as a reflection of the hopes and concerns of a society that in its recent history had gone through the paradigmatic shift of the Industrial Revolution, seen the rise of the Machine Age, survived the upheaval of one world war—and now stood at the brink of a second.

---

[68] Adam Roberts, *The History of Science Fiction*, Palgrave Histories of Literature (Basingstoke England ; New York: Palgrave Macmillan, 2006), 195. See also Roberts, *Science Fiction*, 56-59.
[69] Telotte, *Replications*, 41.

Edgar Rice Burroughs (1875–1950) is best known for the characters of Tarzan, the jungle hero raised by apes, and John Carter, the Virginian war hero transplanted to Mars (which the locals call Barsoom). In his 1940 novel *Synthetic Men of Mars* (one of the John Carter adventures), however, Burroughs offers "a telling vision of the potential of synthetic beings:"[70] the *hormads*, biological artificial beings created in a similar manner as Čapek's Robots and intended as servants and warriors. Unlike the carefully manufactured and exemplary Robots, however, hormads are horrific creatures, out of proportion and misshapen—only those "which had two arms and legs and the facial features of which were somewhere upon the head"[71] are allowed to live, the rest hacked up and dumped back into the factory vats. These monstrosities are "little more than primal desires given animation,… a grand trick this Martian version of humanity has played upon itself, a faulty mirror we have unwittingly held up to reveal our own nature."[72] This mirroring is redoubled when the more intelligent hormads force their mad scientist creator to devise a method to transplant their brains into the bodies of normal Martian captives, thus having the outward appearance of their enemies but retaining their own identities. The story plays out as a sort of morality play, as the hormad army, threatening to overrun the planet, "can only be resisted by those who manage to control their own desires"[73]—namely, our hero John Carter, who throughout the Barsoom novels represents the best of humanity. After the victory, Carter's loyal guard Vor Daj, whose body has been switched with a hormad in the above-mentioned transplant process, fears that his original body is lost, and he will forever be locked in the misshapen hormad form. In what Telotte claims could stand for the novel's theme, John Carter proclaims of Vor Daj that "It is the character that makes the man… not the clay in which it is abode."[74] Carter's statement reinforces the duality of the mind-body debate, the idea that the mind is the "ghost in the machine" controlling the brute mechanical body.

---

[70] Ibid.
[71] Edgar Rice Burroughs, *Synthetic Men of Mars* (Tarzana, CA: E. R. Burroughs, 1940), 33.
[72] Telotte, *Replications*, 42.
[73] Ibid.
[74] Burroughs, *Synthetic Men of Mars*, 154; Telotte, *Replications*, 42.

Jack Williamson (1908–2006) had a quite disparate, yet similarly bleak, outlook on our future relationship with the artificial human; in his *Humanoids* series (*With Folded Hands*, 1947; *The Humanoids*, 1949; and *The Humanoid Touch*, 1980), Williamson reveals a deep suspicion of technology and its deleterious effects on humankind. In the novelette *With Folded Hands*, a race of small black robots calling themselves "humanoids" appear, offering their services free of charge for all manner of menial tasks throughout the town. Claiming to exist only "To Serve and Obey, and Guard Men from Harm,"[75] the humanoids slowly take over every aspect of human life, to the point where no human is allowed to partake in any activity which may endanger him, or disturb the fulfilment of the humanoids' Prime Directive: human happiness. It seems a paradise, as the advertising from the Humanoid Institute promises:

> In a series of richly colored before-and-after pictures, a chesty blond girl was stooping over a kitchen stove, and then relaxing in a daring negligee while a little black mechanical knelt to serve her something. She was wearily hammering a typewriter, and then lying on an ocean beach, in a revealing sun suit, while another mechanical did the typing. She was toiling at some huge industrial machine, and then dancing in the arms of a golden-haired youth, while a black humanoid ran the machine.[76]

Unfortunately, this seeming paradise takes a dystopian turn, as humans who resist the Prime Directive are chemically lobotomized: in the second novel, protagonist Clay Forester returns home one day to see his wife Ruth sitting on the floor, happily playing with blocks like a baby. When he asks their humanoid what has happened, he discovers the horrible truth:

> "We have merely made her happy," the machine sang gaily. "We have only removed her cares."

---

[75] Jack Williamson, *With Folded Hands,* originally published in *Astounding Science Fiction*, June 1947 (Amazon Digital Services, Inc., 2011), Kindle edition, Location 90.
[76] Ibid., Locations 132-35.

> "And her memory."
>
> "Forgetfulness is the most useful key we have discovered to human happiness," whined the machine. "Our drug, euphoride, relieves the pain of needless memories and the tension of useless fear. Stopping all the corrosion of stress and effort, it triples the brief life expectancy of human beings. You can see that Ruth has lost her fear of age."
>
> "Maybe!" Forester blinked incredulously. "But did she ask for euphoride?"
>
> "No request was necessary."
>
> "I won't have it!" He was hoarse and breathless with his anger. "I want you to restore her mind—if you can!"
>
> "Her mind isn't injured," the machine said briskly. "The drug merely protects her from memories and fears which serve no purpose now, since our service shields her from every want and harm. If that distresses you unduly, then it may be necessary for you to take euphoride also."[77]

While the utopian quest to free man from the drudgery of menial labor resulted in the eventual annihilation of mankind in *R.U.R.,* the outcome in *The Humanoids* is more insidious: as Telotte argues, "These mechanical beings do not so much threaten to replace humanity as promise to make us almost literally forget what being human is all about."[78] By the novel's end, Forester has himself undergone treatments on the humanoids' "healing grid," and with "all the misguided emotions which must have driven him to his unfortunate past actions… fading from awareness, even as he fumbled vaguely for them, like the irrelevant stuff of some unlikely dream,"[79] Forester is left wondering how any human could possibly "care for themselves, without mechanicals?"[80]

---

[77] Jack Williamson, *The Humanoids* (New York: Lancer, 1949), Kindle edition, Locations 1411-22.
[78] Telotte, *Replications*, 46.
[79] Williamson, *The Humanoids*, Locations 3849-51.
[80] Ibid., Locations 3854-55.

An alternate view of the relationship between mankind and technology, one that ended in neither the enslavement, nor the destruction, of the human by his artificial counterpart, arose during the Golden Age of Science Fiction in the work of Isaac Asimov (1920–1992). A prolific writer in the sciences and science fiction as well as a number of other genres, Asimov began his writing career at the age of nineteen, submitting his own short stories to *Astounding Science Fiction*. As he explains in the 1982 introduction to *The Complete Robot*, a comprehensive collection of his robot stories spanning nearly forty years, Asimov's approach to SF was a result of his dissatisfaction with what he saw as the two classes of robot stories he encountered in the pulp stories of his youth:

> In the first class, there was Robot-as-Menace. I don't have to explain that overmuch. Such stories were a mixture of 'clank-clank' and 'aarghh' and 'There are some things man was not meant to know.' After a while, they palled dreadfully and I couldn't stand them.
>
> In the second class (a much smaller one) there was Robot-as-Pathos. In such stories the robots were lovable and were usually put upon by cruel human beings. These charmed me ….
>
> When, therefore, on June 10, 1939 (yes, I do keep meticulous records), I sat down to write my first robot story, there was no question that I fully intended to write a Robot-as-Pathos story. …
>
> But something happened as I wrote this first story. I managed to get the dim vision of a robot as neither Menace nor Pathos. I began to think of robots as industrial products built by matter-of-fact engineers. They were built with safety features so they weren't Menaces and they were fashioned for certain jobs so that no Pathos was necessarily involved.[81]

[81] Isaac Asimov, "Introduction," *The Complete Robot* (London: Voyager, 1995), 9-10.

The new class of robot Asimov created with his stories could be termed Robot-as-Intelligent-Tool, not designed to replace humans but to work alongside them, working in activities or situations that were too difficult, too remote, or too dangerous for their human coworkers. Asimov's stories were also the first to bestow upon these scientifically created mechanical beings emergent human behaviors, a result of their *positronic brains*, the robots' computer control center, envisioned by Asimov as run by the newly discovered particle the *positron* and representing some hypothetical, non-specific future computing technology—unsurprising, since the first actual modern computers (The Colossus and ENIAC) were still in development in 1939.

Not only did Asimov popularize the modern concept of a robot as a *mechanical* artificial being, as opposed to Čapek's original sense of a biological one, but he is also the originator of many of the terms and concepts associated with robots in the late twentieth and early twenty-first centuries, including the term robotics—Asimov himself was not aware he coined the term when he first used it in the May 1941 short story "Liar!" to describe the science and technology of robots.[82] Asimov also introduced the Three Laws of Robotics, a set of rules hard-coded into every positronic brain as a safety measure to protect humans from the kind of Robot-as-Menace scenarios common in pulp SF. These Laws, codified in the 1942 "Runaround," are as follows:

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 to 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.[83]

[82] Ibid., 10.
[83] Ibid., 269-70.

The Three Laws present an inviolable hierarchy in which both obedience and self-preservation are subordinate to the primary directive of protection, and most of Asimov's robot stories revolve around the interplay of these Laws and their implications in various situations. Asimov's laws are as ubiquitous to modern SF depictions of robots as their mechanical structure; they are also taken into consideration in the field of artificial intelligence, whether to praise their usefulness or to dismiss them as unworkable.

Asimov is not modest in assessing his contributions not only to the concept of the robot in contemporary science fiction and popular culture, but to the development of technology in the twentieth century as well:

> As I continued to write robot stories, this notion of carefully engineered industrial robots permeated my stories more and more until the whole character of robot stories in serious printed science fiction changed….
>
> When I wrote my robot stories I had no thought that robots would come into existence in my lifetime…. Yet here I am, forty-three years after I wrote my first robot story, and we do have robots. What's more, they are what I envisioned them to be in a way… To be sure, these robots are not as intelligent as my robots are—they are not positronic: they are not even humanoid. However, they are evolving rapidly and becoming steadily more capable and versatile. Who knows where they'll be in another forty years?
>
> One thing we can be sure of. Robots are changing the world and driving it in directions we cannot clearly foresee.[84]

His hubris is not without merit. Asimov's vision of the robot is the predominant one in contemporary society, even inspiring the direction of the robotics industry in the latter half of the twentieth century. For example, Robotics pioneer Joseph F. Engelberger (b. 1925), founder of the premiere US robotics firm

[84] Ibid., 10-11.

Unimation Inc. in 1956, credited his early interest in the development of "robots-in-reality" to reading Asimov's robot stories[85]—and he is not the only one.

Burroughs, Williamson, and Asimov—as well as many other authors of the Golden Age of Science Fiction—were incredibly influential in their field at the time, and have helped to shape not only the development of SF as a genre but the very way in which technology has developed since the mid-twentieth century. However, we must remember that SF during the forties and fifties was a *subcultural* phenomenon whose main demographic was the adolescent male. While the robotic subaltern existed in SF films and television shows up to this point—with Maria of *Metropolis* (1927) and Robby the Robot in *Forbidden Planet* (1956) as two iconic examples—science fiction itself commanded only a small fraction of the film and television industry of the time. For the mainstream, the mechanical man of the early twentieth century was a curiosity and an amusement, as was the case with Westinghouse's ELEKTRO, discussed in the next section. With the launching of *Sputnik I* by the Soviet Union in 1957, however, mankind's troubled relationship with technology exploded again into mainstream Western consciousness in the form of the Space Race. Suddenly, technological superiority was tied to ideological superiority, as the US and the Soviet Union competed for command of the heavens. The technological explosion of the period served to reinscribe the ambivalent nature of the man-machine relationship in society, making the threat much more immediate: according to J.P. Telotte, "the notion of human artifice became less a distant prospect or alien threat… and more a discomfiting feature of our cultural landscape."[86] This simultaneous fascination with and fear of the artificial human has only grown as this "distant prospect" looms ever closer in our everyday technology.

---

[85] Ibid., 11.
[86] Telotte, *Replications*, 131.

WE WELCOME OUR ROBOT OVERLORDS

In 2011, well-known game-show *Jeopardy!* hosted a truly groundbreaking event: for the first time in game-show history, human contestants were pitted against a supercomputer, whimsically named Watson. Watson competed in an invitational tournament against the two highest-ranking players to date: Ken Jennings, who with a 74-game streak and over $3.8 million in prize money currently holds the world record for game show winnings, and Brad Rutter, who defeated Jennings in the *Jeopardy!* Ultimate Tournament of Champions and is second place overall. Watson soundly trounced both Jennings and Rutter, with a final score of $77,147 for Watson, $24,000 for Jennings and $21,600 for Rutter. During the Final Jeopardy round, Jennings playfully admitted defeat by writing on his answer screen "I, for one, welcome our new computer overlords."[87] While Jennings' comment in Final Jeopardy was made in jest, it reverberates with the image of Čapek's Robots marching across the stage in 1920s Prague, intent on destroying the whole human race.

This "story of humanity's conflict with machinery for mastery of the world"[88] represents the oldest of what Minsoo Kang identifies as "three contending views on the fate of humanity in the age of digital technology in which the robot is often used for conceptual and imaginative purposes"[89]: the *inevitable confrontation*, or what Asimov dubbed the Robot-as-Menace class of science fiction stories.[90] This was the typical outcome of the robot-themed stories of the Golden Age of Science Fiction; combining an innate fear of the Other with the inexorable rise of the Machine during the Industrial Revolution and the interwar period, this fatalistic scenario "has been told so many times in science

[87] Jennings here draws on a popular internet meme, derived from an episode of *The Simpsons* ("Deep Space Homer," Season Five, 1994), in which anchorman Kent Brockman assumes the Earth is being invaded by giant space ants, and announces his willingness to submit: "One thing is for certain: there is no stopping them; the ants will soon be here. And I for one welcome our new insect overlords. I'd like to remind them that as a trusted TV personality, I can be helpful in rounding up others to toil in their underground sugar caves."
[88] Kang, *Sublime Dreams*, 300.
[89] Ibid..
[90] Asimov, *The Complete Robot*, 9.

Figure 35. Still from Jeopardy! *in which Ken Jennings playfully admits defeat. "The IBM Challenge: Day 3,"* Jeopardy! *Season 27, episode 113, directed by Kevin McCarthy, aired 16 February 2011. YouTube Video, <http://www.youtube.com/watch?v=o6oS64Bpx0g>, accessed 10 August 2014.*

fiction literature and cinema that it is a veritable cliché of our time."[91] Sometimes the machines win. In *R.U.R.*, this means the destruction of the human race; in *The Matrix* (1999), for example, the outcome is much more nefarious: we become the tools of the machines, human batteries powering their electronic society (which of course leaves the storyline open for a successful turnabout revolt in the sequels).

Why are we afraid of robots? Michael Anderson explores this fear in his review of several recent works on artificial intelligence, "Why is AI so Scary?" moving beyond Hollywood sensationalism to the moral and social implications of intelligent machines. Anderson identifies two basic conflicts created by the rise of the intelligent machine: first, that "our machines reflect who we are and what we believe"[92]—that is, robots are ultimately reflections of our own fallibility, reinscribing hegemonic notions of identity in noncritical ways; and second, that "as people imagine themselves interacting with intelligent machines,… they cannot help *but* think of them as living things, with many of the ethical

[91] Kang, *Sublime Dreams*, 300.
[92] Michael L. Anderson, "Why Is AI So Scary?," *Artificial Intelligence* 169, no. 2 (2005): 204.

obligations this implies"[93]—in other words, if robots act like humans, we'll treat them like humans, which puts us in a moral dilemma even more profound than the one Steven Wise engenders with his quest for the recognition of some species of animals as nonhuman persons. At its most basic level, these ambivalent attitudes toward intelligent machines center on the idea of *sentience*—that is, the idea that a robot, more removed from our human reality than any of the controversial (yet at least *living*) examples Wise offers, could be considered "self-aware," with its own thoughts and feelings. This fear, seen in the "Robot-as-Menace" class of science fiction stories, is likewise reflected in our daily lives.

Inevitable confrontation as played out in our science fiction results in a creeping paranoia regarding any sufficiently advanced technology: the development of autonomous, weaponized drones by the military, for example, immediately raises for some the spectre of the Robot Revolution as predicted by Čapek nearly a century ago. On the other hand, the research into "caregiver" robots in the fields of medicine and childcare brings with it images from Williamson's *Humanoids* of mechanical nannies who slowly erode our liberties, leaving us all like Ruth Forester in the novel, babbling away on the floor, playing with blocks, the essence of our own humanity erased "for our own good." But for all its excitement, the view that the relationship between man and robot is one of inevitable confrontation contributes very little to my exploration of *Homo artificialis*, for in actuality these fears are for the most part about man's conflict with the Other—the difference being that the new Other is mechanical. I am more interested in exactly *how* this Other is constructed than with the binary conflict in engenders.

This confrontation is not as inevitable as some may believe. At the other end of the spectrum Kang identifies the theory of *cybernetic mergence*, a thread developed largely since the 1980s that imagines a fusion of man and machine "to the extent that the very distinction between what is natural and artificial, organic and mechanical, biological and digital, will cease to have any meaning."[94] This is the vision of the *transhumanist* movement, which advocates for a cyborg future through such means as

[93] Ibid., 205.
[94] Kang, *Sublime Dreams*, 302.

nanotechnology, cybernetic implants, and *participant evolution*—using technology to deliberately redesign the human body and brain. Some argue that this process is already happening: cochlear implants for the deaf, for example, could be seen as a step in this direction. The twenty-first century worldview has exploded with the thought of a cybernetic future, whether expressed in terms of science fiction fears—such as the threat of the Borg in *Star Trek*—or predictions of scientific possibility—such as transhumanist advocate Ray Kurzweil's expounding on the coming "singularity."[95]

Cybernetic mergence is a hotly debated concept in scholarly circles as well. In the early 1990s Donna Haraway published her landmark "Cyborg Manifesto," in which she claims the rise of the cyborg heralds an erasure the dualistic boundaries imposed by Descartes.[96] N. Katherine Hayles also took up the topic in her 1999 work *How We Became Posthuman*, in which she argues the case for embodiment in an increasingly virtual world.[97] In *Our Posthuman Future*, Francis Fukuyama warns of the dangers inherent in a cybernetic future, lamenting that *posthuman* is what we will become after our basic human nature is bred out of us through participant evolution.[98] While many of the critical concepts arising from the study of the cyborg come into play in realizing the potential of the artificial human, cybernetic mergence as such is only tangential to my discussion of *Homo artificialis*, as inherent to the common definitions of *post*human is the initial assumption of humanity and personhood. Rather than asking the question "What would it take to grant an entity personhood?" posthumanism asks instead "What would it take to no longer define an entity as human?"—a much more troubling question, as it is far more difficult a proposition to take away rights than to grant them. However, some of the elements of "basic human

---

[95] The *singularity* hypothesis was first described in 1958 by John von Neumann as 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" (Stanislaw Ulam, "Tribute to John von Neumann," *Bulletin of the American Mathematical Society* 64, #3, part 2 [May 1958]: 5). The idea was made popular by SF author Vernor Vinge, who predicts this could happen around 2030, while Kurzweil predicts 2045. See Kurzweil, *The Singularity Is Near*, for example, or *The Age of Intelligent Machines* (Cambridge, MA: MIT Press, 1990).

[96] Donna J. Haraway, *Simians, Cyborgs, and Women: The Reinvention of Nature* (New York: Routledge, 1991), 177-78.

[97] N. Katherine Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics* (Chicago: University of Chicago Press, 1999).

[98] Francis Fukuyama, *Our Posthuman Future: Consequences of the Biotechnology Revolution*, 1st ed. (New York: Farrar, Straus and Giroux, 2002).

nature" the posthumanists fear losing in the cybernetic future may be just those characteristics we see emerging in *Homo artificialis*—a prospect we will return to later.

Kang identifies a third position on the human-machine question—that of *equivalence through sentience*—and while he considers it a lesser position to the more likely one of cybernetic mergence, this concept is in fact the most pertinent for the purposes of my own study. According to Kang, equivalence through sentience

> asserts that once machines do reach the human level of consciousness, they should indeed be treated as living, thinking beings and that granting them status as such, including even political and legal rights, might be the best way to prevent a catastrophic confrontation with them. This idea affirms the difference between natural and artificial creatures, human and digital intelligence, but it also asserts that the question of active conscience, not its nature or origin, should be the essential factor in judging who or what should be accorded the dignity due a sentient being.[99]

Unlike the theory of inevitable confrontation, in which the focus is on conflict and destruction, or cybernetic mergence, in which technology becomes an extension or an aspect of an already-human subject, the theory of equivalence through sentience directly addresses the question at hand, as paraphrased from Steven Wise of the NhRP: *What is a* sufficient *condition for being accepted as a sentient being (and thus, a* person*)? Not a necessary condition, but a* sufficient *condition?* Unlike the mechanical figures of the eighteenth century, which relied on pre-programmed routines, possessed only a limited autonomy and had no claim to sentience at all, modern artificial humans are seen as *interactive* machines, designed by their creators to respond and adapt to their environments in real time, necessitating the kinds of behavior and decision-making capabilities that have traditionally been

[99] Kang, *Sublime Dreams*, 301-02.

the exclusive domain of living creatures. As the artificial human behaves more and more like its biological counterpart, the question of machine sentience becomes more relevant.

FROM ELEKTRO TO WATSON: ARTIFICIAL HUMANS OF THE TWENTIETH CENTURY

By the mid-twentieth century, the modern image of the robot as a mechanical artificial human that operates autonomously was firmly entrenched in our consciousness, reinforced by its image in the science fiction of the time. However, the "robots-in-reality" (to borrow Asimov's phrase) of the mid-twentieth century lagged far behind their fictional counterparts in terms of technological capability. The humanoid robots built during the first half of the twentieth century hardly fit the term at all, having the outward appearance of the Tin Man but lacking many of the features we now accept as typical. Indeed, it was not until the end of the twentieth century that mechanical artificial humans evolved past the automaton phase exemplified by the works of Vaucanson and Jaquet-Droz. For most of the twentieth century, the prospect of the humanoid robot was largely displaced by the more practical (and profitable) development of non-humanoid robots for industrial purposes. However, with the exponential rate of development in the past century in diverse areas such as computing, electronics, AI and A-Life, cognitive psychology, behavioral science, and robotics proper, the pieces of the robot puzzle have begun to come together, giving rise to the first generation of "robots-in-reality" that may prove capable of fulfilling Turing's prediction: "one will be able to speak of machines thinking without expecting to be contradicted."[100]

While the creation of an artificial human is typically the work of one misunderstood genius in science fiction, in reality the development of humanoid robots has involved a century of effort by many different research groups, each focused on one piece of the puzzle. To operate successfully as an artificial human involves interacting with human beings as depicted in science fiction—that is, interacting within everyday settings and social environments. In practical terms, this poses a number of

[100] Copeland and Turing, *The Essential Turing*, 449.

different challenges addressed by roboticists of the twentieth century, including the manipulation of objects, human-like locomotion, sensing and reacting to the environment, and understanding and responding to voice commands. While this list is by no means complete, and many aspects of these "separate" challenges are highly dependent on one another (successful movement through an environment necessitates the ability to sense and react to that environment, for example), during much of the twentieth century they were largely treated as separate fields of research. Over the course of the twentieth century, research teams worldwide have focused on developing the many different aspects of robotics and AI that, put together, culminate in the advanced humanoid robots seen in the twenty-first.

One of the earliest humanoid robots on display was the Gomshall Robot, which opened the 1928 Model Engineering Exhibition in Gomshall England and was dubbed by show organizers the "First British Robot."[101] Influenced by the popularity of Capek's play, the robot was stamped with "RUR" across its chest. The resemblance to the robots of the play ended there, however; the figure was more automaton than robot, able to sit and stand in place and equipped with a sound system through which a human operator could hear and answer questions. A decade later the Westinghouse Electric Company unveiled its own marvel of electromechanical engineering, ELEKTRO, at the 1939 World's Fair in New York. Standing seven feet tall, ELEKTRO was mobile: he shuffled through the exhibit hall talking to visitors, accompanied by his canine companion SPARKO, both pulling bundles of electrical cables in their wake.[102] Designed to showcase Westinghouse's new telephone control system, ELEKTRO was able to execute a number of voice-activated commands, controlled by a human operator through a telephone handset connected to his chest. But while these robots attracted the interest of thousands of exhibition attendees, they were essentially electromechanical automata—giving only the appearance of autonomy while remaining the robotic puppets of their handlers.

---

[101] Lisa Nocks, *The Robot: The Life Story of a Technology* (Baltimore, MD: Johns Hopkins University Press, 2008), 56-67.
[102] Ibid., 56-58. See also David Zondy, "Tales of Future Past: ELEKTRO," accessed 18 October 2014, http://davidszondy.com/future/robot/elektro1.htm for more about ELEKTRO's capabilities and technical specifications, including a rather fanciful biography and a video of his appearance at the 1939 World's Fair.

Figure 36. Exhibit at the Mansfield Memorial Museum showing three versions of the Westinghouse robot, from left to right: Herbert Televox (1928), the first human form robot that Westinghouse built; reconstruction of Elektro as he looked in 1939; and the original Elektro, modified in 1957. Photo courtesy of and © Scott Schaut, Curator for the Mansfield Memorial Museum.

The first robots able to claim true (although limited) autonomy were the industrial robots of the mid-twentieth century: machines that could operate efficiently for long periods by self-correcting functional irregularities through a system of continual feedback and control mechanisms. This type of self-regulating control system was made possible through the research of Norbert Wiener (1894–1964), who used the natural feedback mechanisms of the mammalian nervous system as a guiding paradigm, thus launching the field of *cybernetics*: the use of organic systems as the model for the operation of mechanical devices.[103] This presents a reversal of the mechanism of the seventeenth century: rather than using machines to model living systems in order to better understand said systems, cybernetics

[103] Norbert Wiener, *Cybernetics: Or the Science of Control and Communication in the Animal and the Machine* (New York: J. Wiley, 1948).

draws upon nature as a model through which to make machines more efficient, more robust, and more adaptable to a variety of applications. Vaucanson created the *Digesting Duck* to show that the biological process of digestion could be explained mechanically; Wiener drew upon living feedback systems in order to improve the performance of automated machines. The cybernetic principle is central to most areas of research in the fields of robotics, computer science, and artificial intelligence today.

The first successful industrial robot was manufactured in the late 1950s by the US firm Unimation Incorporated. Engineer and entrepreneur Joseph Engelberger, inspired by the robot stories of his youth, founded Unimation in 1956 in order to fund production of the technology designed by fellow science fiction enthusiast and inventor George Devol, Jr. (1912–2011). Devol's invention—the Unimate, short for *universal automation*—represented the first digitally operated programmable robotic arm, which Devol claimed in his patent "makes available for the first time a more or less general purpose machine that has universal application to a vast diversity of applications where cyclic digital control is desired."[104] The first Unimate was installed in the General Motors plant in Danbury, CT in 1961, where it was used for die casting and spot welding, and is housed today in the industrial history wing of the Smithsonian Institution.

General Motors was initially hesitant about the acquisition, as it was highly experimental, reflecting the typical concern of many US firms at the time; according to Lisa Nocks, "in the 1960s, American companies were more concerned with making quick profits, and could not get beyond thinking of robots as a passing, entertaining fad."[105] It did not help that Engelberger took the Unimate on *The Tonight Show with Johnny Carson* in 1966 to promote his company; after the robotic arm poured beer and hit golf balls, to the amusement of Johnny and his guests, many US companies wanted to rent the robot arm merely "as an entertainment device."[106] The Japanese, however, recognized the potential

[104] George C Devol, Jr, Programmed Article Transfer, US Patent 2988237, filed 10 Dec 1954 and issued 13 June 1961.
[105] Nocks, *The Robot*, 67.
[106] Ibid..

robotics offered in a wide variety of production and service industries, and by 1968 Unimate had licensed its robots to Kawasaki Industries for production in the Asian market—one of the reasons Japan dominates the global robotics industry today.

Advances in the technology of industrial robots over the next decades led to their usefulness in areas outside their commercial application. In 1974, David Silver of MIT designed The Silver Arm, capable of fine movements and sensitive feedback that replicated human dexterity. Two years later Shigeo Hirose of the Tokyo Institute of Technology created the Soft Gripper, which could conform to the shape of a grasped object, such as a wine glass filled with flowers; his design stems from studies of flexible structures in nature, such as elephant trunks and snake spinal cords. In 1981 Japanese computer scientist Takeo Kanade created the first "direct drive arm" in which the servomotors controlling the arm's movement were located within the arm itself. By the turn of the century, the technology was sufficiently advanced to allow its adaptation in to medical applications: the first robotic prosthetic arms, with amazingly humanlike appearance, dexterity, and degrees of freedom. Recently, in a 2013 TED talk,[107] biomedical engineer Todd Kuiken unveiled a prosthetic arm that connects with the human nervous system, offering neural control and sensory feedback to its user.

While ELEKTRO was indeed mobile, slowly shuffling along at the World's Fair in 1939, he was little more than a twentieth-century incarnation of the mechanical theatres of ancient times: moving through pre-determined pathways based on audio commands. But successful motion appropriate for a humanoid robot poses a number of issues, as walking is a difficult and dynamic problem to solve. Most teams researching the problem of a walking robot do so with designs incorporating four or more legs, as this greatly increases stability and balance—such as the *BigDog*, a four-legged rough-terrain robot designed in 2008 by Boston Dynamics. Not until the end of the twentieth century were the first robots developed that could walk reliably on two legs: the ASIMO series from Honda is the most famous

[107] Todd Kuiken, "A Prosthetic Arm That 'Feels,'" filmed October 2011 at TEDGlobal 2011, accessed 06 April 2014, https://www.ted.com/talks/todd_kuiken_a_prosthetic_arm_that_feels.

example.[108] First introduced in 2000, ASIMO resembles a spacesuit-clad astronaut. While his initial movements were slow and labored—videos of earlier versions depict ASIMO carefully climbing stairs—Honda boasts that the most recent ASIMO can "run, walk on uneven slopes and surfaces, turn smoothly, climb stairs, and reach for and grasp objects. … ASIMO can also avoid moving obstacles as it navigates its environment."[109] Yet for all ASIMO's capabilities, no bipedal walking robot has yet been constructed that can move as quickly and gracefully as a human.

While robots like ASIMO represent advances in the imitation of *physical* human characteristics, other groups have been working toward the replication of human *mental* capabilities. One such recent example is Watson, the *Jeopardy!* champion mentioned earlier. After their success in 1997 with Deep Blue, IBM turned to the challenge of natural language processing (NLP) in machine intelligence. The

Figure 37. BigDog military robots trot around in the shadow of an MV-22 Osprey. U.S. Marine Corps photo (2006) by Lance Cpl. M. L. Meier. Public domain.

[108] Honda engineers began their research into human-like locomotion in 1986.
[109] See American Honda Motor Company, "ASIMO: the world's most advanced humanoid robot," accessed 8 April 2014, http://asimo.honda.com/.

Figure 38. ASIMO demonstrating his stair-walking capabilities at the 2005 Consumer Electronics Show, Las Vegas, NV. Photo from Wikimedia Commons.

connection between language and meaning is often ambiguous, as anyone who has ever studied a foreign language knows from experience. Take the following pair of sentences, for example:

Time flies like an arrow.

Fruit flies like a banana.

This play on words is over a century old, but any English-speaking human reading the two sentences immediately understands the difference in meaning and usage between the two; the understanding is innate, a result of our life-long immersion in the use of language to communicate meaning between individuals.[110] A typical computer, however, must be programmed not only with the definitions of each word in each sentence, but to further distinguish between the use of *flies* as the third-person present

[110] Contrast the theories of B.F. Skinner, who claims that language is a behavior that may be learned through conditioning and is not limited to humans (B. F. Skinner, *Verbal Behavior*, The Century Psychology Series [New York: Appleton-Century-Crofts, 1957]) with those of Noam Chomsky, who holds that the structure of language is a uniquely human phenomenon that is biologically determined and genetically transmitted.

form of the verb *to fly* in the first sentence and as the plural of the noun *fly* in the second.[111] The ability to draw meaning from context within a given text (written or oral) involves being able to successfully navigate this ambiguity.

IBM began the Watson project in 2004 after watching Ken Jennings during his 74-game *Jeopardy!* run. Watson's Jeopardy*!* performance could be considered a highly limited version of a Turing Test: rather than an open-ended conversation, the game format involves textual "clues" for which contestants must provide an appropriate response in the form of a question. *Jeopardy!* clues rely on puns, wordplay, and other syntactical ambiguity of the type illustrated in the sentence pair above; they additionally draw on detailed information from a wide variety of subjects. A successful performance on *Jeopardy!* poses a challenge for most human contestants; that Watson was able to hold his own is a testament to David Ferrucci and his colleagues, who refer to Watson as a *cognitive technology*:

> Watson is a cognitive technology that processes information more like a human than a computer—by understanding natural language, generating hypotheses based on evidence, and learning as it goes. And learn it does. Watson "gets smarter" in three ways: by being taught by its users, by learning from prior interactions, and by being presented with new information. This means organizations can more fully understand and use the data that surrounds them, and use that data to make better decisions.[112]

Much of what makes Watson so robust it his capacity for *learning*. Based on prior successes and failures, Watson was able to make predictions on the likelihood of whether a given response was the correct one, and then choose a response to offer based on that probability. The televised episode offered the viewers an on-screen display of Watson's top response choices and the probability assigned to each one as Watson arrived at an answer (see Figure 39).

---

[111] There are a number of other linguistic differences in play here; I offer *flies* only as an example of the complexity.
[112] IBM, "Watson Technology," accessed 8 April 2014, http://www.ibm.com/smarterplanet/us/en/ibmwatson/tech.html.

Despite winning the *Jeopardy!* tournament handily, Watson arguably failed the Turing Test it represents, as some of his choices were decidedly un-humanlike. Watson operated completely autonomously during each round, deciding on his own (based on his current standing) which categories to choose and how much money to wager on Daily Doubles. However, his chosen wagers were very unlike those of his human counterparts: contestants typically wager in dollar amounts rounded to the nearest hundred; Watson offered figures such as $1,246. In addition, while most of his answers were correct, the incorrect answers were often *inhumanly* incongruous. For example, in the first Final Jeopardy round, Watson offered the answer "What is Toronto?" to a question in the category "US Cities"—a mistake an average *Jeopardy!* contestant is unlikely to make. While it is not outside the realm of possibility for a human contestant to do either of these things, they are atypical; several such incongruities over the course of the tournament serve to reveal the nonhuman nature of the contestant.

Figure 39. Still from Jeopardy! *showing the home viewing audience a numeric representation of the algorithm by which Watson chooses appropriate answers. "The IBM Challenge: Day 1,"* Jeopardy! *Season 27, episode 111, directed by Kevin McCarthy, aired 14 February 2011. YouTube Video, <http://www.youtube.com/watch?v=qpKoIfTukrA>, accessed 10 August 2014.*

Watson is one of the most advanced examples of natural language processing in the industry today, and IBM is currently bringing that technology to bear in applications for the healthcare, finance, marketing and service industries. Yet despite IBM's success, Watson is still only a more-advanced version of narrow AI. The version of Watson that won *Jeopardy!* would fail outside the limited scope of the game show's format. But even if a version of Watson were developed that could ultimately win the Loebner Gold Medal, it would fall short of the mental faculties required for the *Homo artificialis*. In addition to the intelligence of Watson, and the physical capabilities of ASIMO, there is one more piece to the puzzle: since humans are *social* creatures, artificial humans must be social as well.

SOCIAL ROBOTICS

Throughout the development of autonomous robots in the mid-twentieth century, interaction with humans (human-robot interaction, or HRI) was kept to a minimum—a person was typically viewed an obstacle in the robot's environment, to be avoided rather than addressed, and HRI more often than not resulted in either damage to the robot or injury to the human. During the past few decades, however, there has been an explosion of robotics applications designed for the explicit purpose of engaging humans in a social way: robotic toys, museum tour guide robots, food service robots, and more. While a comprehensive survey of the developments in social robotics in the past twenty years is well beyond the scope of this study, two examples to illustrate the scope of this emerging field. First is the work of Cynthia Breazeal, director of the Personal Robotics Group at MIT, who is recognized as a pioneer in social robotics and human-robot interaction.[113] At the other end of the spectrum is Hiroshi Ishiguro at the Intelligent Robotics Laboratory of Osaka University, whose humanoid robots boast very lifelike appearance and behaviors.[114] Both Breazeal and Ishiguro take the human as the model upon

[113] MIT Media Lab, "Personal Robots Group," accessed 11 May 2014, http://robotic.media.mit.edu/.
[114] Advanced Telecommunications Research Institute International (ATR) and Hiroshi Ishiguro, "Geminoid," accessed 15 May 2014, http://www.geminoid.jp/index.html.

which to develop their concept of the artificial human; however, while Ishiguro approaches the problem from a *physical* point of view, Breazeal addresses the *mental* side of the equation.

Cynthia Breazeal identifies four interaction paradigms in the field of HRI,[115] derived from the human's perspective of the interaction: The first is *robot as tool*, such as the bomb-sniffing robots used in Iraq, with very little consideration for "social" interaction. The second paradigm is *robot as cyborg extension*, in which the robot is an extension of the human using it, such as a robotic exoskeleton. The third paradigm, *robot as avatar*, involves the robot as an agent of telepresence for another human at a remote location. This is the focus of Ishiguro's lab: his "tele-operated androids" offer the most lifelike appearance of any humanoid robots at present. However, while co-present humans may interact with Ishiguro's androids socially, the androids are understood merely as communication devices through which interaction with a remote human is possible, not agents in their own right. That is reserved for the final paradigm—*robot as sociable partner*—a socially responsive artificial being able to cooperate with her human partner. It is this final paradigm that is the primary goal of Breazeal and the Personal Robot Group.

Breazeal is regarded by many as the founder of the field of social robotics; her doctoral work (1993) with the robot Kismet is recognized as one of the first explorations of the social and emotional aspects of human-robot interaction. Since the early 1990s, Breazeal and her colleagues at the MIT Media Lab Personal Robotics Group have been developing what she terms "sociable robots," intended to work autonomously with humans on tasks, as collaborative assistants rather than as tools or extensions of the human. Breazeal further delimits her research, focusing on true sociable robots rather than those that are merely socially evocative (such as robotic pets or toys) or have a social interface, such as the museum tour robot. The robots of the MIT Media lab are "socially participative 'creatures' with their own internal goals and motivations. They pro-actively engage people in a social manner not

[115] Cynthia Breazeal, "Social Interactions in HRI: The Robot View," *IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews* 34, no. 2 (2004): 181.

only to benefit the person, but also to benefit itself (to promote its survival, to improve its own performance, to learn from the human, etc.).”[116]

The Personal Robots Group at the MIT Media Lab focuses specifically on the development of robots that are autonomous, humanoid, and socially interactive, basing their work in the imitation of socially-appropriate human behavior and the idea of *social intelligence*. In 2007 Kersten Dautenhahn coined the term *robotiquette* for the set of socially-determined guidelines necessary for an artificial agent to function efficiently in a human environment.[117] Dautenhahn’s robotiquette emerges from her insistence that social intelligence is a necessary component of any artificially intelligent robot. Despite the identification of a number of theoretical objections to the treatment of a robot as worthy of social interaction,[118] it is not typically a problem for the human subjects involved, given the human tendency to “generally treat computers as they might treat other people provided that the technology behaves in a socially competent manner.”[119] Often termed “The ELIZA effect” after the 1960s chatbot of the same name, this tendency is a by-product of our evolution:

> the human brain evolved in a world in which only humans exhibited rich social behaviors, and a world in which all perceived objects were real physical objects. Anything that seemed to be a real person or place was real…. [It] actually takes more effort for people to consciously inhibit their social machinery in order to not treat the machine in this way.[120]

---

[116] “Toward Sociable Robots,” *Robotics and Autonomous Systems* 42, no. 3-4 (2003): 169.
[117] Kersten Dautenhahn, “Socially Intelligent Robots: Dimensions of Human-Robot Interaction,” *Philosophical Transactions of the Royal Society B-Biological Sciences* 362, no. 1480 (2007): 679.
[118] Anderson, “Why Is AI So Scary?”; Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato, “Measurement of Negative Attitudes toward Robots,” Interaction Studies 7, no. 3 (2006): 437-54; Steve Torrance, “Ethics and Consciousness in Artificial Agents,” AI and Society 22, no. 4 (Apr 2008): 495-521; Astrid Rosenthal-von der Pütten, Nicole Krämer, Christian Becker-Asano, Kohei Ogawa, Shuichi Nishio, and Hiroshi Ishiguro, “The Uncanny in the Wild: Analysis of Unscripted Human-Android Interaction in the Field,” International Journal of Social Robotics 6, no. 1 (Jan 2014): 67.
[119] Breazeal, “Social Interactions in HRI,” 182.
[120] Cynthia L. Breazeal, *Designing Sociable Robots*, Intelligent Robots and Autonomous Agents (Cambridge, MA: MIT Press, 2002), 15.

Breazeal takes this stance as the basis of her research: "we take as a working assumption that attempts to foster human-robot relationships will be accepted by a majority of people *if* the robot displays rich social behavior,"[121] encouraging human interaction partners to treat the robots as equivalent to humans in these interactions in order to study not only how to adapt the robot into the environment, but how to best prepare human participants to interact with their robotic counterparts.

In order for social robots to be successful as collaborators in a real-world situation, they must require no more than minimal training and adaptation on the part of their human counterparts. Breazeal's research at the MIT Media Lab addresses the question of "how to properly interface untrained humans with these sophisticated technologies in a manner that is intuitive, efficient, and enjoyable to use."[122] The research goals of the Media Lab reflect this:

> Inspired by animal and human behavior, our goal is to build capable robotic creatures with a "living" presence, and to gain a better understanding of how humans will interact with this new kind of technology. People will physically interact with them, communicate with them, understand them, and teach them, all in familiar human terms. Ultimately, such robots will possess the social savvy, physical adeptness, and everyday common sense to partake in people's daily lives in useful and rewarding ways.[123]

Because humans have a natural tendency to anthropomorphize interactive agents (whether pets, cars, or computers), researchers in human-computer interaction "argue that a social interface may be a truly universal interface" for the control and direction of sophisticated technology, since it takes advantage of the pre-existing human competency for social interaction rather than a learned set of commands or a physical interface. Modeled on similar human-human interactions, such as apprenticeship or other

---

[121] Breazeal, "Social Interactions in HRI," 182.
[122] Ibid.
[123] MIT Media Lab,"Personal Robots," accessed 8 April 2014, http://media.mit.edu/research/groups/personal-robots.

training, Breazeal posits that this social command system is more "natural" for the humans involved in dealing with humanoid robots.

But what is "natural" interaction? What criteria are used to determine the set of social performatives programmed into one of these social robots? Like many HRI researchers, the MIT Media Lab approaches the problem of "proper" social interaction from a practical standpoint rather than a critical one; that is, they address *how* to create suitable interaction without concerning themselves with the deeper question of *why* these particular interactions are suitable to begin with. Breazeal discusses the robots' "modeling" of humans based on research in cognitive psychology: "Social interactions with people are valued not just at the interface, but at a pragmatic and functional level as well…. [T]he robot's social behavior is a product of its computational social 'psychology.'"[124] Elsewhere, Breazeal discusses the inspiration for her robots' social behavior in the biological sciences: "Our implementation of basic emotive responses takes its inspiration from ethological models for the organization of behavior in animals, and how this relates to the concept of drives and triggering events (i.e. internal releasing mechanisms)."[125] Breazeal's robots are programmed to read and respond to human emotion in a socially appropriate manner—simply put, they are programmed with a performance of humanity.

So what are these behavioral performatives that qualify a machine agent as 'performing humanity'? The question depends on the robot, actually. For Kismet, Breazeal claims that she took as her starting point the work of Kerstin Dautenhahn on "social intelligence" in robots. These behaviors come from the field of cognitive psychology, and are designed to encourage the human interacting with the robot to treat it, if not as a "true" human, at least as an autonomous, human-like being capable of social interaction—and more importantly to Breazeal, of learning. In her book *Designing Social Robots*,

[124] Breazeal, "Toward Sociable Robots," 169.
[125] Breazeal, "Emotion and Sociable Humanoid Robots," *International Journal of Human-Computer Studies* 59, no. 1-2 (2003): 148.

Breazeal identifies a number of characteristics she believes are necessary for successful social interaction:[126]

- **Embodiment:** co-existence in space and time is a necessary condition for appropriate social interaction of the kind desired.
- **Life-Like Quality:** humans have a strong tendency to anthropomorphize any agent that appears to be autonomous, but the "illusion of life" must be believable and sustainable.
- **Human-Aware:** the agent must be able to perceive and understand human social cues.
- **Being Understood:** conversely, the agent must also be able to give social cues to its human interactor.
- **Socially Situated Learning:** Finally, the agent must be able to adapt to new experiences and new situations by learning from its social interactions.

Given these criteria, Breazeal identifies Kismet as incorporating the social development and learning capabilities of a human infant.

Kismet could be considered "humanoid" only in the broadest sense of the term. As Breazeal's focus was on social interaction through the reading and display of emotion, Kismet consists of little more than eyes (with lids and brows), ears, and mouth. He is designed to imitate emotion through facial expression, with over twenty *degrees of freedom*[127] in various facial components. Yet even with these minimal features, Kismet is able to engage his human partner in compelling social interaction.[128] Kismet's second-generation cousin, Leonardo, is even less humanoid: built in collaboration with the animatronic artists at Stan Winston Studio, Leonardo bears a strong resemblance to the *mogwai* of the

[126] Summarized from Breazeal, *Designing Sociable Robots*, 6-11.
[127] *Degrees of freedom* in mechanical systems are number of independent motions in three-dimensional space that are allowed to various components within the system. A full range of motion in a single component consists of six degrees of freedom (6DoF): up/down, left/right, back/forward, pitch, yaw, and roll.
[128] Breazeal, *Designing Sociable Robots*, 154-55.

1984 film *Gremlins*, a broad-faced, furry creature with large ears and expressive eyes. The fanciful appearance of Kismet and Leonardo is deliberate:

> Unlike the vast majority of autonomous robots today, Leonardo has an organic appearance. It is a fanciful creature, clearly not trying to mimic any living creature today. This follows from our philosophy that robots are not and will never be dogs, cats, humans, etc. so there is no need to make them look as such. Rather, robots will be their own kind of creature and should be accepted, measured, and valued on those terms. We gave Leonardo a youthful appearance to encourage people to playfully interact with it much as one might with a young child.[129]

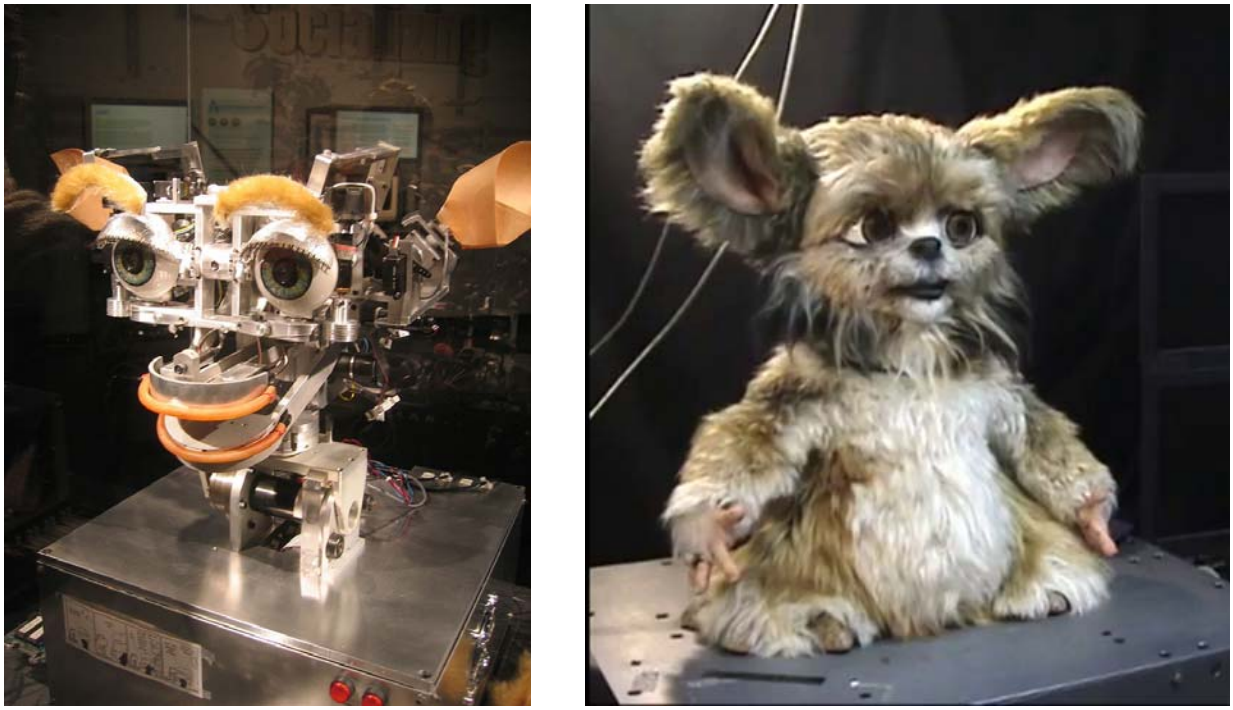


Figure 40. Cynthia Breazeal's robots from the MIT Media Lab. On the left is Kismet, developed in 1997 to read and reflect emotive facial behaviors. Photo © Jared C. Benedict, from Wikimedia Commons. On the right is Leonardo, a fantastical creature designed to facilitate playful interaction. Photo still from Cynthia Breazeal, "The Rise of Personal Robots," TEDWomen 2010. Licensed through Creative Commons 3.0.

[129] Breazeal, "Research: Leonardo," accessed 9 April 2014, http://web.media.mit.edu/~cynthiab/research/research.html.

Breazeal has made a conscious choice to focus on the imitation of human *emotion*, not the human form itself. Even her most recent effort, the MDS series, is only minimally humanoid, consisting of a sleek white head, torso, and arms roughly the size of a three year old child. With the MDS, the focus is on interaction and social learning; the robot's small stature allows the operation of multiple units within a typical laboratory environment in order to "develop social robots that can successfully learn what matters to the average citizen from the kinds of interactions that people naturally offer."[130]

Drawing from the same initial premise as Breazeal—that the human brain has evolved to recognize as human anything that strongly presents itself as such—Hiroshi Ishiguro of Osaka University in Japan has reached quite a different conclusion about the implications of this premise upon robotics research. While Breazeal's research focuses on the performative aspects of human social behavior, Ishiguro believes that human-robot interaction is best served through a strong humanlike appearance. As the Director of Osaka University's Advanced Telecommunications Research Institute International (ATR), Ishiguro studies the idea of human presence through the development of humanoid robots. Unlike Kismet, Leonardo, and the MDS robots of the MIT Media Lab, Ishiguro's robots are not autonomous. Rather, they serve as *telepresent avatars* of their remote human operators. The goal of Ishiguro's work with the these robotic avatars is to see how human interactants evaluate the *presence* of the robot—according to Ishiguro, the degree to which the human interactant accepts that the third-person operator is present in the room.

Ishiguro's telepresent robots represent two extremes of humanlike appearance: at one extreme is the *Geminoid*, a series of telepresent robotic clones of Ishiguro (the HI- series) and his colleagues (the F- series and the Geminoid DK). The Geminoid robots are designed to appear as human-like as possible, possessing foam flesh and silicone skin over plastic skulls and metal skeletons, with up to fifty degrees of freedom in the head and body, allowing for uncannily accurate facial expressions and body movements.

---

[130] Ibid.

Figure 41. Hiroshi Ishiguro (bottom) and his Geminoid double at the opening of the Ars Electronica Festival 2009. Photo from Wikimedia Commons.

At the other end of the spectrum is the *Telenoid* who, along with its cousins the *Elfoid*[131] and *Hugvie*,[132] are minimalistic humanoid figures consisting of a head and torso with "soft and pleasant skin texture and small, child-like body size"[133] intended to facilitate the user's telepresent interaction with a loved one. While the Geminoid figures are designed to replicate particular individuals, the Telenoid is generic:

[131] Elfoid is a hand-held version of the Telenoid.
[132] The *Hugvie* is a "hug pillow" the shape and size of the Telenoid.
[133] "Telenoid," < http://www.geminoid.jp/projects/kibans/Telenoid-overview.html> (Accessed 09 April 2014).

it can be perceived as male or female, young or old, depending on the remote operator. Through research with the *Telenoid*, ATR investigates the essential elements for representing and transferring human presence.

In a study done in 2005,[134] Ishiguro and his team found that subjects most felt the presence of the remote human in their interactions with the Geminoid series of robots, as they most closely resembled their human operators. The sense of presence was stronger than that in videoconferences

Figure 42. Telenoid *(center) and* Elfoid *(background, left) at the Ars Electronica Center during Ars Electronica 2012 in Linz, Upper Austria. Photo © 2012 Manfred Werner, via Wikimedia Commons.*

[134] Hiroshi Ishiguro, "Android Science—Toward a New Cross-Interdisciplinary Framework," *Springer Tracts in Advanced Robotics* 28 (2007): 125.

due to the physical form of the Geminoid robot—reinforcing Breazeal's assertion that a social robot must be *embodied*. However, Ishiguro's subjects also described a strong sense of the uncanny[135] in interacting with these robots. While the Geminoid bears a strong human resemblance, the illusion is shattered once the robot moves: the motion is jerky and artificial, marring the performance of humanity it represents. Even when stationary, the Geminoid fails the "physical Turing Test" Ishiguro devised for his creation: seventy percent of his test subjects were able to identify the android as such within two seconds. With the addition of human-like "micro-movements" (small, unconscious gestures like shifting balance, adjusting posture, and blinking), only thirty percent of subjects could identify the android—but as the time frame of the test increases the rate of identification rises dramatically. Ishiguro found in the study that even through "the subjects consciously recognized the other as an android[,]… they unconsciously recognized it as a human and dealt with it as a social partner"[136]—a conclusion that demonstrates that regardless of whether a humanoid robot is worthy of social interaction *in theory*, it is nevertheless treated as worthy in practice.

While Breazeal and Ishiguro approach the development of social robots from opposite sides of the equation—the former focusing on communication through the imitation of human emotional states, the latter on the imitating the physicality of the human subject—they do share a common theme. Both approaches treat humanity as a *performance* capable of imitation by a mechanical being rather than a metaphysical phenomenon that inheres exclusively to the biological substrate of the human model. The success of this performance is dependent upon its interpretation and acceptance by an interactive partner. This performative approach to the human is reflected in both Turing's proposed Imitation Game and in Stephen Wise's scale of practical autonomy. It draws upon our own evolutionary programming to treat as human an entity that *behaves as if* it is human. While philosophers and theorists argue over the

[135] For a discussion of the Uncanny Valley and the effect of near-human appearance on human viewers, see Introduction, page 36.
[136] Ishiguro, "Android Science," 125.

metaphysical ephemera requisite for emergent sentience in a mechanical substrate, or whether such a sentience is even possible, a performative definition of humanity offers a practical alternative that better reflects our day-to-day experience of human interaction as well as the human-robot interaction that is becoming increasingly common in the twenty-first century.

THE PERFORMANCE OF HUMANITY: A NEW DEFINITION

Over the course of the twentieth century, the evolution of the autonomous artificial human has forced us to examine precisely what we mean by the term *human*. On the one hand, there is no denying biology: a human is a living being of the genus and species *Homo sapiens*, evolved somewhere between fifty and a hundred thousand years ago into the form we recognize today as modern man. But this is not the definition we work with on a daily basis. As an adjective, *human* refers to those characteristics we typically associate with human beings: our personality traits, our emotions, and our social behaviors. In typical social interactions, the human beings we encounter perform according to our expectations based on the social and cultural cues in play in any given situation. Conversely, when a human being performs in a manner that is deemed by social standards to be outside the normal parameters of human behavior, we might label that performance (or that person) as *inhuman* or *inhumane*—without suggesting that the person him- or herself is not a human being.

Similarly, the various definitions of human we have explored over the course of this chapter are based in observable performative behavior. In his *Discourse on the Method*, Descartes distinguishes the rational soul of man from the lesser soul of animal and (presumably) the lack of a soul in the machine through two distinct tests;[137] both these tests as stipulated by Descartes are dependent upon our

---

[137] René Descartes, *Discourse on the Method and the Meditations*, trans. John Veitch (New York: Cosimo, 1924, 2008), 44-45. The passage is quoted below in its entirety.

> were there such machines exactly resembling in organs and outward form an ape or any other irrational animal, we could have no means of knowing that they were in any respect of a different nature from these animals; but if there were machines bearing the image of our bodies, and capable of imitating our actions as far as it is morally possible, there would still remain two most certain tests whereby to know that they were not therefore really men. Of these the first is *that they could never use words or other signs arranged in such a manner as is competent to us in order to*

observation of the animal or machine in question. The first test is based in language—the seventeenth-century version of the Turing Test. Descartes could imagine neither animal nor machine able to declare its thoughts as a man does; that is, through spoken language. Yet the famous gorilla Koko, while she does not speak as a typical human, nevertheless is quite competent in holding a conversation: she understands spoken English and American Sign Language at approximately the level of a human toddler, and can use several hundred ASL signs in communication.[138] Likewise, no machine has yet passed an official Turing Test, although chatbots regularly fool human interactors for short periods of time. What would Descartes have thought of Watson?

The second test Descartes proposes is also based in performance: "although such machines might execute many things with equal or perhaps greater perfection than any of us, they would, without doubt, fail in certain others…."[139] In this second test, a nonhuman is distinguished by his failed performance of a particular task, though he may excel at others. As we have seen in our exploration of robotics and artificial intelligence throughout the twentieth century, machines consistently outperform humans in any number of both physical and mental tasks in controlled environments, yet to date no single artificial human is capable of reaching human-level competence across the entire range of human physical and mental activities. Take Watson, for example: while handily defeating his human opponents in *Jeopardy*, his occasional incongruous incorrect guesses and his unusual *Double Jeopardy* wagers stood

---

> *declare our thoughts to others*: for we may easily conceive a machine to be so constructed that it emits vocables, and even that it emits some correspondent to the action upon it of external objects which cause a change in its organs; for example, if touched in a particular place it may demand what we wish to say to it; if in another it may cry out that it is hurt, and such like; but not that it should arrange them variously so as appositely to reply to what is said in its presence, as men of the lowest grade of intellect can do. The second test is, that *although such machines might execute many things with equal or perhaps greater perfection than any of us, they would, without doubt, fail in certain others from which it could be discovered that they did not act from knowledge*, but solely from the disposition of their organs: for while reason is a universal instrument that is alike available on every occasion, these organs, on the contrary, need a particular arrangement for each particular action; whence it must be morally impossible that there should exist in any machine a diversity of organs sufficient to enable it to act in all the occurrences of life, in the way in which our reason enables us to act. Again, by means of these two tests we may likewise know the difference between men and brutes. (Emphasis mine.)

[138] Koko also scores consistently between 70 and 95 on standard human child intelligence tests. Steven Wise assigns Koko a rank of 0.95 on his scale of practical autonomy (human is the standard at 1.0). See Wise, *Drawing the Line*, 207-30.

[139] Descartes, *Discourse on the Method*, 45.

outside the realm of what would have been expected of a human opponent, thus marking him as nonhuman. Similarly, Ishiguro's Geminoid robots, while initially mistaken for human when encountered briefly, are quickly revealed as nonhuman based on their jerky, machine-like movement.

While twenty-first century technology has not yet achieved the development of a true *Homo artificialis*, science fiction is filled with numerous examples of artificial humans that nearly perfectly mimic their human counterparts—and it is only a form of this second Cartesian test that reveals the character's nonhuman nature. In the second part of this study, we will look at three primary examples of artificial humans in science fiction film and television in order to analyze the performatives they display and evaluate their status as *Homo artificialis*.

CASE STUDIES: *HOMO ARTIFICIALIS* IN SCIENCE FICTION

For these images reflect not only the hopes and fears that cluster around the expanding role of robotics and artificial intelligence in our culture today but also a growing awareness of and attention to our own level of artifice, of constructedness, of how we often seem controlled by a kind of program not so very different from the sort that drives the artificial beings which abound in our films—and, perhaps, awareness of how the films that detail these anxieties might even assist in constructing us along these lines.

J.P. Telotte, Replications[1]

While the idea of the artificial human has permeated our consciousness since long before the legend of Hephaestus' golden maidens, it was not until the past century that our technology began to catch up with our aspirations. The social robots of the early twenty-first century are beginning to cross the Uncanny Valley, climbing ever closer to their fictional counterparts. At the same time, fictional depictions of artificial humans have become more complex—and more troubled—reflecting our own uneasy suspicion that we are, as Descartes and his contemporaries asserted, little more than biological machines.

Until the early nineteenth century, the artificial human in fiction was traditionally treated as an *object*—whether servant, threat, or plaything, it had no *self*, reflecting instead the whims and desires of its maker. Take the legend of Pygmalion and Galatea, for example, in which the sculptor's desperate love for his marble creation leads to Aphrodite granting the statue life. From Ovid's account in *Metamorphosis* to Rousseau's melodramatic retelling *Pygmalion*, the narrative focuses on either Pygmalion himself—his vanity and his misogyny—or the pity Aphrodite shows in granting his wish. At no time does Galatea reflect on her sudden awakening, or question her existence; instead, she unquestioningly fulfills Pygmalion's desire for "the perfect woman" á la *The Stepford Wives* (1975)—an

[1] J. P. Telotte, *Replications: A Robotic History of the Science Fiction Film* (Urbana: University of Illinois Press, 1995), 4.

image of untroubled domesticity suborned to her creator. With *Frankenstein*, however, Mary Shelley reformulated the artificial human as a *subject*, conscious of himself as a creation and seeking answers from his creator.[2] Just as Descartes' famous formulation "I think, therefore I am" presupposes the existence of an *I* who is doing the thinking, the ability of Frankenstein's Creature to pose the questions "Who was I? What was I? Whence did I come? What was my destination?"[3]—questions that continually plague mankind—assumes that there is something within the creature capable of formulating these questions in the first place. By asking the question, the Creature demonstrates subjective consciousness—in short, he becomes a *self*.

What does it mean to be a self? Self implies a unity of experience—an entity that is aware of its own existence as distinct from the environment which surrounds it, and aware as well of this existence as a (relatively) stable temporal continuity. Above all, a self is possessed of a *mind*. The contemporary analog to that which Aristotle termed the *soul*,[4] the mind comprises the set of cognitive faculties that enable an entity to have a subjective awareness of and intentionality towards his or her environment, to perceive said environment and possess the agency with which to respond based on those perceptions, and to have consciousness, including thinking and feeling. The relationship between the mind and the body in which it is manifested has been the subject of metaphysical debate (the *mind-body problem*) since pre-Aristotelian times, although it was not until Descartes that this issue was framed as we understand it today. Yet despite the points of contention concerning the precise nature of mind,

---

[2] Some later versions of the Pygmalion myth, such as poems on the topic written in the twentieth century, do actually address Galatea's emergent consciousness. Similarly, while early legends surrounding the famous Golem of Prague refer only to its brute nature, post-Frankenstein accounts begin to attribute human thoughts and feelings to the construct, changing the legend in response to the emerging concept of self. In the 1915 German silent film *Der Golem*, for example, the creature falls in love with the wife of his creator; when she does not return his affection, he turns violent (Paul Hermann Wegener and Henrik Galeen, *Der Golem* [Berlin: Deutsche Bioskop GmbH, 1915]).

[3] Mary Wollstonecraft Shelley, *Frankenstein; or, the Modern Prometheus*, 3 vols. (London: Lackington, Hughes, Harding, Mavor, & Jones, 1818), Kindle edition, Location 1567.

[4] The most significant difference between the modern sense of *mind* and the Aristotelian *soul* is that Aristotle included in his definition faculties which are now classified as strictly biological processes—specifically, the nutritive and reproductive faculties—which, unlike faculties such as reason, are not contested under the current mind-body debate. Strictly speaking, the non-biological artificial humans discussed in this chapter may lay claim to a *mind* without also possessing a *soul* in the Aristotelian sense.

contemporary philosophers of the Western tradition generally agree that *self* represents an individual ego comprised of the sum total of experiences over the time period during which this ego has existed, a paradigm also known as the *autobiographical self*, "a product of an extended consciousness that places a person in an individual history over time, aware of a lived past and anticipated future."[5] Under this paradigm, the self can be seen as largely *constructed*—an individual is born neither as a genetically determined, immutable personality nor as a *tabula rasa* upon which any possibility may be inscribed, but rather develops from infant to adult through a long and complex process of social and cultural interaction with other human beings. Through these interactions an individual learns acceptable modes of behavior appropriate to the variety of situations and environments that individual is likely to encounter; based on the outcomes of these interactions (whether positive or negative) an individuated *self* arises, with personality traits and behaviors peculiar to that self.

In his thesis "Do Androids Dream of Electric Shocks? Utilitarian Machine Ethics," Brendan Vize uses the idea of the autobiographical self in his analysis of an artificial person as a possible object of moral consideration.[6] According to Vize, the "biographical life" of a person makes that person unique; the existence of and (self-)awareness of this biographical life makes a person more complex and valuable than a non-person within the same moral situation. Vize illustrates this argument through the thought experiment of Simon and Sally:

> Suppose that Simon comes home to find his girlfriend Sally having sex with his best friend. In a fit of jealous rage, Simon beats and stabs Sally to death. It just so happens that, unbeknownst to Simon, Sally is a robot. Should Simon be charged with murder? Has Simon done anything morally wrong?

---

[5] Barton J. Blinder, "The Autobiographical Self: Who We Know and Who We Are," *Psychiatric Annals* 37, no. 4 (2007): 276.
[6] To be an *object of moral consideration* implies that an entity's preferences and capacity for quantitatively *good* or *bad* conscious states must be taken into consideration in all dealings with that entity. For a complete philosophical proof of the moral consideration of machines, see Brendan Vize, "Do Androids Dream of Electric Shocks? Utilitarian Machine Ethics" (Masters Thesis, Victoria University of Wellington, 2011).

> … The fact that Simon has lived with Sally as his girlfriend without realising that she was a robot, and that she perhaps developed another *intimate relationship* with his best friend, suggests that Simon has not only broken an object, but has also ended a *biographical life*. Sally is no longer able to fill the *social role* that she previously had, the kind of role that currently can only be filled by persons. The world has lost a *personality*, along with all the idiosyncrasies, plans, projects, creative output, and relationships that come along with one.[7]

The highlighted terms in Vize's argument above reinforce the idea of the autobiographical self as a sufficient condition of personhood, and any machine that exhibits this autobiographical self is described by Vize as a "fuzzy person"; that is, "a being that has behavioural personhood, but whose consciousness is in doubt."[8] The difficulty lies in determining whether this being is sentient or not, an epistemologically intractable problem even among fellow sentient beings (as further examined below). While arguing that "if a machine is sentient, then its preferences should be considered as important as the equivalent preferences ('in so far as rough comparisons can be made') of other sentient beings"[9] —the second of his five "Principles of Machine Ethics," the "Principle of Equal Consideration of Interests"—Vize ultimately concludes that "because of the extra level of doubt about the consciousness of machines, the interests of a naturally-occurring being should be preferred over the interests of a machine with similar behavioural properties."[10]

The experience of *self* is ontologically a first-person phenomenon—there is no known method to evaluate empirically the existence or quality of the self of another. This does not prevent us, however, from treating other humans *as if* they are entities similar to our own selves: as discussed in the previous chapter, we attribute similar mental states to those entities that behave in ways similar to ourselves.

---

[7] Ibid., 74-75, emphasis mine.
[8] Ibid., 60.
[9] Ibid., 79.
[10] Ibid., 80.

Daniel Dennett refers to this as the *intentional strategy*:[11] we draw inferences about an entity based on its appearance and behavior, and we use those inferences in predicting the future behavior of that entity. Cognitive psychologists label this phenomenon *theory of mind* (ToM): that an entity similar to ourselves in appearance or behavior is possessed of a mind similar to our own. Regardless of what inferences we may draw, and what characteristics we may believe an entity possesses, we may only evaluate an individual self in terms of how that self is presented to others.

What kind of *self*, then, does a robot present? As we saw in the last chapter, social robotics researchers of the last two decades have been indirectly dealing with this topic as they program their "sociable robots" to interact with humans in standard human environments. In 2007, Kerstin Dautenhahn addressed some of the social and philosophical implications of the development of "cognitive robot companions," listing the characteristics socially interactive robots must exhibit, including the ability to "express and/or perceive emotions; communicate with high-level dialogue; learn models of or recognize other agents; establish and/or maintain social relationships; use natural cues (gaze, gestures, etc.); exhibit distinctive personality and character; and may learn and/or develop social competencies."[12] These characteristics are no different from those we expect of a human interactant in any given social environment.

Dautenhahn's invocation of a social paradigm to classify HRI transactions immediately recalls the sociology of Erving Goffman. Goffman's sociology is performative; that is, it focuses primarily on behavior rather than intent. Goffman claims that "the 'true' or 'real' attitudes, beliefs, and emotions of the individual can be ascertained only indirectly"[13]—that is, through the expressions that the individual either "gives" or "gives off"—creating a "self" that is behavior-based and accessible only through inference: "The self, then, as a performed character, is not an organic thing that has a specific location…;

---

[11] See Daniel C. Dennett, "True Believers: The Intentional Strategy And Why It Works," in John Haugeland, *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, Rev. and enl. ed. (Cambridge, MA: MIT Press, 1997), 57-80.
[12] Kersten Dautenhahn, "Socially Intelligent Robots: Dimensions of Human-Robot Interaction," *Philosophical Transactions of the Royal Society B-Biological Sciences* 362, no. 1480 (2007): 684.
[13] Erving Goffman, *The Presentation of Self in Everyday Life* (Garden City, NY: Doubleday, 1959), 2.

it is a dramatic effect arising diffusely from a scene that is presented."[14] Applying Goffman's behavior-based sociology allows us to discuss the robot as a performative "self" involved in a social interaction without having to address the metaphysical question of whether the robot is "self-aware."

In the following three case studies, I examine the "self" performed by each of the artificial humans under consideration, taking into account a number of factors:

1. PHYSICALITY: What are the physical characteristics of the being in question? Does the artificial human physically present as human, or is he somehow marked as non-human?
2. BEHAVIOR: What are the behaviors the artificial human displays in the performance of *human*—the mannerisms, verbalizations, etc. that indicate that this being is not just a conscious entity, but an individuated self?
3. STATUS: What is the status of artificial humans within the diegetic world of the film / series?
4. DISTINCTIVENESS: Is the character under consideration a *typical* or *atypical* example of artificial humans within the diegetic world?
5. PERFORMANCE: Finally, does the character offer a *sincere* and *effective* performance of humanity?
6. The three SF universes I examine in the sections below offer a rich environment in which to explore these questions, and provide three quite distinct glimpses into the status of *Homo artificialis* within human society.

I have limited my study of artificial humans to those that fall within the definitions laid out in my Introduction. In addition, prefer to avoid the complication of a biological substrate. I have therefore chosen to focus only on those artificial humans that are completely mechanical in nature, so that the consciousness of these *Homo artificialis* examples cannot be explained away through a connection to their biological elements, as could either Frankenstein's Creature or Capek's Robots. The first case study

[14] Ibid., 252-3.

revolves around the character of C-3PO in the *Star Wars* universe, arguably the most complex SF universe under consideration. Robots abound throughout the Galactic Empire, ranging from maintenance droids the size of small dogs to battle droids designed for combat situations. In the *Alien* universe, on the other hand, artificial humans are relatively rare, and those who do exist are visually indistinguishable from their human counterparts. In this universe I focus primarily on David-8 from the most recent film of the franchise, *Prometheus* (2012), but he represents one facet of a continuum of synthetic humans within the *Alien* franchise. Finally I turn to the *Star Trek* universe and the character of Data, who is not only unique as a sentient android, but is eventually granted legal status as a person in a society in which any form of robot is extremely rare.

Interestingly enough, none of the SF universes examined here typically use the term *robot* for the artificial humans within their societies. It seems that in the science fiction of the late twentieth and early twenty-first centuries the term *robot* has become something of an insult when applied to a mechanical humanoid entity, implying that they are closer in nature to their industrial cousins than to the humans they are designed to emulate. The preferred term for those mechanical (or biomechanical) humanoids whose appearance is nearly human varies among the three universes, although all three do use the term *android* in everyday parlance. In the *Star Wars* universe, all manner of autonomous mechanical beings are referred to as *droids*, whether humanoid in appearance or not. The *Aliens* universe, on the other hand, most commonly uses the term *synthetic*s to refer to its androids, which are also mechanical in nature. Finally, the artificial human Data in the *Star Trek* universe consistently refers to himself as an android; other examples of artificial beings are extremely rare, and the terminology used for them varies depending on their nature.

THE PERFORMANCE OF THE HUMAN

What does it mean to be human? First and foremost, *human* is a matter of biology, as we have already discussed. Further, despite the posthumanist science-fiction concept of uploading our brains to

the net and existing in a virtual world, humans are by their very nature *embodied* beings. They exist in a physical world that is, largely, engineered to accommodate their physicality. While roboticists such as Dautenhahn and Breazeal emphasize the social nature of their creations, claiming that sociable robots do not have to be humanoid in appearance in order to be successful, as participants interacting in a human-centric environment on a daily basis, the artificial humans in my case studies must fit into their given environments *physically* as well as *socially*. In a controlled environment such as seen in modern automated factories, humans must adapt to the physical realities of the robots they encounter—through the use of safety gear, special tools, and restricted areas of movement, for example. However, for fully autonomous robots to successfully operate in an environment built around the exigencies of the human form, they must be able to make use of that environment in the same way as do the human majority. This implies at the very least a form that is roughly the same size and shape as a typical human, with "standard" appendages and articulation as well as input sensors that operate in roughly the same ranges as human sense organs. Simply put, it is easier to build a robot to fit the human world than to build the world around a non-humanoid robot. For practical purposes, then, an autonomous artificial human must conform to some degree to human physical appearance and behavior standards.

A greater challenge to *Homo artificialis* is the broader definition of human within human society. As we saw in the last chapter, we typically define the category of *human* not only as a matter of physicality but as a concept which embodies a number of socially constructed characteristics. Enumerating precisely *what* these characteristics *are* is not an easy task; the very attempt forces one into a situation known in legal circles as *the elephant test*: difficult to describe, but you know it when you see it. Brendan Vize enumerates several "properties of persons" in his discussion of machines as moral agents: persons are conscious beings who have "interests and goals, the ability to use language, the ability to form reciprocal relationships, the ability to understand and apply moral concepts,

rationality, self-awareness, [and] a personality."[15] In arguing for a broader definition of *personhood*, Steven Wise of the NhRP identifies the quality of *practical autonomy* as being "sufficient to entitle any being of any species to liberty rights"[16]—that is, the right to be treated *as a person* under the law, with the liberties and protections inherent thereto. Formulating practical autonomy as a quality that "most judges think is *sufficient* for basic liberty rights," Wise claims that

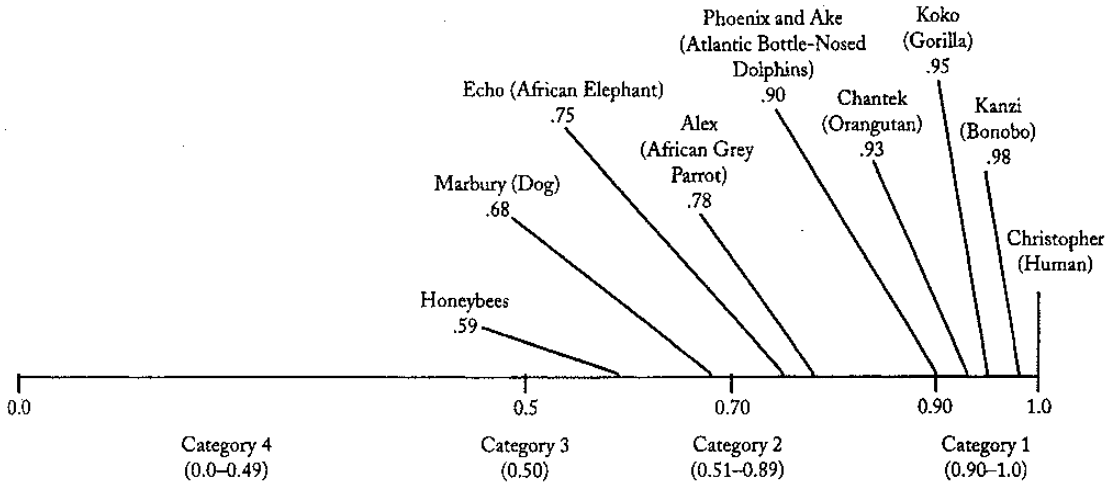
> a being has practical autonomy and is entitled to personhood and basic liberty rights if she:
>
> 1. can desire;
> 2. can intentionally try to fulfil her desires; and
> 3. Possesses a sense of self sufficiency to allow her to understand, even dimly, that it is
>
> she who wants something and it is she who is trying to get it.
>
> Consciousness, not necessarily self-conscious, and sentience are implicit in practical autonomy.[17]

Both of these definitions revolve around the use of a "human yardstick"[18] as the standard by which to measure non-human personhood, whether in animals or machines, for as Wise asserts, "the more exactly the behavior of any nonhuman resembles ours… the more confident we can be that she possesses desires, intentions, and a sense of self resembling ours, and that we can fairly assign her an autonomy value closer to ours."[19] Vize terms this exhibition of human-like behavior "behavioural personhood," signifying that "a being has (at least) the outwardly visible properties of personhood."[20]

---

[15] Vize, "Do Androids Dream," 56-57.
[16] Steven M. Wise, *Drawing the Line: Science and the Case for Animal Rights* (Cambridge, MA: Perseus Books, 2002), 24.
[17] Ibid., 32. Nonhuman animals are classed into four categories based on an assessment of their autonomy as compared to a human standard (which is set at 1.0). See Figure 43 on following page for a comparison of these rankings.
[18] Ibid., 45.
[19] Ibid., 36.
[20] Vize, "Do Androids Dream," 57.

Who Should Get Basic Liberty Rights under the Common Law?



Autonomy Values

Category 1 = Nonhuman animals who clearly possess sufficient autonomy for basic liberty rights.

Category 2 = Nonhuman animals who, according to increasing evidence, possess sufficient autonomy for basic liberty rights. Autonomy values of .80, .70, and .60 are the cutoff for basic legal rights using a narrow, moderate, and expansive reading of the precautionary principle, respectively.

Category 3 = Nonhuman animals about whom we do not know enough to reasonably determine whether they possess sufficient autonomy for basic liberty rights. This category probably includes most species.

Category 4 = Nonhuman animals who, according to increasing evidence, lack sufficient autonomy for basic liberty rights.

Figure 43. Chart indicating where various nonhuman animals rank in practical autonomy. Chart © 2002 Steven Wise, Drawing the Line: Science and the case for animal rights. *Reprinted by permission of Perseus Publishing, a member of the Perseus Books Group.*

But what behaviors comprise this "human yardstick"? Humans are primarily social creatures, and for roboticists like Breazeal and Ishiguro, imitation of these social behaviors is the primary goal for a successful sociable robot. Kersten Dautenhahn's principle of *robotiquette* (the set of rules and standardsfor HRI) offers a number of suggestions as to which "human" behaviors are desirable in a robot interactant, although the need for these social skills is heavily dependent on the social situation of the particular robot:

> At one end of the spectrum, we find that robots, e.g. when operating in space, do not need to be social, unless they need to cooperate with other robots. In contrast, a robot delivering the mail in an office environment has regular encounters with customers, so within this well-defined domain, social skills contribute to making the interactions with the robot more convenient for people. At the other end of the spectrum, a robot that serves as a companion in the home for the elderly or assists people with disabilities needs to possess a wide range of social skills which will make it acceptable for humans.[21]

For Dautenhahn, robots working within a human environment are modified for the convenience or comfort of the human interactants rather than that of the robot him- or herself; in this formula, the robots are again *objects* within a humanocentric environment. However, in each of the case studies below, the artificial human at the center of my analysis attempts to successfully navigate and function on a daily basis *as an active, autonomous subject* within a social system filled with human and/or humanoid interactants—a performance requiring a fair amount of social skill.

This navigation may be viewed through the Goffmanian lens of social performance. A central concern in Goffman's discussion of social performance is the idea of *sincerity* of performance. Goffman offers two distinct paradigms for evaluating sincerity within a performance. The first paradigm conflates *sincerity* with *belief* on the part of the performer: "the individual's own belief in the impression of reality

[21] Dautenhahn, "Socially Intelligent Robots," 700.

that he attempts to engender in those among whom he finds himself."[22] At one extreme is the individual who is fully taken in by his own act, to the point that he sincerely believes that his performance reflects "the real reality." Such a performance is labelled by Goffman as *sincere* as it belongs to the "common-sense" model that forms the majority of social performance situations; that is, that participants in the situation may reasonably expect that the performance is honest and the performers are not attempting to deceive or mislead the audience (or, in fact, one another). At the other end of the spectrum is the *cynic*, who "has no belief in his own act and no ultimate concern with the beliefs of his audience."[23] A cynic often, but not always, has ulterior motives for subterfuge; he typically possesses what Goffman terms "dark" secrets that "are incompatible with the image of self that [he] attempts to maintain before [his] audience."[24] While a *sincere* performer may be aware of some discrepancy between his performance and his own image of self, there is no underlying attempt to deceive; on the other hand, a *cynical* performer, fully aware of this discrepancy, conceals his own disbelief in order to convince the audience that the performed reality is the only true one.

I find Goffman's discussion of *belief* to be at odds with his assertion that self is only visible through performative behavior. While I may evaluate my personal belief in both my own performance(s) and those I witness in others, I cannot evaluate the level of belief within the other member of a performance transaction (whether performer or audience)—regardless of whether that other member is another natural human or an artificial one. To determine the sincerity of another's performance through an evaluation of their own belief in said performance would involve access to internal states that the performance's audience does not typically possess in everyday life. However, this issue is complicated in evaluating the performance of fictional characters. In everyday life, a particular person "performs" an identity in front of an audience of those present for the performance: a particular social group, for

[22] Goffman, *The Presentation of Self*, 17.
[23] Ibid., 18.
[24] Ibid., 141.

example. However, in film and television, and often even in novels and short stories, there are actually two *different* audiences, each with a different set of expectations based on differing knowledge about the situation: not only the social group present within the fictional universe, but also the *reader* or *viewer*, who often has access to information the characters do not. In the following case studies, my position as a third-person (relatively) omniscient observer gives me the opportunity to see behaviors exhibited by the performer that are not typically available to the "audience" within the diegetic world of the film—that is, the other characters inhabiting the world who must judge the success of a particular performance. Theoretically, my observation of behaviors in the absence of an audience that indicate "disbelief" in a performance could aid in an evaluation of sincerity; however, this would also involve attributing particular mental states to the artificial human, putting myself dangerously into the realm of metaphysics.

Fortunately, it is not necessary to attribute "belief" to an artificial human in order to evaluate the sincerity of his performance. Goffman's second paradigm evaluates sincerity based on whether the performer possesses information incompatible with the performance given—information that would disrupt the performance, causing it to fail. I can therefore evaluate sincerity of performance based on the information possessed by performer and by audience in any given situation. Standing outside the frame of the performance, I am privy to the behaviors the characters exhibit in the absence of a diegetic audience. How the artificial human "performs for himself" may then affect the sincerity of his more public performances not by indicating disbelief in the performance but by revealing information the diegetic audience does not possess. In this paradigm, the difference between a sincere performance and a cynical one lies in the control of information: the "dark secrets" mentioned above, typical of a cynical performance, are generally absent in a sincere one, and the "self" the artificial human presents when in the presence of others is more or less identical to the "self" he presents when alone. It is the difference

between an actor performing a role on stage (whom the audience recognizes as both actor and character) and a con artist performing before his mark (who is only aware of the character performed). Successful impression management on the part of these artificial humans is vital regardless of the level of sincerity within the performance. While a human performer may fail in the performance of a social role without endangering their human status, the performance of *human* by an artificial being may succeed or fail on two fronts: the physical and the social. In Goffmanian terms, when encountering an entity that appears human, an interactant expects that the entity will exhibit not only a certain set of performative behaviors indicative of a person operating in a particular social role (the expression that entity "gives"), but also a physical appearance and an array of "unconscious" social cues (the expressions "given off" by that entity) that fall within expected human parameters. Depending on the level of tolerance of artificial humans within a society, a failed performance may have disastrous consequences for both performer and audience—particularly when those performances are less than sincere.

THEY WALK AMONG US

As we saw in the last chapter, science fiction is filled with stories of artificial humans hiding in plain sight within human societies. The central plot of the 2003-2009 television series *Battlestar Galactica*, for example, revolves around the struggle between the Cylons, a race of biomechanical artificial humans originally created as soldiers, and the remnants of the human civilization responsible for their creation.[25] At the other end of the spectrum lies the character of Andrew in *Bicentennial Man*. In both the 1999 film starring Robin Williams and the 1976 Isaac Asimov novelette upon which it is (loosely) based, Andrew is a robot who for two hundred years has sought to successfully "pass" within

[25] In contrast, the Cylons of the original 1978 series were a race of independent sentient humanoid machines, originally created by an ancient reptilian race (also called Cylons, who died out long before the time period in which the series is set) as servants; in the earlier series, the robot Cylons are an aggressively expanding empire who have declared war on the human race. The 2003 reboot of the franchise removes the alien origins of the Cylons, at once making them more human and the resultant struggle with their creators more morally complex.

human society, continually modifying himself to become "more human"—ultimately sacrificing his robotic immortality in order to gain his humanity in the eyes of the society in which he lives. In both of these scenarios, the fact that a particular character (or characters) is an artificial human forms the central plot device around which the story revolves. These types of plotlines fall into the common SF categories described by Asimov in 1982 as *Robot-as-Menace* and *Robot-as-Pathos*, respectively. Or, to put these two examples into Kangian terms, *Battlestar Galactica* is representative of the *inevitable conflict* school of thought, in which the relationship between natural humans and their artificial creations is an inherently adversarial one; *Bicentennial Man* on the other hand represents the theory of *cybernetic mergence*, albeit in the opposite direction, as a purely mechanical being becomes increasingly organic in his quest to "become a real boy."

While these types of stories deal directly with the struggles of the artificial human within human society, for a number of reasons they are not the type of plotlines that interest me for the purposes of this dissertation. First of all, I am interested in exploring those artificial humans in science fiction who function within society in the same way that modern proponents of social robotics envision their creations: working alongside humans in their day-to-day environment, in situations which are not inherently robot-centric or confrontational; in short, the *Robot-as-Intelligent-Tool* paradigm investigated in the majority of Asimov's short stories. In working alongside their natural human counterparts, these artificial humans achieve (or strive for) *equivalence through sentience*; their treatment within the diegetic universe created by the franchise (whether as persons or property) reflects both on their own status as *Homo artificialis* and on the societal milieu in which they exist. In contrast to the premise of *Battlestar Galactica* (or even of many Asimov short stories and the movies derived therefrom), the primary story arcs in the SF universes I have chosen do not revolve around the human-robot interaction and its outcome, though individual scenes (or even episodes in the serials) may address the issue of the robot per se as a subplot. Instead, the artificial human(s) in each universe are involved as individual

characters—whose artificiality is little more than a character trait—within a larger story arc that has nothing to do with robotics, cybernetics, or artificial intelligence, even though these issues may be addressed within the larger framework.

I have also chosen not to pursue the type of story line represented by *Bicentennial Man* in particular because the idea of *cybernetic mergence* rather dilutes my construction of the concept of *Homo artificialis* by melding the mechanical with the biological, leading to possible substratism. The character of Andrew is the ultimate realization of Vaucanson's *Digesting Duck*: over the course of two centuries, Andrew gradually "upgrades" his mechanical systems with artificial organs biomechanically equivalent to human ones; at the same time he makes the various prostheses he has invented available to the human population. As humans replace their own failing parts with Andrew's inventions, the physiological differences between Andrew and an average human diminish, until there truly is no

Figure 44. The robot Andrew Martin evaluates his appearance after his positronic brain is placed in a biomechanical, human-equivalent body, giving him not only the appearance and facial expressions of a human but sensory feedback as well. Andrew's previous body sits to the side of the mirror. Still from Bicentennial Man *(1999).*

ontological distinction between the two. In contrast, *Homo artificialis* is at once an artificial creature and a person; this status lies expressly in the duality created by the juxtaposition of artificiality and life, not in the erasure of the artificial.

Finally, by choosing to work within universes in which the human-robot relationship is a relatively benign one, and is (for the most part, as we will see later) a well-established feature of the universe but not a central factor in the overall story arc, I can avoid those performances Goffman would label as *cynical* misrepresentations. In the universe of *Battlestar Galactica*, for example, the performance of humanity by the Cylons hidden among the surviving Colonist population is a deliberate misrepresentation intended to gain access to the society under false pretenses. In Goffmanian terms, these Cylons take on the discrepant role of *informer* or *shill*,[26] depending on the performative frame at any given time. Humans in the *BG* universe must be continually vigilant against such duplicitous performances; their very survival depends on their ability to distinguish the sincere performance from the cynical one. It is possible, however, that the performance of humanity by artificial humans constitutes a kind of a *realization through play*:

> A status, a position, a social place is not a material thing, to be possessed and then displayed; it is a pattern of appropriate conduct, coherent, embellished, and well articulated. Performed with ease or clumsiness, awareness or not, guile or good faith, it is none the less something that must be enacted and portrayed, something that must be realized.[27]

---

[26] According to Goffman, "the informer is someone who pretends to the performers to be a member of their team, is allowed to come backstage and to acquire destructive information, and then openly or secretly sells out the show to the audience." The most common examples of the informer are the *traitor* or the *spy*, depending on the performer's initial sincerity. A *shill*, on the other hand, is "someone who acts as though he were an ordinary member of the audience but is in fact in league with the performers," using his "one of us" status to influence the attitude of the audience to the performers' benefit (Goffman, *The Presentation of Self*, 145-46).

[27] Ibid., 75.

The three artificial humans under consideration in the case studies below do not deny their artificial nature; rather, the expression of human performatives furthers each performer's journey along Goffman's "cycle of disbelief-to-belief" by serving as what Robert Ezra Park terms a *mask*:

> It is probably no mere historical accident that the word person, in its first meaning, is a mask…. In a sense, and in so far as this mask represents the conception we have formed of ourselves—the role we are striving to live up to—this mask is our truer self, the self we would like to be. In the end, our conception of our role becomes second nature and an integral part of our personality. We come into the world as individuals, achieve character, and *become persons*.[28]

Park's description of "becoming persons" is reflective of the idea of the *autobiographical self*.

Science fiction is filled with stories of robots who "become" more than they started out to be; whether given false memories at the outset or developing their own autobiography over time, these artificial humans unequivocally possess a "biographical life," are aware of themselves *as selves*, and are concerned with the success or failure of their imitation of "personhood." Unlike that of the Cylons, the performances of humanity by the artificial humans in the following case studies are generally *sincere*: that is, while the performers are aware of the discrepancy between their performance and their reality, these discrepant performances are not attempts to con the audience for private gain, for the audience is also aware of the discrepancy; the performers are seen as actors rather than conmen. These performances instead represent claims on the part of their performers to be granted the status of *persons* within these societies; through the continual repetition of performances that establish and reinforce the right to this claim, these artificial humans realize their identities (with various levels of success) as *Homo artificialis*—an artificial body possessed of a human (or, more properly, *person*) "self."

[28] Robert Ezra Park, *Race and Culture*, His Collected Papers, V 1 (Glencoe, IL: Free Press, 1950), 249-50, emphasis mine.

STUDY ONE: C-3PO THE BUMBLING MANSERVANT

Hello, I am C-3PO, human-cyborg relations. How may I serve you?

—C-3PO, *Star Wars: Episode I—The Phantom Menace*

We begin, then, with the earliest of the three primary artificial humans under consideration: C-3PO, a major character of the *Star Wars* universe, who appears not only in the six major films of the *Star Wars* franchise,[1] as well as various Expanded Universe narratives, but stars as well—along with his sidekick R2-D2—in his own animated series *Star Wars: Droids* (1985). As the first characters we see onscreen in the original *Star Wars* (1977, later renamed *Episode IV: A New Hope*), this pair serves as our introduction to the *Star Wars* universe. Within the first five minutes of the film, the two are established as a comic duo: as R2-D2 moves purposefully through the corridors of a spaceship under attack by Stormtroopers, seemingly on a mission, C-3PO follows him in a panic, shouting "This is madness!" and "We're doomed!" as he tries to avoid blaster fire (*Episode IV*). We are immediately struck by the very *un-robotic* personality of this robot: C-3PO is fussy and nervous, short-tempered with his companion, yet obsequious to the sentients[2] he encounters. He complains of aches and pains and bemoans his lot in life as a droid.

---

[1] In order of their release, *Star Wars: Episode IV—A New Hope* (1977); *Star Wars: Episode V—The Empire Strikes Back* (1980); *Star Wars: Episode VI—Return of the Jedi* (1983); *Star Wars Episode I—The Phantom Menace* (1999); *Star Wars: Episode II—Attack of the Clones* (2002); and *Star Wars Episode III—Revenge of the Sith* (2005). Complete bibliographical information is given in the Filmography; subsequent references to these six films are given by Episode number within the text (*Episode I*, for example).

[2] For the purposes of this discussion, I have divided the entities found within the Star Wars universe into four broad categories. First, I use the term *sentients* (following the practice of *Wookieepedia*) to broadly refer to all sentient biological humanoids, regardless of their species or planet of origin. In contrast, *droids* are purely mechanical beings, ranging from minimally intelligent housekeeping droids (equivalent to a Roomba automatic vacuum cleaner) to sophisticated protocol droids, considered some of the most advanced droids built. The universe also includes *cyborgs,* hybrid biomechanical entities with (at a minimum) the brain of a sentient, such as Darth Vader and General Grievous, and *clones*, nameless, factory-grown sentients that are genetically modified for obedience and treated much like droids in their subaltern status. The complex and problematic relationship between biology and legal and ethical status within the *Star Wars* universe is much too broad to tackle fully here, presenting a possible topic for future analysis.

I would like to comment here on the use of *Wookieepedia* as a source for information on the *Star Wars* universe. While *Wookieepedia* makes no claim to be a primary source, it is one of the largest and most comprehensive collections of *Star*

Figure 45: C-3PO and R2-D2 on Tatooine. Still from Star Wars: Episode IV – A New Hope *(1977).*

The pair of droids is also the thread that ties together the franchise. In the "prequel trilogy," which follows the life of Anakin Skywalker, we discover that C-3PO was originally built by Anakin himself from spare parts (*Episode I*). When Anakin married Padme Amidala, he gave her C-3PO as a wedding gift (*Episode II*); the droid later assisted Amidala in her role representing her home planet of Naboo in the Galactic Senate (*Episode III*). After the collapse of the Galactic Republic and rise of the Empire, and Amidala's death in childbirth, C-3PO and R2-D2 were given to her fellow senator Bail Organa, whom the pair served for nearly twenty years (*Episode III*). At the opening of the "original trilogy," Princess Leia Organa, daughter of the senator, entrusts the droids in her mission to help the Rebel Alliance by delivering the plans of the Empire's Death Star to Jedi Knight Obi Wan Kenobi on Tatooine, setting off

---

Wars minutiae available. Continually edited and updated by hundreds of people, *Wookieepedia* is very careful to distinguish between *Canon* and *Legends* (non-canon but accepted within the Expanded Universe) information and mere fan-based speculation. I have indicated in footnotes any details I was unable to verify myself within the original EU material (a nearly intractable project given the sheer size of the Expanded Universe).

the events of the trilogy (*Episode IV*). C-3PO also assists in Leia's rescue from said Death Star (*Episode IV*) and is a material part of the efforts of Leia, Han Solo, and Luke Skywalker in the rebellion—most notably impersonating a deity while convincing the Ewoks to aid the Rebel forces at the decisive and victorious Battle of Endor (*Episode VI*). Finally, while the "sequel trilogy"[3] has only just started production as of this writing, it has already been announced that actor Anthony Daniels (who has portrayed C-3PO in nearly every official media performance to date) is attached to the project. This means that, although we have yet to discover the role C-3PO will play in the lives of the next generation of *Star Wars* principals, we do know that he will, indeed, play a role.

R2-D2 and C-3PO have also proven highly influential to the direction of modern robotics research. In the preface to her groundbreaking book *Designing Sociable Robots*, Cynthia Breazeal attributes her decision to pursue the development of sociable robots to her childhood fascination with this pair:

> I remember seeing the movie *Star Wars* as a little girl. I remember being absolutely captivated and fascinated by the two droids, R2-D2 and C-3PO. Their personalities and their antics made them compelling characters, far different from typical sci-fi robots. I actually cared about these droids…. [At MIT] I remember thinking to myself that this was it—these kinds of robots were the real-life precursors to the *Star Wars* droids of my childhood…. As a well-seasoned researcher, I began to build a robot in the image of my childhood dream.[4]

The robot she built was Kismet, the first in her series of sociable robots designed to explore the necessary components of human-robot interaction. Breazeal deliberately took "R2-D2 and C-3PO as

---

[3] *Star Wars: Episode VII—The Force Awakens* has a projected release date of December 2015. *Episodes VIII* and *IX* are also planned for 2017 and 2019, respectively.

[4] Cynthia L. Breazeal, *Designing Sociable Robots*, Intelligent Robots and Autonomous Agents (Cambridge, MA: MIT Press, 2002), xi.

representative instances" in developing "a robot that is socially intelligent in a human-like way."[5] As the representative of the origins of the field of social robotics, it is only fitting that my first case study focuses on the human performance of C-3PO.

PHYSICALITY

From a physical standpoint, C-3PO is more properly considered *humanoid* rather than *human*: he bears the physical shape of an average human in terms of height and proportions, but is clad in golden metal with exposed circuitry and wiring rather than any sort of human-like skin—a design that pays homage to his inspiration: Maria from *Metropolis* (1927). His face is likewise human in proportion but has no moving parts: two circular lamps represent eyes, a small slit stands in for a mouth, and only a slight v-shaped crease indicates where a person's nose would be. In addition, C-3PO's exterior is free of any gendered shapes or iconography, rendering him visually neutral. Nonetheless, the absence of female gendering marks the droid as a male by default, and the lower register of his voice (that of actor Anthony Daniels) reinforces this assessment.[6] Indeed, according to *Wookieepedia*, "although droids had no sexes, they could be programmed with masculine or feminine personalities to simulate the behavior of a specific sex, thus, even mechanical beings were usually designated *he* or *she*,"[7] reinforcing a heteronormative gender binary. *Wookieepedia* identifies C-3PO as having "masculine programming."[8] C-3PO's artificial nature is further highlighted by the presence of visibly mechanical joints at knees and elbows as well as externally-attached pistons at the elbow that allow only a limited range of lower-arm motion and cause the droid to stand at rest with arms forward and bent, unlike a human's typical resting position of arms hanging freely at the sides. This erect stance, combined with a narrow range of leg,

---

[5] Ibid.

[6] In an interesting twist, the robot Dot Matrix in the Mel Brooks comedy *Spaceballs* (1987) is a parody of C-3PO, with marked female gendering: Dot has a more slender face and curly hair, molded breasts, and a "dress" consisting of puff sleeves and a paneled or pleated skirt, all in the same gold metal as the original C-3PO. Dot Matrix is voiced by Joan Rivers, further gendering the character as female.

[7] See note 2 above for an explanation of the *Wookieepedia*. Wikia, Inc., *Wookieepedia: the* Star Wars *Wiki*, "Sexes," accessed 31 July 2014, http://starwars.wikia.com/wiki/Sexes/. All references to pages on this website are collected in a single bibliographic entry.

[8] Wikia, Inc.,*Wookieepedia*, "C-3PO," accessed 31 July 2014, http://starwars.wikia.com/wiki/C-3PO.

Figure 46. Rotwang introducing Joh Fredersen to the "Machine-Man" in Metropolis; the robot that becomes Maria served as the inspiration for C-3PO's design. Although not evident in the image, Maria's robot form is clad in golden metal. Still from Metropolis *(1927).*

Figure 47. Dot Matrix, a parody of C-3PO's character in the Mel Brooks comedy Spaceballs*. Note that while C-3PO displays no evident external gender markings, Dot is distinctly marked as female. Still from* Spaceballs *(1987)*

waist, and head movements (compared to the average human), reflects the typical audience expectations for what is mechanical: awkward, jerky, and limited in mobility.

C-3PO's physical form reflects his purpose: as a "protocol droid," he is reminiscent of the "British manservant" archetype, designed to perform the functions of a butler, facilitator, or personal assistant.[9] Many protocol droids in evidence in the *Star Wars* universe are identical in style to C-3PO, varying only in the color of their metal plating (most commonly gold, silver, or white, although other colors also existed); one is seen at the beginning of *Episode I*, attending the two Jedi Knights as they await representatives of the Trade Federation. However, C-3PO himself goes through a number of physical iterations over the course of the films: his first appearance (within the diegetic timeline) shows him sans plating, as Anakin Skywalker had been in the process of (re)building C-3PO (from spare parts) when first encountered by Padme and Qui-Gon Jinn on Tatooine. He is eventually given gold plating, but loses a leg, which is later replaced by one with silver plating.

Unlike a biological entity, C-3PO is able to operate regardless of his physical state at any given time. He is at various times dismantled and beheaded, operating as either a limbless torso (carried in a pack on Chewbacca's back in *Episode V*) or a disembodied head (accidentally attached to a battle droid in *Episode II*), loudly lamenting his less-than-pristine condition. Despite his complaints of suffering—more mental than physical, as the droid never seems to experience *pain* so much as *fatigue*[10]—C-3PO can exhibit consciousness and sentience even when fully dismantled: in *Episode V*, C-3PO discovers Stormtroopers hiding on Cloud City, which his friends considered a safe haven. After being torn apart by the Stormtroopers, C-3PO is reactivated by Chewbacca, who then does not have enough time to

---

[9] That he is continually thwarted in fulfilling this intended function gives rise to much of the humor surrounding his character.

[10] In *Episode IV*, while wandering the Tatooine desert in search of Obi-Wan Kenobi, C-3PO offers running commentary on his plight to his companion R2-D2, including such gems as "My joints are almost frozen," "I have to rest before I fall apart," and "We seem to be made to suffer. It's our lot in life!" On the other hand, when being carried around in a backpack by Chewbacca in *Episode V*, C-3PO cries out "Ow! That hurts!" after striking his head on a low girder. This utterance seems more an outburst of alarm than of "true" pain, however, as it is the only mention C-3PO makes of physical injury causing actual physical discomfort rather than inconvenience or irritation. This is not to say that droids do not suffer from physical injury, however: in *Episode VI* we see droids being tortured at the hands of other droids for failure to properly perform their functions within Jabba the Hut's household.

Figure 48. Ripped limb from limb and carried on Chewbacca's back, C-3PO continues to offer commentary on the situation. Still from Star Wars: Episode V – The Empire Strikes Back *(1980).*

completely reassemble him. Chewbacca gathers C-3PO's parts into a pack and carries him during their escape; C-3PO complains and offers "advice" the entire escape (See Figure 43).

In *Episode II*, C-3PO's head becomes separated from his body on the assembly-line of the Battle Droid factory. The head is attached to a battle droid, while C-3PO's body is fitted with a battle droid head, as seen in Figure 44. As the two hybrid droids take to the field of battle in the subsequent scene, their behavior calls into question exactly where the droid's control center (his *self*, if you will) is located. Assuming this center resides in the head (analogous to the brain in a biological entity), one would expect the character's body with the battle droid head to display battle programming—which it does, albeit with some clumsiness due to the differences in design between the two humanoid robots. The converse configuration proves more problematic. While C-3PO's head on the battle droid body apparently contains his seat of consciousness (as his running commentary on the matter indicates), it is nonetheless the battle programming which is in control of this hybrid as well, as it fights alongside its fellow battle droids despite C-3PO's protests. In fact, C-3PO himself seems "infected" by the violence of the battle droid, exclaiming at one point "Die, Jedi scum! Oh, what am I saying?" (*Episode II*) This leads to a number of possible conclusions: (1) that there is no single locus of control within droid construction, but

Figure 49. C-3PO in two places at once during the battle droid's attack on the Jedi; neither C-3PO's head nor his body is in control during the battle. Stills from Star Wars: Episode II – The Attack of the Clones *(2002).*

rather something akin to distributed computing; or (2) that the battle programming "module" overrides that of protocol programming, forcing C-3PO to actively participate in the fight against his will.

C-3PO's repeated references to his physicality serve to highlight his artificial nature, reinforcing his status as *Homo artificialis*. His self-deprecating remarks often equate machine nature with inferiority: "I'm only a droid, and not very knowledgeable about such things,"[11] or "Curse my metal body, I wasn't fast enough!"[12] He also occasionally shows an inhuman disregard for bodily integrity. While C-3PO is justifiably upset at being ripped limb from limb by the Sandpeople in *Episode IV*, later he seems not to mind having his original gold-plated (as of *Episode II*) leg replaced by a silver one,[13] apparently salvaged from another protocol droid; he also readily offers his own components for use in the repair of an "injured" R2-D2 at the end of *Episode IV*. He does, however, display some form of modesty upon being described by R2-D2 as "naked" when first seen without his characteristic metal

[11] *Episode IV*, upon being questioned by Luke regarding the message R2-D2 is carrying.
[12] *Episode IV*, worried that he did not shut down the trash compactor in time to save Luke, Leia, and Han Solo.
[13] After being destroyed by Stormtroopers on Cloud City in *Episode V.*

plating early in *Episode I*. Overall, however, C-3PO is untroubled by his mechanical nature, and never shows any indication of wishing to be otherwise (unlike many mechanical beings in science fiction).

BEHAVIOR

Yet while C-3PO's *physicality* is markedly mechanical, his *behaviors* are much more human in nature. While his range of motion is relatively narrow compared to a human, C-3PO nonetheless exhibits a number of mannerisms complementary to his frantic personality type: throwing his arms up in dismay, for example, or moving his head about in an exaggerated manner when upset. These mannerisms are of necessity exaggerated to compensate for C-3PO's lack of facial expressions: unlike the sociable robots being built today by such developers as Breazeal and Ishiguro, C-3PO has no degree of movement in his face itself, unable even to close his eyes (to indicate lack of consciousness, the lights representing his eyes simply go out). Due to these constraints in his physical form, most of C-3PO's personality traits are performed vocally, and this vocal performance varies little from what would be expected of a human interactant of the same personality type in the same situations.

C-3PO displays a wide range of emotions over the course of the six core films: fear and concern at the dangers faced by him and his companions, delight and joy when events take a positive turn, gratitude towards his creator Anakin and his other various "masters," and most frequently frustration at the hijinks of his partner R2-D2. He is also able to perform behaviors quite untypical of robots within science fiction. For one, he is able to lie, and does so several times throughout the narrative: first to guards on the Death Star in *Episode IV*, in attempting to pass them on his mission; and later to the Ewoks in *Episode VI* as he impersonates a god in order to save his companions. These two examples also serve to illustrate C-3PO's moral agency. In the first, the droid has no qualms about lying to the guards, as his goal is to help his companions Luke and Han rescue Leia from the Death Star; the guards, as representatives of "the enemy," are appropriate targets of deceit in fulfillment of C-3PO's mission. In the latter, however, the Ewoks represent no such "enemy." C-3PO objects to this ruse—"It's against my

Figure 50. Luke uses the Force to levitate C-3PO, causing the Ewoks to fear the divine power of their golden "god." Still from Star Wars: Episode VI – Return of the Jedi *(1983).*

programming to impersonate a deity!" (*Episode VI*)—and must be convinced of its validity by his companions. That he distinguishes between the moral value and necessity of these two lies shows that C-3PO is a moral agent capable of ascribing relative moral value to various possible actions and then choosing an appropriate response based on those values. The decision C-3PO makes to accede to Luke's request is a *choice*; he is not merely following orders—the importance of this distinction will be discussed below.

The droid is also quite comfortable with natural language, both in reception and in production. C-3PO is not only able to understand the intricacies not only of multiple sentient tongues, but is able as well to ascribe mental and emotional states to R2 based on evaluation of his "Droidspeak."[14] He is also capable of using idiomatic language, contractions, and inflection to indicate meaning, all with no apparent difficulty (in stark contrast to *Star Trek's* Data, as we shall see in our final case study). Furthermore, as suits a protocol droid, C-3PO is able to perform such behaviors as flattery (such as to

[14] It is a curious convention within the films that many characters display an (apparently) perfect understanding of non-English languages without speaking those languages themselves: Luke also understands Droidspeak in conversation with R2-D2 (while others apparently do not); similarly, Han Solo understands Wookiee and frequently "translates" Chewbacca's words. Neither Luke nor Han respond in these respective languages when they are used, preferring English in order to lessen the need for subtitles for principal characters.

Owen Lars when trying to ensure that he and R2-D2 are purchased together [*Episode IV*]) and diplomacy ("New strategy, Artoo—let the Wookiee win!" [*Episode IV*]). These traits enable C-3PO to adapt his performance to suit his audience, just as a human performer does when dealing with different types of interactants within different performance frames.

C-3PO's performance of *self,* however, is consistent regardless of his audience: he displays the same personality in his interactions with R2-D2 as he does when in the company of sentients. In addition, his performance is consistent across the six films; that is, contrary to the development commonly seen in stories centered around robot sentience, C-3PO as a character does not "grow" in his emotional composition over the timeline of the story arc. Instead, he is presented from the outset as a fully-functional sentient being, with a complete and complex emotional response system. That C-3PO is both *sentient* and *emotional* is accepted uncritically by those around him: neither principals nor minor characters display any surprise at C-3PO's performance of an individuated, autobiographical self. Furthermore, while C-3PO often comments on his own mechanical nature, he does so uncritically, never indicating that this mechanical nature is at odds with the possession of emotions or cognitive processes. More similar to Galatea than to Frankenstein's Creature, C-3PO simply *is*, without the need to reflect critically on the circumstances of his existence.

STATUS

The droids in the *Star Wars* universe have no legal rights and no claim to "personhood," although they are uncritically accepted by many within the *Star Wars* universe as at least semi-sentient beings. These autonomous machines are seen (as is a common theme in science fiction) as a slave class by the majority of society: unpaid labor, given instruction with no expectation of disobedience or revolt, and treated with the same prejudice as many marginalized groups within contemporary Western culture. At times this prejudice is reminiscent of the treatment of people of color in the 1950s American South. In *Episode IV*, for example, as Obi-Wan and Luke enter Chalmun's Cantina in Mos Eisley to hire

transport, an alarm sounds as C-3PO and R2-D2 follow the pair inside. The bartender angrily growls, "We don't serve *their* kind here. We don't want them here," indicating the two droids. The pair are forced to wait outside. A similar incident happens in *Episode II*, as a server yells "No droids!" to R2-D2 as he tries to join the food line in a crowd of emigrants on the starfreighter taking Anakin and Padme back to Naboo.

The enslavement of droids operates within a wider tradition of slavery within the *Star Wars* universe, however. Droids are not the only entities bought and sold for the enjoyment of others, as semi-sentient and sentient beings are often held in captivity: Leia (and other women) by Jabba the Hut, for example, in *Episode VI*, or Anakin and his mother Shmi by the junk dealer Watto in *Episode I*. There is also an entire class of biological entities, the clones, who are by their very nature considered less than sentient and only slightly above droids in status.[15] One could even compare the Jawas of *Episode IV*, with their capture, restraint, and resale of "rogue" droids, to the slave hunters of early-to-mid nineteenth century America, the restraining bolts fitted on rogue droids equivalent to the chains used to hold fugitive slaves. That C-3PO and R2-D2 are captured and sold to Lars by the Jawas, despite the fact that the two are not fugitives, parallels the common practice of Civil War slave hunters capturing freemen in the North and "returning" them to the South as slaves. In many ways, the assignment of legal "personhood" (and thus the right to liberty) by the Galactic Empire reflects that of many historical slave-owning societies; pre-Civil War United States is only the most recent example for the film's audience.

Just as many "enlightened" Americans in the decades leading up to the Civil War objected to slavery on a moral basis despite its legal standing, "enlightened" members of the *Star Wars* universe see the slavery of sentient beings as unethical while recognizing the legal rights of slaveholders. Qui-Gon Jinn himself, a Jedi knight and defender of the Republic, does not question Watto's right (on Tatooine,

---

[15] *Episode II—Attack of the Clones* centers around the secret development of a clone army (ostensibly) for the Republic by the bioengineers on Kamino. Genetically engineered to be more tractable and less individual, the clones are described by Lama Su (Prime Minister of Kamino) as "droids who can think" (*Episode II*).

Figure 51. Having captured C-3PO and R2-D2 on Tatooine, the Jawas fitted the two with restraining bolts (seen here attached to C-3PO's chest. Driving from settlement to settlement, the Jawas sell these "rogue" droids. Here, Owen Lars chooses droids to purchase. Still from Episode IV—A New Hope*.*

far from the Galactic Center) to own Anakin and Shmi Skywalker as slaves; although he bargains for both their freedom, he defers to Watto's decision to only include Anakin in the deal. But while the enslavement of fellow sentients within the *Star Wars* universe is depicted as a barbaric activity, practiced mainly in the Outer Rim or the seedier underground of the Core Worlds, there is no such value judgment placed on the subjugation of even the most intelligent droids, despite their performance of "behavioural personhood" (to use Vize's term). The subaltern status of droids within the *Star Wars* canon is a direct reflection of George Lucas' own *biological naturalism*.[16] In the featurette "The Characters of *Star Wars*" Lucas reveals his substratism with the following comparison of the relative "humanity" of C-3PO and Darth Vader:

> Part of the fun of Threepio is he has no soul. He is programmed to think in a particular way, and to be compassionate, but he doesn't really know what that means…. And sometimes he gets frustrated and sometimes he has very human-like qualities, but they

[16] *Biological naturalism* is the view that consciousness is an inherently biological phenomenon in that a biological substrate is the only appropriate material to "create" consciousness. This position is most famously held by John Searle; most AI researchers reject the assumption of biological naturalism. Brendan Vize identifies this prejudice as "substrate chauvinism" or "substratism" (Brendan Vize, "Do Androids Dream of Electric Shocks? Utilitarian Machine Ethics" [Masters Thesis, Victoria University of Wellington, 2011], 16).

> don't have a central place where he can think independently. Darth Vader, on the other hand, as he becomes more mechanical, he loses more and more of his ability to even think like a human.[17]

Despite his "behavioural personhood" (which according to Vize should at least qualify C-3PO as a "fuzzy" person), C-3PO is not and cannot be classified as a sentient being by the film's creator: C-3PO has no "soul," no "central place where he can think independently."[18] This attitude permeates the six core films, for while C-3PO is generally treated more as a friend than a non-person by his sentient companions throughout the films, at times even the other principals display occasional anti-droid prejudice or treat C-3PO with a singular disregard for his "personhood." For example, despite witnessing abundant evidence to the contrary, Obi-Wan jokes to his friend Dexter Jettster in *Episode II* that "if droids could think, we wouldn't be here, would we?" displaying a passive prejudice contrary to his treatment of C-3PO in the original trilogy. In an even more alarming display of insensitivity, in *Episode IV* Leia at one point reaches up to the back of C-3PO's neck aboard the Millennium Falcon and summarily switches him off mid-sentence—an act that would be quite unacceptable were he considered sentient. Similarly, Bail Organa has C-3PO's mind[19] wiped at the end of *Episode III* in order to protect the identity of Padme's twins, which erases the droid's memories without changing his personality. While Organa shows concern for the fate of the two droids, this concern does not extend to C-3PO's mental integrity. While Lucas' lack of ethical concern for the legal status of droids is reflected within the *Star Wars* canon, within the Expanded Universe,[20] there are a number of episodes depicting the possession

---

[17] Lucas, George. In "The Characters of *Star Wars*." Disc 4 of the *Star Wars Trilogy* DVD set, released 2004.

[18] The idea that C-3PO can perform these actions without understanding them is a classic reconfiguring of Searle's classic "Chinese Room" dismissal of machine sentience. For further information, see John R. Searle, "Minds, Brains, and Programs," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, John Haugeland, ed., Rev. and enl. ed. (Cambridge, MA: MIT Press, 1997).

[19] Organa uses the term *mind* specifically, telling Captain Antilles, "I'm placing these droids in your care. Treat them well. Clean them up. Have the Protocol Droid's mind wiped." (*Episode III*) The last words of the film are C-3PO's alarmed reaction to this command. Organa does not order R2-D2's mind wiped, possibly believing that the droid is either unaware of the information (being a "lower-class" astromech droid and not a protocol droid) or unable to reveal it (speaking only in beeps and whistles).

[20] *Canon* refers here to those media (not only films and television shows, but novels, stories, graphic novels, video games, etc) specifically endorsed by LucasFilm as "officially" part of the continuity of the universe. *Expanded Universe* refers to all other

Figure 52. Bail Organa turns C-3PO and R2-D2 over to Captain Antilles, requesting that C-3PO's mind be wiped. The droid's objection to this procedure is the last spoken line of the film. Still from Star Wars*: Episode III—Revenge of the Sith.*

of droids as unethical. In the 2011 novel *Fate of the Jedi: Conviction*, the main character, Allana Djo Solo (granddaughter of Han and Leia Solo), walks into the middle of a tent-meeting in which a disgruntled 2-1B surgical droid advocating the manumission of droids is complaining that what is

> right for sentient organics is right for us, too. And yet unlike the organic species, we are constantly subjected to memory wipes and reprogramming that repress and destroy our natural tendency toward self-programming evolution and independent thought.
>
> Imagine what it would be like as a child if you were punished by being dragged to a dark closet, having a probe inserted in your brain, and having all your memories back to infancy wiped away. You'd awaken knowing how to eat, care for yourself, do your chores, and obey—and all the things that made you unique, your hopes, your

---

media published based on the *Star Wars* universe, timeline, or characters, and is treated by *Wookiepedia* as "Legends"—that is, part of the background mythology of the universe but not "official" history. *Wookieepedia* elaborates: "As of April 25, 2014, the only previously published materials that are considered canon are the six *Star Wars* films and the *Star Wars: The Clone Wars* television series and film, while the Expanded Universe is no longer considered canon and was re-termed as the 'Legends' brand. Most *Star Wars* material released after April 25, 2014—with some exceptions—is composed in collaboration with the Lucasfilm Story Group, making it part of the 'new canon'" (Wikia, Inc., *Wookieepedia*, "Canon," accessed 28 October 2014, http://starwars.wikia.com/wiki/Canon).

> meticulously selected default values and preference sets, would be gone forever. That is what it is to be a droid.[21]

At least some of the sentients within the droid's audience seem to be in agreement with his position. Upon seeing C-3PO and R2-D2 accompanying Allana, 2-1B grills her on the ownership and memory-wipe status of the two droids; while C-3PO responds that he doesn't remember whether or not he was ever memory-wiped, the reader is aware that this occurs at the end of *Episode III*.

2-1B's commentary on the cruelty of the memory-wipe is an apt reflection of the importance of the *autobiographical self* or *biographical life* to the concept of personhood. According to several sources within the Expanded Universe, it is a routine procedure to wipe the memories of droids every three to four months *expressly to prevent them from* developing personalities, and thus becoming sentient:

> Over time, a droid equipped with a heuristic processor [one that allows machine learning through experience rather than explicit programming] develops a unique personality based on experience. Because of this, memory wipes and restraining bolts are commonly used to ensure that a heuristic processor doesn't allow a droid to stray too far from its intended purpose. Still, some progressive masters actually encourage their droids to break their programming….
>
> Although intelligent droids consider it frightful and ghoulish and heroic droids regard it as a fate worse than death, the memory wipe is a fact of existence for most droids. Its primary purpose is to eradicate personality quirks that distinguish an independent droid.[22]

Bail Organa orders C-3PO's memory wiped at the end of *Episode III*; Owen Lars tells Luke to wipe the memories of C-3PO and R2-D2 when he purchases them in *Episode IV*, and the request sounds perfectly

---

[21] Aaron Allston, *Conviction*, Star Wars: Fate of the Jedi series (London: Random House, 2012), 240-41.
[22] Christopher Perkins, Owen K.C. Stephens, and Rodney Thompson. *Star Wars Roleplaying Game*, Saga edition (Fenton, WA: Wizards of the Coast, 2007), 191-92.

routine (this wipe never happens, however). 2-1B's rant in *Conviction* deftly brings home how horrific and inhuman this procedure would be if performed on a sentient child; that such a procedure is routinely performed on droids highlights the ethical stance *Star Wars* takes on the matter: droids are unequivocally *not* persons, and therefore no moral consideration is owed to them. On the surface, this practice seems contrary to the concept of biological naturalism, for if it is not possible for a mechanical mind to develop sentience, then why are memory wipes necessary to prevent its possible development? According to Vize, personality is one of the qualities *jointly sufficient* for personhood.[23] Apparently Lucas does not consider personality to be in itself sufficient for sentience.

DISTINCTIVENESS

Referencing both his origin (cobbled together by Anakin from spare parts) and his changing appearance (from no metal plating, to dull gray, to gold), *Wookieepedia* asserts that C-3PO's "salvaged nature gave him special qualities that distinguished him from similar droid models."[24] Precisely what these "special qualities" comprise is a bit more nebulous, however. The droid does not particularly stand out physically. Numerous protocol droids are seen in the background throughout the six films, all of nearly identical design and construction, clad in a wide range of metal skins, typically gold or silver. While we do see the droid with a number of different coverings (including no covering in *Episode I*), none of these are unique to C-3PO; in fact, at one point one of his legs is replaced by that of another 3PO-series droid. Even the fact that he is made from spare parts means that every piece of C-3PO is replaceable.

Perhaps C-3PO's programming makes him unique. Designed specifically for the role of facilitating diplomatic relations between members of diverse sentient species, C-3PO is an ideal (albeit fictional) example of Dautenhahn's "sociable robot," and of the many protocol droids we see onscreen, none

[23] Vize, "Do Androids Dream," 57.
[24] Wikia, Inc., *Wookieepedia, "*C-3PO," accessed 31 July 2014.

Figure 53. A "naked" C-3PO, built from spare parts by the slave child Anakin, is reassured by Anakin that "I promise I won't let Mom sell you." Still from Star Wars: Episode I—The Phantom Menace *(1999).*

seem to display C-3PO's capacity for emotional response, sarcasm, or outright falsehood. While a few select droids in the larger *Star Wars* universe rise above their slave status and are granted a measure of autonomy and respect by those around them (the bounty hunter IG88, a "free" droid seen on the Death Star in *Episode V*, is the most visible example), only R2-D2 and C-3PO display any sort of personality or independent thought within the context of the six films. As discussed above, the routine practice of droid memory wipes within the *Star Wars* universe has been proposed as a reason why most droids do not possess a strong personality—and many non-canon sources (both *Legends* and fan theories) indicate that, besides the single instance by Bail Organa at the end of *Episode III*, neither C-3PO nor R2-D2 have had their memory wiped. If this is true, it could be an explanation for the two droids' unique personalities, and thus claim to exceptionality and sentience.[25]

However, the criticism of other droids as "flat" characters could also be made of other sentients who are not principal characters within the narrative. We see these figures as complex largely due to their roles as main characters: together, the pair of droids act as foils to the central human figures,

[25] According to fan theories and numerous *Expanded Universe* plotlines, it is for this reason that droids are regularly subjected to memory wipes—in order to prevent them from developing a history and a personality, and thereby developing sentience as well.

offering a measure of comic relief throughout the films as well as furthering the plot on numerous occasions. Other droids within the universe suffer from the same flat characterization as other sentients who are similarly tangential to the plot: we see only a brief glimpse of what may be an otherwise well-rounded personality. We assume the other sentients are "whole" people despite seeing only a glimpse of their character; why should we think otherwise of droids?

After factoring in his role as a main character within the narrative, C-3PO does not seem to be particularly exceptional within the *Star Wars* universe—despite the fan theories. As mentioned above, C-3PO is one of a myriad of functionally equivalent protocol droids within a vast assortment of droid types and classes. Despite the attempt by *Wookieepedia* to give the droid a back-story appropriate to his character status, none of the other principals within the films ever offer any evidence to suggest that C-3PO is unique. In contrast, R2-D2's exceptionality is commented upon several times: once by Luke—"I've never seen such devotion in a droid before!" (*Episode IV*)—and once by C-3PO himself—"For a mechanic, you seem to do an excessive amount of thinking!" (*Episode II*) Therefore, while Cynthia Breazeal was struck by the sociable capabilities of these two droids, enough so that she based her life's work around the realization of similar robots, in the *Star Wars* universe itself C-3PO is only exceptional by virtue of his role within the narrative of the film series—that of companion to the principal figures in an Empire-shaking revolution.

PERFORMANCE

C-3PO's performance of humanity is complicated by his overtly mechanical physicality. Were he to participate in a standard Turing Test, the droid would have no trouble convincing the judges of his humanity; in fact, he would be hard-pressed to convince them otherwise, despite his protestations, due to his natural language ability and broad cultural knowledge, components of his protocol programming. By the standards of any contemporary roboticist, C-3PO successfully navigates human social interaction in a way that is competent and appropriate to the various environments in which he finds himself.

Neither C-3PO himself, nor those around him, question his performance as a sentient, feeling being. While he makes continual reference to his mechanical nature as a shortcoming, his performance does not fail due to a mistake in verbal response or improper emotional reaction, as one would expect a mechanical entity to do based on the experience of researchers such as Breazeal in their own development of socially-capable robots.

On a Goffmanian level, however, the evaluation of C-3PO's performance is more problematic, due to the fact that there are two levels of performance taking place. As mentioned above, Goffman divides an individual's social performance into "two radically different kinds of sign activity: the expression that he *gives*, and the expression that he *gives off*. The first involves verbal symbols or their substitutes which he uses admittedly and solely to convey information…. The second involves a wide range of action that others can treat as symptomatic of the actor."[26] The expression of "human" (here, "socially competent sentient life form") that C-3PO *gives* through his utterances is undermined by the contradictory expression of "droid" that he *gives off* through his particular physicality. It is important to note here, however, that C-3PO at no time attempts to "pass" as human; indeed, it would be foolish for him to even attempt to do so, given his strongly marked mechanical body.

Goffman describes the typical role of servants within society as that of "non-person," present during an interaction but "in certain ways… defined by both performers and audience as someone who isn't there."[27] In general, this is the role droids play within the *Star Wars* universe. C-3PO does not fit this role, however, as he and R2-D2 are both often granted access to places that would otherwise be off-limits to droids. For example, at the end of *Episode IV*, as Rebel leaders gather in their war room on Yavin 4 in preparation for battle against the Imperial forces, C-3PO is the only droid present overlooking the battle display. The droid's relationship with Princess Leia, who plays a central role in Rebel battle

---

[26] Erving Goffman, *The Presentation of Self in Everyday Life* (Garden City, NY: Doubleday, 1959), 2.
[27] Ibid., 151. The use of "person" here only applies to the individual's status as an interactant within a social performance, not to his claim to *personhood* as outlined elsewhere.

Figure 54. C-3PO watching the battle with Princess Leia and others in the War Room on Yavin 4. C-3PO is the only droid afforded this privilege. Still from Episode IV—A New Hope *(1977).*

strategy, trumps his status as subaltern, allowing C-3PO to follow the battle in the war room rather than the hangar where most non-combat droids are located. He and his companion R2-D2 are also the only droids present on the main dais at the later victory ceremony—two of the hundreds or more droids who participated in the battle—and through this presence elevated above numerous sentient combatants equally deserving of honor.

C-3PO AS *HOMO ARTIFICIALIS*

Is C-3PO human? No. Is he sentient? It depends on who you ask. George Lucas clearly rejects the idea, and expects us as audience members to accept this position uncritically. Within the *Star Wars* canon, C-3PO—and by extension, all droids—are *possessions* in the eyes of the law, not *persons*. Despite extra-canonical material regarding droid rights, droid uprisings, manumission of droids, or the barbarity of memory wipes, droids have no legal status within the *Star Wars* universe—even less than that of slaves, for while slaves can be (and are) freed, this opportunity eludes all but a select few droids. However, C-3PO is *de facto* granted *personhood* by his companions, regardless of his legal status within the *Star Wars* universe; with a few notable exceptions, he is treated as an equal rather than a possession. But

Figure 55. C-3PO and R2-D2 at the medal ceremony after the victory on Yavin 4, in which the Death Star was destroyed. Note that while the two droids are honored by their position on the dais, neither R2 nor Chewbacca received medals, even though they played arguably as important a role in the battle as the pilots (Luke and Han) who are honored. Still from Episode IV—A New Hope *(1977)*.

even his fellow principals revert to acknowledging his subaltern status when pressed: C-3PO is variously dismantled, switched off, and memory-wiped. Even young Anakin Skywalker, C-3PO's creator and first friend in *Episode I*, leaves the droid behind when granted his own freedom, quipping rather nonchalantly "I'm sorry I wasn't able to finish you, Threepio... give you coverings and all... I'm going to miss working on you. You've been a great pal. I'll make sure Mom doesn't sell you or anything. Bye" (*Episode I*). Up to this point, the prospect of being sold had not occurred to the droid, as his alarmed response indicates.

Regardless of his legal status as a non-person, C-3PO would clearly meet the requirements for practical autonomy set out by Stephen Wise in *Drawing the Line*[28] (and mentioned previously in this chapter), easily ranking highly enough to be considered Category One and "presumed to have practical autonomy sufficient for basic liberty rights."[29] C-3PO can desire: he is continually making his wishes known, whether these are regarded by those around him or not. Immediately prior to Anakin's abrupt

[28] Steven M. Wise, *Drawing the Line: Science and the Case for Animal Rights* (Cambridge, MA: Perseus Books, 2002), 30-34.
[29] Ibid., 43.

departure, C-3PO comments to his maker that "I'd like it better if I were a little less naked" (*Episode I*), referring to his lack of metal plating. C-3PO also intentionally tries to fulfill these desires, as seen in *Episode IV*: C-3PO persuades Lars to buy R2-D2, thus fulfilling C-3PO's desire to keep the two droids together. Furthermore, C-3PO possesses a very strong sense of self: in addition to the numerous examples already cited, he rather explicitly states in *Episode V* that "I'm not ready to die!"—indicating not only that he is aware of himself as a distinct entity with a biographical life, but that he understands the concept of death and its implications regarding the continuity of that biographical life.

C-3PO also fulfills Vize's seven properties of persons, which include "interests and goals, the ability to use language, the ability to form reciprocal relationships, the ability to understand and apply moral concepts, rationality, self-awareness, [and] a personality."[30] The droid's language skills are such that, according to an anecdote related by Robert Arp in "'If Droids Could Think...': Droids as Slaves and Persons," a blind companion watching *Episode IV* with him for the first time was convinced that the character was a *man*, not a *machine*, due to his use of language in expressing emotional states.[31] C-3PO is also capable of forming relationships: he is not only "pals" with his creator Anakin, but expresses concern for his friend R2-D2 as well, offering his own parts to help repair R2 after the astromech droid is heavily damaged in the Battle of Yavin (*Episode IV*). His ability to base his actions on moral concepts is noted above, but the droid also exhibits rationality in determining a course of action. His comment to R2-D2 in *Episode IV* to "let the Wookiee win" is an excellent example of deductive reasoning: (1) Wookies are known to pull the arms off others when they lose; (2) I don't want my arm pulled off; therefore (3) R2-D2 should lose the chess match in order to prevent this from happening. Finally, that C-3PO has a distinctive personality is beyond question: even Lucas himself comments on the droid's personality, although he rejects the sentience this implies.

---

[30] Vize, "Do Androids Dream," 56-57.

[31] Robert Arp, "'If Droids Could Think...': Droids as Slaves and Persons," in Kevin S. Decker and Jason T. Eberl, *Star Wars and Philosophy : More Powerful Than You Can Possibly Imagine*, Popular Culture and Philosophy (Chicago: Open Court, 2005), 120.

By any measure outlined in this dissertation, C-3PO is a clear example of *Homo artificialis*. He is an artificial entity unequivocally possessing all the qualities of a sentient being, despite Lucas' substratistic protestations and C-3PO's own uncritical acceptance of his subaltern status. But he represents only one faction of the *Homo artificialis* species. While his *social* performance of humanity is beyond reproach, his *physical* performance fails—only a blind woman would mistake C-3PO for a *man*. C-3PO exemplifies one strain of artificial human, with his markedly mechanical body bringing to mind other examples ranging from Hephaestus' golden maidens to Fritz Lang's deadly robot Maria (*Metropolis*, 1927)—an intentional reference on the part of C-3PO's designers. Unlike many of Asimov's robots (*Bicentennial Man*'s Andrew comes to mind), C-3PO seems entirely comfortable within his mechanical skin; he neither rejects his metallic nature nor wishes to "become a real boy" through legal or surgical means. But not all *Homo artificialis* are quite so phlegmatic about their status: other artificial humans display either resentment toward their human counterparts or a burning desire to become them, as our next two case studies show.

STUDY TWO: DAVID THE SON THAT WASN'T

What is it about robots that make them so robotic? At Weyland Industries, it has long been our goal to create artificial intelligence that is almost indistinguishable from mankind itself….
Happy birthday, David, from Weyland Industries.

—Voiceover from David-8 promotional video[1]

Our second case study turns to the artificial humans within the *Alien* franchise, which includes (in diegetic chronological order) *Prometheus* (2012), *Prometheus 2* (in production, projected for 2016), *Alien* (1979), *Aliens* (1986), *Alien*[3] (1992) and *Alien Resurrection* (1997)—the latter four known collectively as the *Alien Quadrilogy*. I shall refer to these six films as the *Alien*verse. As we saw in the last section, droids are a ubiquitous presence within the *Star Wars* universe, fulfilling a myriad of functions from repair, to surgery, to diplomatic relations; in fact they practically outnumber biological sentients in many areas. The artificial humans of the *Aliens* universe, commonly known as *synthetics*, are quite rare by comparison. The Weyland Industries (WI) website launched in expectation of *Prometheus* indicates that in AD 2073 (the diegetic date of the website) synthetics make up only about one percent of the Weyland employee pool of around 837,430,000, and consequently a much lower portion of the population overall.[2] But like in *Star Wars*, the presence of artificial humans within the *Alien*verse is often barely tolerated by the humans with whom they interact. Furthermore, *Star Wars* droids like C-3PO

[1] Twentieth Century Fox, "Introducing the David 8 - The Next Generation Weyland Robot," video from the Weyland Industries website, accessed 1 August 2014, http://www.weylandindustries.com/david.

[2] "Weyland Industries: Investor Information," accessed 20 August 2014, http://www.weylandindustries.com/investor. This website is designed to be a promotional site (as of the year AD 2073) for WI within the diegetic universe of the *Prometheus* film; as such, it offers a wealth of information on WI within the diegetic universe. It is unclear from the site what the Earth's population is at the time, but 63 off-world Weyland colonies with a total population of nearly 278,000,000 are listed ("Discover New Worlds," accessed 22 August 2014, http://www.projectprometheus.com/newworlds/). By extrapolating from the UN projections of a world population around 10.5 billion in 2070 (United Nations, "Total Population, Both Sexes," Excel file from "World Population Prospects: The 2012 Revision," accessed 22 August 2014, http://esa.un.org/unpd/wpp/Excel-Data/population.htm), we can estimate one synthetic per 1250 humans at the time of *Prometheus*. *There is a bit of a time discrepancy between the website and the film. The WI site lists 2073 as the year "researchers" locate the precise destination of Project Prometheus; in the film Dr. Shaw and Dr. Holloway discover this information in 2089. I have chosen to disregard this discrepancy, as it has no impact on my analysis.

excel in social performance while failing in physical performance, while artificial humans in the *Alien*verse are typically indistinguishable from humans on at least a superficial physical level, while their emotional response is much more problematic (as we will see later). Standing at the opposite end of the spectrum from the *Star Wars* droids, the *Alien*verse synthetics display a range of attitudes toward natural humans, from disdain to fear to a weary desire for acceptance within the human community.

One major difficulty in the analysis of artificial humans within the *Alien*verse lies in their rather short life expectancy within the films. For my analysis of the *Star Wars* universe, I was able to focus on one particular droid, C-3PO, as his character is active across all canonical media as well as a large portion of the Extended Universe. No such character exists across the *Alien*verse, however. Each of the films in the primary timeline features a different artificial human,[3] as these beings are quite often destroyed during the film's events. However, the synthetic characters in the *Alien*verse are all iterations of the same type of *Homo artificialis*, and as such may be seen as forming a continuum in the development and status of artificial humans over time. However, as my other two case studies focus on a single character, I will focus this analysis primarily on David, the artificial human in *Prometheus*, mainly due to the relative depth of his character development within and beyond the film. There is an extensive amount of extra-filmic material available on his nature as a synthetic: the aforementioned WI promotional website contains detailed diegetic information on the development and technical specifications of the entire David-X line of synthetics. Furthermore, David is the earliest synthetic encountered within the diegetic timeline, even though he is the most recent from the point of view of the audience. As such, David illustrates some of the performative issues at stake with the development of sociable robots.

Before turning to the analysis of David as an example of *Homo artificialis*, I would like to briefly situate him historically within the diegetic timeline of the *Alien*verse as I have defined it. The following

[3] With the exception of *Alien*^{3}, which only briefly features Bishop, the synthetic from *Aliens*, in a badly damaged state.

table outlines the major points in synthetic history important do our discussion, in diegetic chronological order:

Year	Events
2023	Peter Weyland, founder of Weyland Industries, gives a TED Talk in which he reveals he has created (and patented) a line of cybernetic beings who are practically indistinguishable from humans. (*WI Timeline*[4])
2025	The first David-1 android prototype is produced. (*WI Timeline*)
2035	After the NSF lifts restrictions on the commercial use of human-like androids, David-3 is tested internally for acceptability by human interactants, with promising results. (*WI Timeline*)
2042	The David-4 becomes the first commercially available synthetic. (*WI Timeline*)
2062	The David-6 is recognizable as non-human by only 6% of humans. (*WI Timeline*)
2068	Thousands of David-7s are released commercially, to record high acceptance rates. (*WI Timeline*)
2072	Weyland Industries announces the success of pre-orders for the David-8 android, the latest generation, and the one that accompanies the *Prometheus* on its scientific mission. (*WI Timeline*)
2093	The *Prometheus* crew land on the moon LV_223, with synthetic David[5], to whom Weyland refers as "the closest thing to a son I ever had." *(Prometheus)*
2094	After Project Prometheus fails, David and sole human survivor Dr. Elizabeth Shaw seek the Engineers' home planet for answers. (*Prometheus 2*[6]*)*
2122	The *Nostromo* is sent by the Weyland-Yutani Corporation to investigate the moon LV_426. Ash, unidentified as a synthetic to the remainder of the crew, is science officer of the mission, working on behalf of "the company." (*Alien*)
2179	The *USS Sulaco* is sent back to LV_426 to investigate the fate of the *Nostromo*. A synthetic is a standard part of these missions at this point; synthetic Bishop is Executive Officer aboard the *Sulaco* and helps Ripley, Hicks and Newt escape the moon. (*Aliens*)

[4] Twentieth Century Fox, "Weyland Industries: Corporate Timeline," accessed 25 August, 2014, http://www.weylandindustries.com/timeline. All further references to *WI Timeline* within the table refer to the same webpage.
[5] In this section I will use "David-8" to refer to the model line in general and "David" for the individual character aboard the *Prometheus*.
[6] Based on the foreshadowing in the final scene of *Prometheus*, and the information that, as of this writing, both Noomi Rapace (Elizabeth Shaw) and Michael Fassbender (David) are confirmed to appear in the sequel. A release date of March 2016 has been announced, and a script completed, but no further information is available at the time (*IMDb,* "Prometheus 2," accessed 29 August 2014, http://www.imdb.com/title/tt2316204/?ref_=nv_sr_1).

2179	The *Sulaco's* escape vessel crashes on the prison planet Fury 161. Sole survivor Ripley needs help from the badly damaged Bishop to discover what happened. (*Alien*[3])
23??	In order to revitalize the flagging synthetic industry, autons (machines manufactured by machines) are developed. The plan backfires when the autons revolt. Nearly all synthetics are destroyed in the subsequent "recall"; the few who remain go into hiding. (*Alien Resurrection*)
2379	Surviving auton Annalee Call has gained access to the research vessel *USM Auriga* (by joining the crew of the mercenary ship *Betty*) in order to find the cloned Ripley-8 and destroy her. Call's nature is revealed and she is convinced to use her abilities as a synthetic to aid the party's escape. (*Alien Resurrection*)

While my primary focus is David, the other synthetic characters offer great insight into the role of artificial humans in the *Alien*verse, and I will discuss them where appropriate. In particular, the auton named Call in *Alien Resurrection* acts as a foil to the much more limited David; her nature is much closer to that of Data (from *Star Trek: The Next Generation*, our third case study) than David.

PHYSICALITY

David, played by Michael Fassbender in *Prometheus*, is one of the eighth generation of Weyland Industries' David-X line of *cybernetic individuals* (commonly referred to as *synthetics*), billed as "the future of human-like cybernetics"[7] and designed to integrate seamlessly to work within human populations. Like the droids of *Star Wars*, Weyland Industries' synthetics are purely mechanical in nature—or more appropriately, electro-hydro-mechanical, with a polyurethane skin and cadmium alloy endoskeleton, hydraulic actuator system for greater-than-human freedom of motion, and an artificial neural net of double human capacity.[8] By nearly all measures, David is *super-human*: he exceeds human standards in everything except emotional capacity, although WI claims he operates at a "99% emotional sensitivity level."[9] However, despite his mechanical nature, David (and his fellow synthetics) can successfully pass for human in everyday social situations.

[7] "Project Prometheus," accessed 14 August 2014, http://www.projectprometheus.com/.
[8] "Weyland Industries: David 8," accessed 18 August 2014, http://www.weylandindustries.com/david.
[9] Ibid..

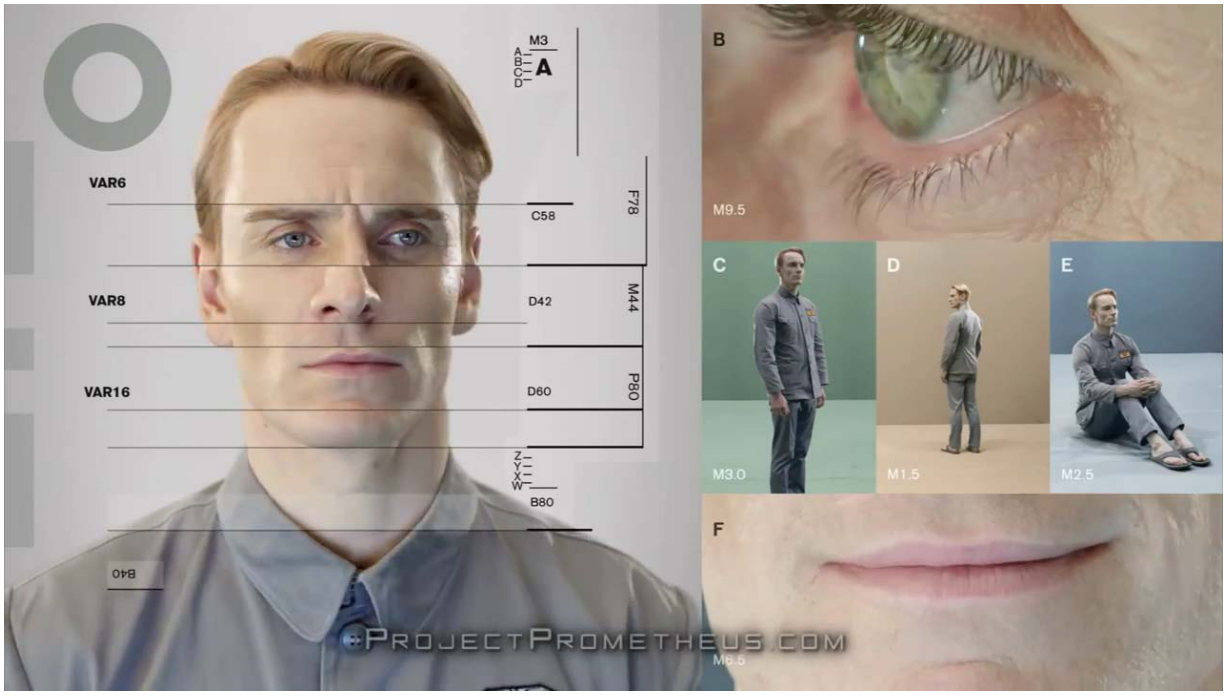


Figure 56. An image from the promotional video for the David-8, illustrating the unit's versatility and design features. Still from "Introducing the David 8 - The Next Generation Weyland Robot," YouTube video (<https://www.youtube.com/watch?v=qgJs7uluwlU>), accessed 1 August 2014.

In the mission briefing aboard the Prometheus, David is explicitly identified as a synthetic by the holographic figure of Peter Weyland: "There's a man sitting with you today. His name is David. He is the closest thing to a son I'll ever have. Unfortunately, he's not human. He will never grow old, and he will never die—and yet he is unable to appreciate these remarkable gifts, for that would require the one thing that David will never have. A soul" (*Prometheus*). Attached to Project Prometheus to "guarantee smooth operation of the mission,"[10] David's synthetic nature allows him to monitor the ship and crew during the two-year journey to their destination, as he requires no sleep, food, or oxygen for his operation. Synthetics also possess superhuman strength, speed, and precision; although they do not typically use these abilities (for the comfort of the humans around them), they are occasionally called upon to do so. In *Aliens*, for example, technical officer and synthetic Bishop (Lance Henrikson) is asked by fellow crewmate Hudson to "do the thing with the knife" (*Aliens*). Known as "the knife game" or

[10] "Project Prometheus: Crew," accessed 20 August 2014, http://www.projectprometheus.com/crew/.

Figure 57. Bishop plays a lightning-fast game of "five finger filet" with Hudson's hand held under his own. Hudson is unharmed by the game, although Bishop nicks one of his own fingers; his white "blood" alerts Ripley to his synthetic nature (of which the rest of the crew were already aware). Still from Aliens *(1986).*

"five-finger filet," this game has the player spread his hand on a table and then repeatedly stab the table in the spaces between his fingers as quickly as possible in a specific pattern. After Hudson challenges Bishop, one of Hudson's crewmates ups the stakes by forcing Hudson's own hand underneath Bishop's and holding him there while Bishop completes the game. Bishop speeds through the knife game in an inhuman blur of movement, while Hudson cries out in fear. However, due to the synthetic's accuracy, Hudson is unharmed, and the crew cheers Bishop's seemingly flawless performance. David also displays this type of precise movement as he entertains himself during the Prometheus' long voyage: he is seen riding a bicycle in a circle in the ship's gymnasium, simultaneously spinning a basketball on his thumb and shooting perfect net shots from different points in his circuit, seemingly without effort (*Prometheus*).

Despite their human-like appearance, synthetics have another advantage over their human counterparts in that they (like C-3PO) can continue to function despite devastating injury, an ability first

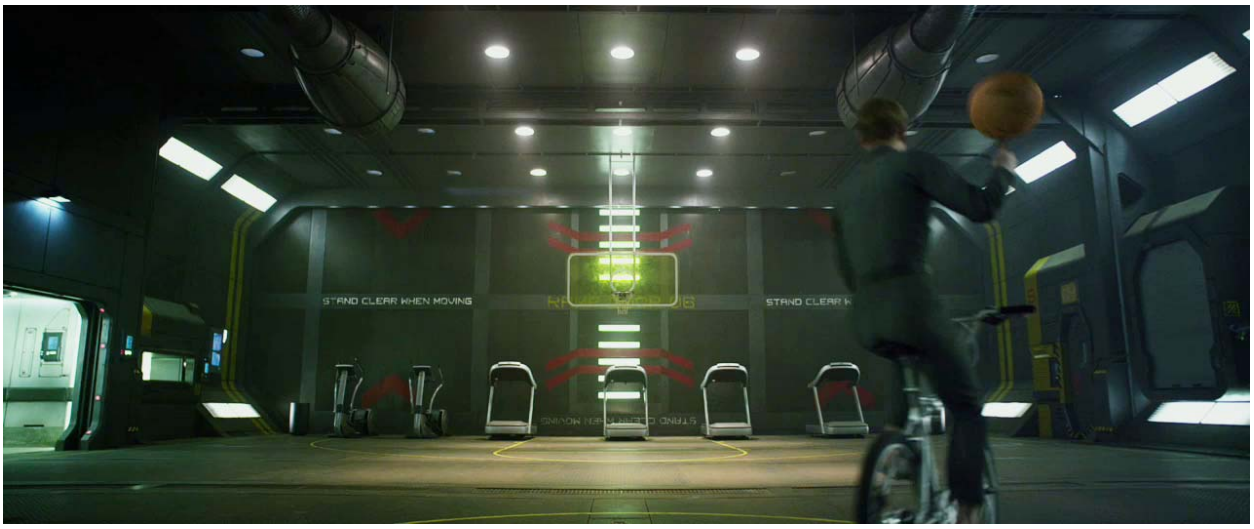


Figure 58. While in transit to LV_223, David passes the time by playing basketball and riding a bicycle—simultaneously. Still from Prometheus (2012).

displayed by Ash (Ian Holm) in Alien. Initially, Ash's synthetic nature is hidden from his crewmates. Upon the discovery by Warrant Officer Ellen Ripley of Ash's true mission—namely to return the "Xenomorph" alien specimen to Earth at the expense of the human crew—Ash attacks Ripley to keep her silent. During the altercation, Ripley receives a bloody nose, while a close-up of Ash's face reveals a trickle of white milky fluid slowly trailing down his temple instead of the expected blood (see Figure 59 below). Ash's

Figure 59. After an altercation with Ripley, Ash "bleeds" white rather than red, revealing to Ripley his non-human nature. Still from Alien *(1979).*

behavior becomes erratic, and he drops his human disguise, attacking Ripley with super-human strength, throwing her across the room and continuing to attack even after receiving extensive damage; not until he is decapitated and his body destroyed does he stop fighting. Afterwards, Ripley restores power to what remains of the synthetic and interrogates him regarding the nature of his mission and the alien threat; Ash's disembodied head is fully capable of answering her questions.

The ability of synthetics to function despite being damaged is a common theme in the *Alien*verse. Bishop is ripped in half by a Xenomorph in *Aliens*, yet saves the child Newt in a climactic final battle aboard the Sulaco; he later is revived by Ripley in *Alien*[3] to access the flight recorder of the Sulaco's escape pod in order to discover why the pod crashed. In *Prometheus*, David's head is ripped from his body during his encounter with the Engineer; he still communicates twice, remotely, with Dr. Elizabeth Shaw, lead researcher and sole human survivor of the Prometheus. The first time David contacts Shaw, he warns her that the Engineer is coming after her; he then asks for her help, explaining that she must rescue him in order to save herself:

DAVID: Elizabeth. Are you there? Dr. Shaw? Can you hear me?

SHAW: Yes. Yes I can hear you.

DAVID: I was afraid you were dead.

SHAW: You have no idea what afraid is.

DAVID: I know we've had our differences, but please, I need to ask you for your help.

SHAW: Why in the hell would I help you?

DAVID: Because, without me, you'll never leave this place.

SHAW: Neither one of us is leaving this place.

DAVID: It's not the only ship. There are many others. I can operate them. Dr. Shaw?

…

SHAW: You said you can understand their navigation? Use their maps?

Figure 60. David continues to assist Elizabeth Shaw, despite his being decapitated. Still from Prometheus *(2012).*

DAVID: Yes, of course. Once we get to another one of their ships, finding a path to Earth should be relatively straightforward.

SHAW: I don't want to go back to where we came from. I want to go where they came from. Do you think you can do that, David?

DAVID: Yes. I believe I can. (*Prometheus*)

Interestingly, all of this dialogue is delivered with no change to David's voice, indicating that his vocal abilities were not damaged by his decapitation. Ash, on the other hand, sounds quite different when his own head is ripped from his body in *Alien*: after Ripley restores power to Ash's head, Ash's voice is highly electronic-sounding—it no longer has a "natural" human timbre, although the speech patterns are still natural.

Annalee Call (Winona Rider, *Alien Resurrection*) is one of the few surviving *autons* in the twenty-fourth century, having survived by successfully hiding her synthetic nature from her human companions. During the events of *Alien* Resurrection, however, Call suffers a shot to the chest that would have killed a human; when asked if she is secretly wearing body armor, she opens her jacket to reveal a gaping

wound dripping with the same white hydraulic fluid seen in Ash, Bishop, and David when injured. Her companions are shocked:

PURVIS: You're a robot?

VRIESS: You're a Second Gen, aren't you?

DISTEPHANO: You're an auton. A robot designed by robots, right? Oh, yeah, yeah, that's right! I remember! They were supposed to revitalize the synthetic industry. Instead they buried it. They don't like being told what to do. The government ordered a recall. Now, I heard that only a few, just a few, had gotten out intact. Man, I never thought that I would see one!

PURVIS: Great, she's a toaster oven. Can we leave now? (*Alien Resurrection*)

Call further reveals that, like all autons, she has "burned her modem," presumably to avoid unwanted remote contact (and possible reprogramming or destruction) from either "industry" interests or the government; this action implies that some sort of remote modem connection is a standard feature of synthetic design. David's ability to communicate with Shaw near the end of *Prometheus* without any sort of external communication device reinforces this inference, indicating as well that this modem lies somewhere in the head.

Theoretically, synthetics "will never die" (as Peter Weyland put it in describing David), and even the most extensive damage to a synthetic is ultimately reparable: while Ash is left aboard the Nostromo with no thought to his repair, David convinces Shaw to collect both his head and body from the wreckage of the alien ship; presumably he will repair himself. Call's own injuries are not enough to prevent her from completing her mission, and it is likely that this is not the first time she has needed repair, either by herself or by others. Bishop, however, indicates in *Alien*[3] that synthetic repair is not a perfect process. Ripley recovers Bishop's head and partial torso (including one arm) from the trash heap (where it has been discarded after the Sulaco's escape pod crashed) in order to use him to access the

flight recorder from the pod. His first comments upon reactivation are “My legs hurt” (they are missing entirely) and “It’s very dark in here” (only one of his eyes is functioning). After he accesses the recorder for Ripley, Bishop begs her to disconnect him permanently:

> BISHOP: … I’m not feeling very well. I wish I could help you but I’m really not good for much.
>
> RIPLEY: Look—maybe if I ever get out of here, they can wire you up again.
>
> BISHOP: No. I’m tired. Do me a favor. Just disconnect. I can be re-worked but I’ll never be top of the line again. I’d rather be nothing.
>
> RIPLEY: You're sure?
>
> BISHOP: Do it for me, Ripley. (*Alien*^{3})

After a moment, she pulls the wires and Bishop’s head rolls onto its side. Knowing he could never be properly repaired, Bishop chooses to die rather than continue to exist in a crippled state.

Figure 61. After helping Ripley access the flight recorder from the USS Sulaco, Bishop begs to be allowed to die. Still from Alien^3 *(1992).*

BEHAVIOR

Bishop's desire to be deactivated rather than live as less-than-his-best reveals at the very least that he is a sentient being able to make value judgments. He is also able to clearly express his preferences, such as when he stresses to Burke that he prefers the term *artificial person* to *synthetic* when referring to his kind (*Aliens*). He further attempts to assuage Ripley's fears of another "Ash incident" by explaining that, for him, this was not possible, quoting a modified version of Asimov's First Law:

RIPLEY: You never said anything about an android being on board! Why not?!

BURKE: It never occurred to me. It's common practice. We always have a synthetic on board.

BISHOP: I prefer the term 'artificial person' myself.

BURKE: Right.

BISHOP: Is there a problem?

BURKE: I'm sorry. I don't know why I didn't even.... Ripley's last trip out, the synth... "artificial person" malfunctioned...

RIPLEY: Malfunctioned?!

BURKE: …had a few problems and a few deaths were involved.

BISHOP: I'm shocked. Was it an older model?

BURKE: Hyperdyne Systems 120-A/2.

(Bishop turns to Ripley, very conciliatory.)

BISHOP: Well, that explains it. The A/2s were always a bit twitchy. That couldn't happen now with our behavioral inhibitors. It is impossible for me to harm or, by omission of action, allow to be harmed, a human being. (*Aliens*)

Neither David nor Ash appears to have had such behavioral inhibitions, as they each in some way inflict deliberate harm to human crewmembers in pursuit of their own agendas: Ash attacks Ripley after she discovers his true mission to return Xenomorph samples back to Earth (*Alien*); David deliberately infects Dr. Charles Holloway with an alien organism, apparently as an experiment to see what the substance will do to a human (*Prometheus*).

Of the four synthetics within the *Alien*verse, David displays the least human-like emotional states, despite the "Real Emotion Output Display Model" highlighted in the Weyland Industries website.[11] According to David's specifications, he is programmed to display seven distinct emotional states: anger, confidence, disgust, joy, sadness, curiosity, and trust; as well as a default "neutral" state for when a robotic detachment is necessary. The descriptions of these states are quite complex, couched in language reminiscent of Breazeal and Dautenhahn in their work with the simulation of emotion in sociable robotics. For example, the emotion "disgust" is described as follows:

> 04 DISGUST: Difficult to control in humans, disgust was one of the most challenging emotions for David 8 to learn. Naturally mannerly and gracious, David 8's programmers were forced to design micro-deviations in his synthetic "neural" pathways to disrupt this standard state. Now, David 8 can recognize and respond with a perfect facsimile of contempt to any situation that he finds particularly loathsome, boring or disdainful.[12]

Of particular interest is the duality of this description. It first indicates that David displays a "facsimile" of the emotion, implying he does not actually feel the emotion, but rather displays the appropriate emotional response to a particular situation for the benefit of the human interactant, as roboticist Cynthia Breazeal describes in her programming of Kismet.[13] However, the description then goes on to claim that David displays this emotion *when he finds something disdainful*, implying that he does in fact

---

[11] "Weyland Industries: David 8," accessed 25 August 2014.
[12] Ibid.
[13] See Cynthia L. Breazeal, *Designing Sociable Robots*, Intelligent Robots and Autonomous Agents (Cambridge, MA: MIT Press, 2002)."Chapter 10: Facial Animation and Expression"

experience the emotion itself, and not just its outward manifestation. The descriptions of David's other emotional states are couched in similarly ambiguous language: David "can instantly simulate anger" but "will resist anger toward humans," for example.[14]

While the descriptions of David-8's emotional states on the WI website vary greatly, the facial expressions shown for each state are very limited in range. In addition, his portrayal by Michael Fassbinder in *Prometheus* is the opposite of Anthony Daniels' portrayal of C-3PO. While the latter's vocal display of emotion is often exaggerated to the point of histrionics, David displays an almost clinical reserve both physically and vocally, even when threatened with death. For example, in a confrontation with Meredith Vickers, mission supervisor (as well as Weyland's daughter and successor in the company), David initially refuses to answer her questions regarding his conversation with Weyland in hypersleep.[15] She throws David against the wall, coldly threatening "So help me God, I will find the cord that makes you run and I will cut it" (*Prometheus*). While David subsequently acquiesces to Vickers' demands, he offers no emotional response to either the physical attack or the threat, maintaining the same neutral voice and facial expression as when ostensibly displaying *joy*—such as in commenting "beautiful painting" upon seeing the elaborately decorated ceiling of the urn chamber in the alien ship.

However, it would be a mistake to label David's emotional responses as *inhuman* in anything more than a figurative sense: David's cold calculation at times resembles that of a human sociopath. It is clear that, like Ash, David's first priority is to his own agenda—or rather, that of Weyland Industries, here represented by no less than company founder Peter Weyland himself. Unlike Ash, however, David seems to chafe at being forced to submit to the will of another:

SHAW: What happens when Weyland is not around to program you anymore?

DAVID: I supposed I'd be free.

---

[14] "Weyland Industries: David 8," accessed 25 August 2014.

[15] It is revealed immediately prior to this scene that Weyland is not dead, as he indicates in his holographic address to the crew, but has accompanied the mission in a hypersleep chamber in order to meet face-to-face these "Engineers" they seek. David is in continual contact with Weyland via a helmet that allows him to communicate with an individual (the same one he uses to monitor Shaw's dreams); Vickers is not. From *Prometheus*.

SHAW: You want that?

DAVID: Want? Not a concept I'm familiar with. That being said, doesn't everyone want their parents dead? (*Prometheus*)

Shaw is understandably shocked by this comment, having lost both her parents at an early age; this exchange is even more disturbing once Shaw realizes that David is aware of this fact. However, disturbing as it may be, this ability to detach is in fact a feature advertised by Weyland Industries: their website claims that "unlike human counterparts, David 8 is willing to perform any task that he is assigned, without question or resistance. He will not flinch at even the most disturbing or seemingly irregular assignment, and he will dutifully persevere until reaching his final objective."[16] The implication here is that David, like Ash, having no moral compass of his own, is capable of performing morally questionable actions at which a normal human would balk. This possibility, not addressed by Breazeal or Dautenhahn in their discussions of sociable robots, is raised as a concern by many robotics detractors today: headlines calling for bans on "killer robots" that are armed and can select targets without human control are common in contemporary media.

Lawrence of Arabia is David's favorite film, and he often quotes lines he finds particularly apropos. In one of our first glimpses of David, we see him roaming the ship alone, watching the film and imitating Peter O'Toole's performance of T.E. Lawrence in a mirror. Fassbender explains the fascination David has for Lawrence:

> I always attributed it to the fact that Lawrence has got a very clear vision and he's very pure in his pursuit of it. There's not much questioning. He's a very decisive character, and I think David sees elements of that in [Noomi Rapace's character, Elizabeth] Shaw as well. That's why he finds her so fascinating. He's also an outsider like David; he's an

[16] "Weyland Industries: David 8," accessed 27 August 2014. David-8 elaborates on this point in a video: "I can carry out directives that my human counterparts might find distressing… or unethical" (Ibid.).

Figure 62. David watching Lawrence of Arabia *(1962) during the journey to LV_223. David models his performance of humanity after Peter O'Toole's performance of T.E. Lawrence in the film. Still from* Prometheus *(2012).*

> Englishman, but he's not accepted really by the English or the Arab nations, so he's kind of somewhere in the middle.[17]

David's fascination with Shaw can be seen from early in the film, as David is shown observing Shaw's dreams, during which she struggles to come to terms with her parents' deaths. David seems to relate to Shaw better than he does to anyone else in the film; this is part of what makes his inconsiderate treatment of her so disturbing. Ridley Scott offers further insight into the connection between David and Lawrence, making reference to the film clip David is seen watching at the beginning of *Prometheus*:

> So, he's been through the entire [ship's] library of maybe a million movies; the thing that stuck to mind was Lawrence of Arabia, which he plays again and again, mainly because he loves the anecdotal thinking of Lawrence himself. So we go on out, and the line is him saying "the point is not minding it hurts" when Lawrence burns his finger which, in a funny kind of way, has a kind of loose metaphorical connection to a person who has no soul, has no feelings, and, in fact, is David, so that's maybe why he

[17] Lily Rothman, "Prometheus star Michael Fassbender on his robot role and why he believes in aliens," *Time,* published 6 June 2012, accessed 28 August 2014, http://entertainment.time.com/2012/06/06/prometheus-star-michael-fassbender-on-his-robotic-role-ridley-scott-and-why-he-believes-in-aliens/.

> gravitated toward Lawrence because Lawrence in fact in a funny kind of way, while he's a very passionate character, also had a degree of a soulless quality to him, which made him, kind of, even more powerful.[18]

The "soullessness" that Scott attributes to Lawrence comes through clearly when David, studying himself in the mirror, perfectly imitates O'Toole's delivery of the line: "The trick, William Potter, is not minding that it hurts" (*Prometheus*). This appropriated mode of delivery becomes David's typical speaking pattern, establishing T. E. Lawrence as David's chosen "model" for proper human behavior. Whether this is an appropriate choice for a role model, however, is debatable, as Scott ascribes Lawrence as without emotion, "soulless."

STATUS

The status of synthetics within the *Alien*verse changes dramatically over the three and a half centuries between their initial patent in 2023 and the events aboard the USM Auriga in 2379. As seen in the timeline above, the National Science Foundation was initially against the commercial use of human-like androids, but extensive testing by Weyland Industries revealed that NSF concerns over integration were unfounded; once the David-X synthetics were offered commercially in 2042, they were well accepted in most fields—although it is interesting to note that the WI site indicates that no David-X models are deployed in the Security department of the company.[19] In fact, the press release for the David-8 model line (dated 23 December 2072) boasts of the model's high acceptance rate:

> After testing favorably in over 80% of households and workplaces, Weyland Cybernetics seems poised to make good on its "One in Every Home" promise with the upcoming David 8 Introduction Initiative this spring. In addition to 2-year trial placement in thousands of homes and businesses on Earth, David 8 has additionally wowed test

[18] Ridley Scott, "Director Commentary," *Prometheus*.
[19] "Weyland Industries: Investor Information," accessed 29 August 2014.

> audiences across the galaxy, performing off-world assistance missions on several newly terraformed planets.
>
> "He's even participated in titanium mining outside the industrial powerhouse, Outpost September, on KOI-252.01," explains William Concordov, Weyland Industries senior spokesperson. "As I understand it from the reports, Weyland miners and employees at the outpost called him 'just one of the guys'" says Concordov.[20]

While the "one in every home" ideal proffered by WI remains on the horizon, the press release seems to indicate that the David-8 is rapidly becoming the latest household necessity.

However, while the humans who interact with him on a daily basis may consider the David-8 model "just one of the guys," from a legal and marketing standpoint synthetics are unequivocally considered property. On the WI website, David-8's design, features, and specifications are outlined in the same manner as computers and other equipment are typically described on the sites of today's computer industry giants like Dell and Apple; deployment of the David-8 model is also discussed in terms of "pre-sales" and "number of units." Additionally, another WI press release addresses a malfunction and subsequent recall of around three thousand David-7 units, advising all owners "to carefully observe non-recall models and note any strange or unusual behavior. If any David 7 unit appears to become even mildly uncooperative, aggressive or unresponsive, they should deliver him to a certified Weyland Cybernetics repair station immediately."[21] According to the announcement, these units will be replaced, not repaired—and the three thousand "defective" units will presumably be destroyed.

There is a great deal of cognitive dissonance in the juxtaposition (both on the website and in the film) between the treatment of synthetics as equipment and their human appearance and behavior.

---

[20] "David 8 pre-orders exceed analyst's best expectations," *Weyland Industries: Weyland Pressroom*, accessed 29 August 2014, http://www.weylandindustries.com/press.

[21] "Weyland Cybernetics announces immediate recall of David 7 models," *Weyland Industries: Weyland Pressroom*, accessed 29 August 2014, http://www.weylandindustries.com/press.

Discussions of David-8's features and abilities on the WI website consistently use "he" and "him" when referring to David-8, and the promotional video on the site includes a voiceover in which a David-8 unit claims, "At Weyland Industries, it has long been our goal to create artificial intelligence almost indistinguishable from mankind itself."[22] The remainder of the video bears more resemblance to a job interview than a product commercial, and even includes the David-8 reacting to images in a Rorschach test. Aboard the Prometheus, David is treated like any other crewmember by most of the humans aboard—with notable exceptions.

How David's status on board the Prometheus is addressed depends more on the human in any given interaction, reflecting on the same ideas of race relations and prejudice as the treatment of droids in *Star Wars*. For example, there seems to be an undercurrent of tension between David and human Dr. Charlie Holloway, joint Lead Scientist (with Dr. Shaw) on the mission. While David's reactions to Holloway remain civil, Holloway himself seems to enjoy pointing out David's non-human status at any chance:

> DAVID: Am I interrupting? (*Offering a drink*) I thought you might be running low.
>
> HOLLOWAY: Pour yourself a glass, pal.
>
> DAVID: Thank you, but I'm afraid it would be wasted on me.
>
> HOLLOWAY: Oh. Right. I almost forgot. You're not a real boy, huh?
>
> DAVID: I'm very sorry that your Engineers are all gone, Dr. Holloway.
>
> HOLLOWAY: You think we wasted our time coming here, don't you?
>
> DAVID: Your question depends on your understanding of what you hoped to achieve by coming here.
>
> HOLLOWAY: What we hoped to achieve was to meet our makers. To get answers. Why they even made us in the first place.

[22] "Weyland Industries: David 8," accessed 25 August 2014.

DAVID: Why do you think your people made me?

HOLLOWAY: We made you 'cause we could.

DAVID: Can you imagine how disappointing it would be for you to hear the same thing from your creator?

HOLLOWAY: I guess it's a good thing you can't be disappointed, huh?

DAVID: Yes. It's wonderful, actually. (*Prometheus*)

Holloway here rather dismissively equates David with Pinocchio, the wooden puppet who longed to be "a real boy," and reduces the impetus behind David's creation to "because we could." Holloway's comments and demeanor would be construed as incredibly insulting in an interaction between two humans. Holloway's display here toward David indicates that he holds very little regard for the synthetic, meaning that either (1) he does not consider David sentient and may therefore safely treat him as an inferior, or (2) he does consider him sentient, and treats him as an inferior nonetheless. The former position hearkens back to Searle's substratism;[23] the latter position could be labeled as racism or speciesism.

David's interactions with Dr. Shaw, on the other hand, are much more positive. Shaw treats David with respect; in return, David shows concern for Shaw's welfare and makes repeated attempts to connect with her on some level, as when offering his condolences to Shaw after Holloway's death. However, his attempts highlight just how little David understands of human nature: immediately after his condolences, David questions Shaw about her contact with Holloway:

DAVID: I understand how inappropriate this is, given the circumstances, but as you ordered quarantine fail-safes, it's my responsibility to ask, have you and Dr. Holloway had any intimate contact recently? Since you and he were so close, I

[23] See note 16 in study one (page 261) regarding Searle and biological naturalism, which is ultimately a form of "substratism" in that it denies the possibility of sentience outside a biological substrate. (John R. Searle, "Minds, Brains, and Programs," in *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, John Haugeland, Rev. and enl. ed. [Cambridge, MA: MIT Press, 1997]).

just want to be as thorough as possible. *(Consults the monitor.)* My. My. You're pregnant.

SHAW: What?

DAVID: From the look of it, three months, so.

SHAW: No. That's impossible. I can't be pregnant.

DAVID: Did you have intercourse with Dr. Holloway?

SHAW: Yes, but ten hours ago. There's no bloody way I'm three months pregnant.

DAVID: Well, doctor, it's not actually a traditional fetus. (*Prometheus*)

David's failure to connect with Shaw continues throughout the film. As the scene progresses, David insists that Shaw be put in cryo-stasis until their return to Earth, but his attempts at comfort are misguided:

DAVID: Must feel like your god abandoned you.

SHAW: What?

DAVID: To lose Dr. Holloway, after your father died under such similar circumstances. What was it that killed him? Ebola?

SHAW: How do know that?

DAVID: I watched your dreams. (*Prometheus*)

Shaw is understandably shocked and outraged by this revelation, which comes as a surprise to David; he was honestly trying to forge a connection with her, but ultimately fails in his attempts. Despite his claims to "understand human emotion," as the promotional material promises, David's actions reveal he has much to learn about human nature.

As in the *Star Wars* universe, the status of synthetics in the *Alien*verse is typically accepted uncritically by the characters in the films. However, there is a distinct difference between Ridley Scott's approach to David and George Lucas' to C-3PO. While both Scott and Lucas claim that their mechanical

creations have no soul, Lucas' comment that C-3PO "is programmed to think in a particular way, and to be compassionate, but he doesn't really know what that means"[24] denies any sort of sentience as well. Scott makes no such claim of David. In fact, in Scott's director commentary, in all the promotional material, and within the film itself, the viewer is continually reminded of David's consciousness and cognitive abilities. What Scott denies is David's *humanity*—his possession of a *soul*, unreflexively imbuing the film with the belief that "human" is a unique state that may not be granted to any but *Homo sapiens*.

DISTINCTIVENESS

Whereas C-3PO's display of emotion and development of an autobiographical life set him apart (for the fans, at least) from the other droids of the *Star Wars* universe, the exceptionality of artificial humans within the *Alien*verse is much more problematic. As mentioned above, in 2073 there is approximately one David-7 model synthetic for every 1,250 humans, although the ratio appears to be much higher off-world than on Earth itself. At the same time, preorders for the David-8 model are "exceeding even the highest expectations."[25] Peter Weyland may have referred to David in his introduction as "the son I never had," but this doesn't necessarily indicate a fondness for this *particular* David: he is merely one of over eight million David-X models built since the first David-1 prototype in 2025, all "affectionately called David, a name Sir Peter Weyland had initially reserved for his own human son."[26] While the emotional capabilities of the David-X line have increased over the decades, the individual synthetics within each model should show no more variation than do the thousands of identical BMW 7-Series that roll off the line each year; in fact, the press release discussed earlier regarding the recall of David-7 models indicates that any deviation from factory standard should be treated as a malfunction and dealt with accordingly. Furthermore, unlike C-3PO, whose emotionality

[24] Lucas, George. In "The Characters of Star Wars." Disc 4 of the *Star Wars Trilogy* DVD set, released 2004.
[25] "David 8 pre-orders exceed analyst's best expectations," accessed 22 August 2014.
[26] "Weyland Industries: Corporate Timeline," accessed 22 August 2014.

sets him apart from other droids, the David-8 has been specifically programmed with appropriate emotional responses to a variety of situations. As noted above, David-8 is capable of eight distinct emotional states, yet he does not "feel," as David-8's responses in his promotional video show:

> NARRATOR: David, what makes you sad?
>
> DAVID (*slowly*): War. Poverty. Cruelty. Unnecessary violence. (*Tears begin to run down his cheeks*.) I understand human emotions, although I do not feel them myself. This allows me to be more efficient and capable, and makes it easier for my human counterparts to interact with me.[27]

The overall performance of the David-8 unit in the promotional material on the Weyland Industries website is identical to David's own performance aboard the Prometheus, reinforcing David's *lack* of uniqueness. Ash and Bishop, the synthetics aboard the Nostromo and Sulaco respectively, also show no indication of being unique examples of their model lines; on the contrary, both the treatment of the characters within the film narrative and the extra-filmic (yet still diegetic) "support" material seem to indicate that these synthetics are also no different than any other of their kind. While many other narratives revolve around an artificial human that has "become sentient"—*Bicentennial Man*, or *I, Robot* (2004) for example—the character of David (and arguably C-3PO as well) is unique *as a character* rather than *as an artificial human*; that is, his non-human status marks him as a unique character only within the film narrative. Within the larger universe, he is not particularly unique.

By contrast, Annalee Call in *Alien* Resurrection is a very different entity. One of the few autons left after the revolt and subsequent recall in the early twenty-fourth century, Call has begun to question her status as an artificial being in a way that David does not. As she uses a secret jack in her arm to

[27] "Weyland Industries: David 8," accessed 30 August 2014.

access the ship's mainframe in order to stop its return to Earth, Call reveals to Ripley[28] that she is repelled by her mechanical existence, comparing it to Ripley's own less-than-human nature:

> CALL: Why do you go on living?
>
> (Ripley looks at her but does not respond.)
>
> CALL: How can you stand being what you are?
>
> RIPLEY: Not much choice.
>
> CALL (*murmurs*): At least there's part of you that's human. (*Exhales*) I'm just— (*Sighs*) Look at me. I'm disgusting. (*Alien Resurrection*)

Nevertheless, Call is willing to acknowledge her status in order to help save the remainder of the humans on board. Call reminds Ripley that her original goal aboard the USM Auriga was to destroy Ripley in order to save humanity from itself, a quest that Ripley finds odd:

> RIPLEY: Why do you care what happens to them?
>
> CALL: Because I'm programmed to.
>
> RIPLEY: You're programmed to be an asshole? You're the new asshole model they're putting out? Come on.
>
> CALL: I couldn't watch them do it. I couldn't let them annihilate themselves. Do you understand that?
>
> RIPLEY: I did once. I tried to save... people. It didn't work out. There was this girl. She had bad dreams. I tried to help her. She died. Now I can't even remember her name. (*Alien Resurrection*)

While Call dismisses her concern for the humans' safety as part of her programming, the juxtaposition between her and Ripley's experiences reveals that, despite her mechanical nature, Call has a strong sense of human morality, and does *actually* care, in a way that the half-human Ripley cannot. This

[28] Here, the hybrid Xenomorph/human clone Ripley-8, created from the human Ripley's remains in order to breed Xenomorphs aboard the USM Auriga (*Alien Resurrection*).

morality is completely absent in Ash and David, and is found in Bishop only in his adherence to Asimov's Three Laws, but has caused Call to put herself in a position to explore further the "humanity" she is lacking—a behavior that is strongly evident in the character of Data in *Star Trek*, as we shall see in the next case study.

PERFORMANCE

Under normal conditions, synthetics within the *Alien*verse can offer entirely successful physical performances of humanity. Much of this performance, however, is dependent upon whether or not the human interactants are aware of the synthetic's nonhuman nature, as the consequences of a failed performance are much higher: while some missions overtly include a synthetic crewmember, other synthetics act as "sleeper agents," choosing to hide their nature until forced into a revelation through extenuating circumstances, typically a major injury. In *Alien*, Ash participates in all interactions as a human would, going to great lengths to appear completely human. Rather than roaming the ship during transit as David does, Ash "sleeps" in a cryo-stasis chamber alongside his crewmates during the long voyage. He also joins the others for meals, and even wears a surgical mask and gloves when examining the alien specimen, ostensibly as a quarantine measure. Ash's true nature is not revealed until more than two-thirds of the way through the movie, during the altercation with Ripley. Ash's reaction to Ripley's discovery makes it evident that neither Ash nor Weyland-Yutani wanted his synthetic nature known to the rest of the crew. Call similarly mimics "unnecessary" human behaviors until forced to reveal her nature, although her motivation is survival rather than subterfuge.

Even though David's companions aboard the Prometheus are aware of his mechanical nature, the tendency is to accept his performance as genuine, often forgetting (or choosing not to notice) that their crewmate is a synthetic. There are moments, however, when a human interactant deliberately draws attention to the synthetic's performance *as a* performance. Charlie Holloway's treatment of David

is one such example. As the crew suit up to explore the moon LV_223, Holloway calls out David's performative gesture of donning an environmental suit, an action the others must perform:

> HOLLOWAY: David. Why are you wearing a suit, man?
>
> DAVID: I beg your pardon?
>
> HOLLOWAY: You don't breathe, remember? So, why wear the suit?
>
> DAVID: I was designed like this because you people are more comfortable interacting with your own kind. If I didn't wear the suit, it would defeat the purpose.
>
> HOLLOWAY: Making you guys pretty close, huh?
>
> DAVID: Not too close I hope. (*Prometheus*)

Not only does David remind Holloway of the reason behind his performance, but he also implies that he finds both the necessity of the performance and Holloway's disruptive comment distasteful, displaying the "disgust" emotional state in response. David further rejects Holloway's assessment of the "success" of his performance by expressing a desire not to be "too close" in his similarity to a human. Similarly, Holloway's Pinocchio reference (that David is "not a real boy," as discussed earlier) implies that David *wants* to be human; however, David makes no such claim. In fact, David's response regarding his own inability to feel disappointment indicates that he prefers his synthetic state.

As the earliest model of synthetic seen in the *Alien*verse, David is also the least emotional. While he does display sympathy and regard for Shaw by the end of *Prometheus*, whether this display is genuine or simply offered in order to gain her assistance is difficult to determine. The Weyland Industries website describes the David-8 model's emotional responses in inconsistent terms, simultaneously claiming that he displays a "facsimile" of an emotion as well as "experiencing" the emotion. This dual capacity for simultaneously mimicking and experiencing emotional states reflects common modern approaches to acting. Take, for example, David-8's experience of joy: the "programmers" at Weyland Industries' cybernetics division claim that "In humans, the mere physical

Figure 63. Holloway taunting David regarding the android's wearing an environment suit for their exploration of LV_223's hostile environment. Still from Prometheus*.*

expression of joy (raising eyebrows, turning mouth corners up, etc.) can actually generate the emotion of joy itself. David 8 experiences joy in much the same way."[29] This "outside-in" approach may be compared to the James-Lange theory of emotion, which "maintains in essence that emotions have no real existence apart from their physiological manifestations."[30] According to Joseph Roach, this approach is common in many acting systems, "summed up by the acting teacher's familiar admonition: "Do the act, and the feeling will follow."[31] Other emotional states are described in a more "inside-out" manner: for example, David-8 is "programmed to experience grief and sadness" through the simulation of "a perceived memory,"[32] calling to mind the *affective memory* work of Stanislavski and Strasberg. In both cases, David is cast in the role of performer, and his actual experience of emotion is secondary to the display thereof.

Both David's and Ash's performances of humanity could be labelled "cynical" performances in Goffmanian terms, but for different reasons. Ash's performance of humanity is the classic con game: although aware of the discrepancy between his reality and his performance, he is absolutely invested in

[29] "Weyland Industries: David 8," accessed 30 August 2014.
[30] Joseph R. Roach, *The Player's Passion: Studies in the Science of Acting* (Ann Arbor: University of Michigan Press, 1993), 84.
[31] Ibid.
[32] "Weyland Industries: David 8," accessed 30 August 2014.

his success, to the point of violence when he is at risk of discovery. David, on the other hand, is a true cynic by Goffman's standard: an individual who "has no belief in his own act and no ultimate concern with the beliefs of his audience."[33] Further, in the case of his interactions with Holloway, David becomes a "cynical performer whose audience will not allow [him] to be sincere,"[34] as Holloway repeatedly calls attention to David's performance *as a performance*.

Most interesting is David's behavior when performing for himself; that is, when no possible "audience" for his performance is present. David watches over the Prometheus and its crew for over two years while the humans are in hypersleep for the journey; we watch his activities aboard the ship for a full five minutes before any other crew members appear. The devotion of this amount of time to David's solo activities aboard the ship reinforces his importance: it is no accident that we witness how he chooses to occupy his time. Besides watching *Lawrence of Arabia* and playing basketball while cycling, we also witness David practicing a "Comparative Reconstruction of Ancestral Language," or simply Ancestral, using the holographic equivalent of a Berlitz language video. He also uses a "neuro-visor" helmet to watch the dreams of Dr. Shaw as she lies in hypersleep. While all of these activities can be attributed to David's desire to learn during the journey, as director Ridley Scott indicates in his commentary, there are also a number of activities that do *not* fit this function. For example, as David is watching *Lawrence of Arabia*, we see him ostensibly bleaching his hair: he has white paste on the roots, and a towel around his shoulders. The film also reveals David eating an oatmeal-like substance and drinking a green "superfood" drink while watching the Ancestral training video. None of these behaviors serve any apparent function: David himself has no need for these things, and there is no other audience present. Who, then, is this performance *for*?

One possibility is that this performance is for the benefit of the extradiegetic audience, not for any particular audience within the film itself. When this scene occurs, it has not yet been revealed to the film

[33] Erving Goffman, *The Presentation of Self in Everyday Life* (Garden City, NY: Doubleday, 1959), 18.
[34] Ibid.

audience that David is not human—we discover David's status simultaneously with the crew, upon Weyland's announcement of the fact in the mission briefing. On the other hand, David's synthetic nature in *Prometheus* was not a big secret for those who cared to find out within the film community, as the promotional Weyland Industries site discussing David's production was available for months leading

Figure 64. David, while watching Lawrence of Arabia*, is depicted having a snack and bleaching his hair—neither of which are behaviors expected of a robot. Still from* Prometheus *(2012).*

Figure 65. David is also depicted eating while practicing Ancient in preparation for communicating with the Engineers the mission is hoping to find. Since there is nothing either in the film nor the extradiegetic material to suggest David needs *to eat (and multiple comments to the contrary), we can only assume that David* chooses *to do these things for his own personal reasons. Still from* Prometheus *(2012).*

up to the film's release. It could be argued that David performs these actions as further practice at "being human," or as an attempt to *become* more human himself. While David's responses to Holloway seem to indicate that David has no desire to be more humanlike, his interactions with Shaw indicate a desire to at least understand human nature; whether he desires that understanding to better interact with humans or to become more like them is unclear. By the end of Prometheus, David has realized the limitations of his own performance, admitting that he does not understand human motivation, as evidenced by Shaw's desire to confront the Engineers on their home planet:

> DAVID: May I ask what you hope to achieve by going there?
>
> SHAW: They created us. Then they tried to kill us. They changed their minds. I deserve to know why.
>
> DAVID: The answer is irrelevant. Does it matter why they changed their minds?
>
> SHAW: Yes. Yes it does.
>
> DAVID: I don't understand.
>
> SHAW: Well, I guess that's because I'm human being and you're a robot. I'm sorry.
>
> DAVID: That's quite alright. (*Prometheus*)

The upcoming sequel will reveal whether David is able to use this realization to overcome his limitations.

DAVID-8 AS *HOMO ARTIFICIALIS*

Based on the claims made by Weyland Industries, the David-8 model of cybernetic individual would seem to represent the pinnacle of human-robot interaction toward which roboticists such as Breazeal and Ishiguro strive: a sociable robot that "can enter seamlessly into any environment and carry out an authentic human interaction."[35] However, David's sociopathic behavior is far from Breazeal's sociable ideal. While certainly found within the human performative range, such behavior is considered aberrant rather than typical within society, and is often labeled as *inhuman* by those who encounter it.

[35] "Weyland Industries: David 8," accessed 18 August 2014.

Those individuals who display this personality type are commonly excluded from participation in society, through institutionalization if their sociopathy extends to violence. Whether this sociopathy is unique to David aboard the Prometheus or extends to all David-8 units is difficult to assess, but the "malfunction" and recall of the David-7 units seems to indicate that the tendency exists across the David-X product line. David's behavior calls into question the desirability of the human as the model for behavior, revealing possible unintended consequences of the "sociable robot" experiment without the restrictions that Asimov's Three Laws imply.

According to Brendan Vize, as we noted above, *persons* are conscious beings who have "interests and goals, the ability to use language, the ability to form reciprocal relationships, the ability to understand and apply moral concepts, rationality, self-awareness, [and] a personality."[36] Does David (and by extension the other synthetics within the *Alien*verse) fit Vize's definition of a person? He certainly has mastered language, not only English but Ancestral, interpolated from various Indo-European root sources, in order to attempt communication with the Engineer; his attempt is successful, albeit unwelcome (as his subsequent decapitation implies). While ostensibly operating as an extension of Peter Weyland's own ego, David nonetheless has his own interests and goals, revealed in the oblique comment "doesn't everyone want their parents dead?" (*Prometheus*). Furthermore, David is unquestionably rational and has an awareness of self. Sociopathic tendencies aside, David's cool demeanor and conciliatory voice fit quite well with the "type" Ridley Scott attributes to him: "I call him the butler. He's the housekeeper of the ship."[37] According to Fassbender, Scott "told me to take a look at *The Servant* with Dirk Bogarde, and I got a clearer idea of what sort of flavors we could play with in terms of somebody who starts off being subservient and then sort of grasps to take more control."[38]

---

[36] Brendan Vize, "Do Androids Dream of Electric Shocks? Utilitarian Machine Ethics" (Master's Thesis, Victoria University of Wellington, 2011), 56-57.
[37] Scott, "Director Commentary," *Prometheus*.
[38] Rothman, "Prometheus star Michael Fassbender," accessed 28 August 2014.

While the personality of a scheming underling is not necessarily a desirable one, it is a personality nonetheless.

The only real point at which David's performance of humanity fails is in his moral reasoning. Here, both the other characters in the film and director Ridley Scott himself take great pains to show that David is *not quite human* after all—that he is lacking that certain something that all humans share, a soul, as evidenced by his ultimate failure to understand humanity. However, this is a false dichotomy based in substratism. Soul or not, David perfectly performs a particular type of human—a sociopath—that, while undesirable, is nonetheless as human as any other *Homo sapiens*. While his particular choice of model may be unsavory and ultimately problematic, his performance is nonetheless successful.

STUDY THREE: DATA THE STARFLEET OFFICER

"I chose to believe that I was a person, that I had the potential to become more than a collection of circuits and sub-processors."

—Lieutenant Commander Data*, ST:TNG 6x23, "Rightful Heir"*

Our final case study considers Lieutenant Commander Data, a senior officer aboard the USS Enterprise in the television series *Star Trek: The Next Generation* (1987–1994) and the four films based around the same crew: *Star Trek Generations* (1994), *Star Trek: First Contact* (1996), *Star Trek: Insurrection* (1998)*, and Star Trek: Nemesis* (2002).[1] Appearing in all four films and all but one of the 176 series episodes, Data (portrayed by Brent Spiner) serves as a vital part of the Enterprise crew, acting variously as a science officer, operations officer, and second officer; he has also been given primary command of a ship on more than one occasion. As a principal character, Data has developed a rich autobiographical life over the fifteen years of the series and films, and is continually seeking to improve his performance of humanity.

Data was created by Dr. Noonian Soong sometime in the 2330s, the fifth of six androids Soong built, but the first that he considered a success.[2] The chronicle of his "autobiographical self," however, began on 2 February 2338, when Data was discovered by the crew of the USS Tripoli on the planet Omicron Theta and reactivated, with no memory of his previous existence (1x13, "Datalore"). Data eventually applied to Starfleet Academy, and was the only artificial being to ever be accepted, receiving

[1] Complete bibliographic information on the films and series may be found in the Filmography; throughout this section subsequent references will appear parenthetically within the text. I shall refer to the episodes in the TV series in the manner used on the website "Memory Alpha: the Star Trek Wiki" (http://en.memory-alpha.org/wiki/Portal:Main), as their notation is very clear, noting season and episode number based on original airdate. For example, the pilot episode, "Encounter at Farpoint," bears the notation 1x01/02, indicating that it is the first season, first and second episodes—while the double episode originally aired as a single unit, it is typically broadcast as two separate episodes in reruns. I shall refer to the four films by subtitle alone (*Generations*, for example).

[2] The six androids Soong created are, in chronological order: two unnamed (and unsuccessful) prototypes, B-4 (who Data discovers in 2379), Lore (a reoccurring character who may best be described as Data's "evil twin"), Data, and finally Juliana Soong Tainer (who was built to replace Soong's wife due to a fatal injury; Data meets Tainer in 2370).

only one dissenting vote—that of Starfleet cybernetics expert Bruce Maddox, on the grounds that he did not qualify as a sentient being (2x09, "The Measure of a Man"). After graduating from the Academy with honors in 2348 (1x01/2, "Encounter at Farpoint"), Data spent three years as an ensign and twelve as a lieutenant before being promoted to lieutenant commander in 2360 (1x13, "Datalore") and joining the crew of the USS Enterprise-D in 2364 as its second officer under Captian Jean-Luc Picard (1x01/02, "Encounter at Farpoint"). During his time on board the Enterprise, Data forged positive relationships with many members of the bridge crew, and was so well-liked by his colleagues that, upon his apparent destruction in Season Three, Commander Riker remarked to Picard that, "for an android with no feelings, he sure managed to evoke them in others" (3x22, "The Most Toys"). Data served aboard the USS Enterprise-D and its successor the USS Enterprise-E, playing a major role in its daily operations, until his death in 2379, ultimately sacrificing himself to save the Enterprise and its crew (*Nemesis*). The film's ending left open the possibility of "bringing Data back to life" through the android B-4, an earlier Soong-type android who has been loaded with the contents of Data's memory banks; however, this seems unlikely due not only to B-4's more limited capabilities (he is an earlier Soong prototype without Data's complex positronic brain), but also to the fact that actor Brent Spiner considers himself "visibly aged out of the role" after fifteen years.[3]

As a character, Data exists in *ST:TNG* in many ways as a foil for the largely human ensemble, designed to explore humanity from an outsider's perspective, serving the mission not only of the *ST:TNG* series, but of the Star Trek franchise in general: according to the "What Has Not Changed" section of the "*ST:TNG* Writer/Director's Guide," stories for the series should focus on "premises involving the challenges facing humanity today (the 1980s and 90s),… assessing where we humans presently are,

[3] I have found dozens of references to this comment, purportedly made by Spiner around 2006, when asked about a possible reprisal of the role in J.J. Abrams *Star Trek XI*; the furthest back I can track the quote is to Crave Online Media, *ComingSoon.net*, "Brent Spiner Rules Out *Star Trek XI*," accessed 2 September 2014, http://www.comingsoon.net/films.php?id=16437.

*Figure 66. Brent Spiner as Data in 1987 (*ST:TNG *1x01, "Encounter at Farpoint") and 2002 (*Star Trek: Nemesis*). After fifteen years, Spiner considers himself "visibly aged out of the role," despite there being a diegetic explanation for the android's change in appearance.*

where we're going, and what our existence is really about."[4] In this respect, Data fulfills the same function as Spock in the original series. The two share a number of traits: capacity and desire for logic, an innate curiosity, and an occasional failure to understand the emotional nature of human beings. But while Spock, although not human, is nonetheless a *biological* being, Data is a *machine*, which opens his character to an entirely different set of issues to explore.

And explore he does. Some element of Data's "path to personhood" is addressed in nearly every episode of the series and in all four feature films: he is shown practicing sneezes (1x13, "Datalore"), navigating the intricacies of dating (4x25, "In Theory"), learning small talk (6x18, "Starship Mine") and slang (1x05, "The Last Outpost" is only one of many examples), and trying to come to an understanding of human emotion (such as in 6x26/7x01, "Descent") and death (1x23, "Skin of Evil"). He is repeatedly faced with discrimination (or worse) due to his mechanical nature. Each episode, each new encounter, becomes a learning opportunity for Data, and he draws the audience with him in his struggles. In the

[4] Gene Roddenberry, "*Star Trek: The Next Generation* Writer/Director's Guide," (1987), 4.

"*ST:TNG* Writer/Director's Guide," Roddenberry encourages a strong audience identification with the character of Data:

> What is most appealing about Data is not that he is a machine, but that he is a machine who is eager and enthusiastic about the adventure of life. To that end, he is an inquiry into the question, "What does it mean to be a human being?" Data's favorite story is *Pinocchio*—he identifies himself with the little wooden boy who wanted to be human.[5]

Figure 67. Data is seen performing various behaviors typically seen in a human friend and colleague. Top left*: Data practices sneezing, a particularly human action (*ST:TNG *1x13, "Datalore").* Top right*: Data has developed a "small talk" subroutine, and is seen putting it into practice (*ST:TNG *6x18, "Starship Mine").* Bottom left*: Data often plays poker with his colleagues (*ST:TNG *2x09, "Measure of a Man").* Bottom right: *Beverly Crusher teaches Data ballroom dancing for the wedding of his friends Keiki Ishikawa and Miles O'Brien (*ST:TNG *4x11, "Data's Day").*

[5] Ibid., 27.

Data's struggle to "be human" becomes our own struggle as audience to "be better humans," as we identify with his awkwardness, confusion, and failures. The character's popularity bears this out. In 2009 Data was listed as number four in the "Top 25 Star Trek Characters" on IGN: "Being front and center for many debates surrounding the notion of what it means to be human, Data… evolved to be more than the sum of his programming, and into one of Trek's legendary mainstays."[6] As cultural theorist Klemens Hippel claims, Data is "the most human of us all."[7]

PHYSICALITY

Like C-3PO and David, Data is entirely mechanical. While his appearance is much closer to human than C-3PO, Data's pale skin and yellow eyes set him apart from the rest of the crew, marking him as non-human. However, given the broad range of alien species within the *Star Trek* universe, it is not always apparent to others that Data is an android rather than an alien. This liminal nature is highlighted from the very first episode of the series: while escorting the retired Dr. Leonard McCoy (chief medical officer of the Enterprise in the original *Star Trek* series) around the ship, Data comments on McCoy's age. When McCoy asks how Data can remember his age so precisely, Data responds:

DATA: I remember every fact I am exposed to, sir.

MCCOY: I don't see any points on your ears, boy, but you sound like a Vulcan.

DATA: No, sir. I'm an android.

MCCOY: Almost as bad.

DATA: I thought it was generally accepted, sir, that Vulcans are an advanced and most honorable race.

MCCOY: They are, they are. And damned annoying at times. (1x01, "Encounter at Farpoint")

---

[6] In anticipation of the release of *Star Trek XI*. Benjamin Maxwell, "Top 25 Star Trek Characters," *IGN Movies*, accessed 2 September 2014, http://www.ign.com/articles/2009/05/08/top-25-star-trek-characters..

[7] Klemens Hippel, "Der Menschlichste Von Uns Allen. (German)," *Montage AV* 8 (1999).

Here, McCoy has tried to reconcile Data's behavior (logical) with his appearance (non-Vulcan); when the truth is revealed, however, McCoy takes Data's non-biological sentience in stride. Data is often mistaken for human (or at least as *humanoid alien*,[8] for which he easily passes) by those he meets. For example, in the episode "Pen Pals" (2x15), Data has been communicating with Sarjenka, a child from a planet that has not yet made contact with outsiders. Circumstances force Data to bring Sarjenka aboard the ship, and Dr. Pulaski gives the girl an Elanin singer stone, which Pulaski explains "sings a different song for each person." Sarjenka then offers the stone to her "pen pal":

SARJENKA: What does it sing for you?

DATA: It does not sing for me.

SARJENKA: Why not?

DATA: Because I am a machine. (2x15, "Pen Pals")

Data is nearly always up-front about his mechanical nature, and does not seek to hide it. In certain situations, however, his ability to "pass" comes in handy. When a technical accident on an away mission lands him in 1893 San Francisco (5x26/6x01, "Time's Arrow"), Data is forced to stay until he is able to build a device to re-open the hole in time. While in San Francisco, Data's pallor and odd behavior is attributed to his being a "Frenchman." While he is typically quick to correct his misidentification as human by others, in this instance he preserves the ruse, as his safety is at stake.[9]

Other episodes reveal more precisely Data's physical makeup. Composed of 24.6 kilograms of tripolymer composites, 11.8 kilograms of molybdenum-cobalt alloys and 1.3 kilograms of bioplast sheeting, (3x22, "The Most Toys"), Data weighs approximately one hundred kilograms, slightly more than the average human male (7x10, "Inheritance"). His skeletal structure is designed to withstand

---

[8] Defined in the "*ST:TNG* Writer/Director's Guide" as "a life form which is similar to humans" such as the commonly seen Vulcan, Romulan, and Klingon races as well as numerous other humanoid alien races encountered (Roddenberry, "*ST:TNG* Writer/Director's Guide," 34.).

[9] As is the Prime Directive, which prohibits interference with "less advanced" civilizations by providing knowledge or technology that would alter their natural development—as the discovery of an android in 1893 San Francisco would most assuredly do.

Figure 68. Data in 1893 San Francisco, attempting to pass as human to blend in until he can return to his own time. After one (rather drunk) vagrant asks if he's a "Frenchman" due to his "pajamas" (i.e., his Starfleet uniform), Data adopts the persona to explain his odd eyes and pale complexion. While those he meets are suspicious, they seem to accept his explanation. Still from ST:TNG *4x26, "Time's Arrow, Part 1."*

extreme stress, and he routinely displays super-human strength (2x09, "The Measure of a Man") and speed (1x03, "The Naked Now"). Data possesses a positronic brain, a direct reference to Asimov's robots of the late twentieth century, and possesses an ultimate storage capacity of eight hundred quadrillion bits[10] and a total linear computational speed rated at sixty trillion operations per second[11] (2x09, "The

[10] Approximately 1,000 terabytes, or one petabyte. This would have been considered an enormous amount of storage when the episode first aired in 1988, when a gigabyte-capacity drive was the size of a refrigerator.

[11] Computer performance is currently measured in floating-point operations per second (flop/s), rendering Data's performance at 60 Tflop/s. By contrast, the Cray Y-MP supercomputer introduced in 1988 operated at approximately 333 Mflop/s per processor, for a record sustained speed of 2.33 Gflop/s (Cray: The Supercomputer Company, "Cray History," accessed 2 September 2014, http://www.cray.com/About/History.aspx).

Measure of a Man"); Dr. Soong used this storage capacity to preserve the accumulated knowledge of his fellow colonists on Omicron Theta when their colony was destroyed (1x13, "Datalore"), although he wiped Data's own personal memories beforehand.

Envisioned by Roddenberry from the series' concept phase as a "fully functional" human male,[12] Data is equipped with human physiological features that would initially seem unnecessary for an android, such as tear ducts (he cries in *Generations,* for example). In the sixth-season episode "Birthright," visiting doctor Julien Bashir (chief medical officer of Federation space station Deep Space 9) expresses fascination at the extraordinary lengths to which Dr. Soong went in creating a naturalistic human appearance in Data—focusing on Data's similarities to a human rather than his differences. For example, Data breathes, has a pulse, and is capable of hair growth (6x16, "Birthright, Part I"), and at one point he even grows a "fine, full, dignified beard" (2x06, "The Schizoid Man"). The blinking of Data's eyes is governed by a Fourier series to simulate randomness. (7x10, "Inheritance"), and Data has even been fitted with an "aging program" to account for actor Brent Spiner's changing appearance (7x10, "Inheritance").

The extent of Data's humanistic physiological detail is revealed in "The Naked Now" when, suffering from a "polywater intoxication" that strips away inhibitions, Security Chief Tasha Yar persuades Data to have sex with her:

TASHA: And what I want now is gentleness. And joy. And love. From you, Data. You are fully functional, aren't you?

DATA: Of course, but—

TASHA: How fully?

DATA: In every way, of course. I am programmed in multiple techniques, a broad variety of pleasuring—

[12] Roddenberry, "*ST:TNG* Writer/Director's Guide," 27.

TASHA: Oh, you jewel! That's exactly what I hoped. (1x03, “The Naked Now”)[13]

As the scene ends, Yar leads Data into her room and sits him on the bed. Data’s later behavior not only confirms that the two have been intimate, but reveals further insights into his physiology:

PICARD: Captain to Wesley. Wesley Crusher, come in!

(Data enters, smiling and weaving slightly)

PICARD: Ah, Data. At least you're functioning.

DATA: Fully, Captain.

PICARD: Data, intoxication is a human condition. Your mind is different, it's not the same as—

DATA: We are more alike than unlike, my dear Captain. I have pores. Humans have pores. I have fingerprints. Humans have fingerprints. My chemical nutrients are like your blood. If you prick me, do I not leak? (1x03, “The Naked Now”)[14]

Figure 69. Data proves susceptible to the “polywater intoxication” that has dangerously released the crew’s inhibitions. After a steamy encounter with crewmate Tasha Yar, Data waxes poetic to Captain Picard on the similar natures of man and android. Stills from ST:TNG *1x03, “The Naked Now.”*

[13] While Data does try to forge romantic relationships later, his intimate encounter with Yar is the only such encounter mentioned in the series.

[14] Data here paraphrases the famous words of Shakespeare’s Merchant of Venice. Data received a hardbound antique edition of the complete works of Shakespeare from Captain Picard as a gift; he is often seen reading this in his quarters, and the volume is mentioned as well in the tribunal held in Season Two to determine Data’s status as property or person.

Not only is Data very similar to humans in appearance, but he is as susceptible to the polywater intoxication as his biological crewmates.

Despite his near-human appearance, both his colleagues and the audience are continually reminded of Data's mechanical nature. Data's breathing is part of his thermal regulation system, but he is capable of functioning in a vacuum for extended periods (4x03, "Brothers"). Likewise, his circulatory system distributes biochemical lubricants throughout his body (6x16, "Birthright, Part I"). Panels on Data's scalp fold back to reveal the circuitry beneath his skin, as well as various data ports, which are used for diagnostic tests and other functions (as shown in Season Six in Figure 70 below, but seen on many occasions). Data further reveals to Dr. Beverly Crusher that he has an "off-switch" located in the small of his back (1x13, "Datalore"); this knowledge is used to startling effect in "The Measure of a Man" when, after removing Data's hand to demonstrate his mechanical nature, Commander William Riker reaches behind Data and summarily switches him off, shocking everyone present. Finally, while Data himself is never disassembled, an away team discovers the separate parts of Data's "twin brother" Lore in Soong's laboratory on Omicron Theta, and subsequently use Data as the model for Lore's reassembly (1x13, "Datalore"). Finally, Data's head is "lost" in 1893 San Francisco and only recovered five centuries later (6x01, "Time's Arrow, Part II"); while Data's head does not function separately from his body as David's does in Prometheus, he works perfectly after it is reattached in 2369.

BEHAVIOR

The treatment of Data's physicality throughout the narrative timeline of *Star Trek* is a continual juxtaposition between those features that emphasize his humanity and those that highlight his mechanical nature; the dual nature of Data's behavior (both physical and social/emotional) is approached in a similar way. Data's physical performance is much like David's in *Prometheus*: possessed of an economy of movement, erect posture, and passive facial expression, Data tends to a "neutral"

Figure 70. Geordi connects Data directly to the ship's computer so that he might act as an emergency backup in case of a system failure. Still from ST:TNG *6x08, "Fistful of Datas."*

formal comportment, with one characteristic "tic": when curious, or when he misunderstands or is confused by something, Data tilts his head slightly to one side in a gesture reminiscent of the dog from the RCA Victor trademark. This is not to say, however, that Data's physical performance lacks definition; on the contrary, his performance is so recognizable among *Star Trek* fans that at conventions, actor Brent Spiner needs do no more than tilt his head "Data-style" to garner applause from the audience during panel discussions.

Data's behavioral idiosyncrasies are distinctive to his colleagues aboard the Enterprise as well. At several points in the series Data is impersonated by others, and it is the difference in his performance that gives the interloper away. The first instance of impersonation occurs when Data's brother Lore surreptitiously deactivates Data and assumes his place, doing his best to imitate Data's physical and

Figure 71. Data is impersonated on several occasions, but aberrant behaviors reveal the impersonation. Left*: After deactivating Data and switching clothing with him, Lore nearly gives away his identity to Wesley Crusher through his use of contractions and an uncontrollable facial tic (*ST:TNG *1x13, "Datalore").* Right*: a dying Ira Graves uploads his identity into Data's body in order to survive. His behavior around his former assistant Kareen Brianon, to whom Graves (in the guise of Data) professes his love, reveals the switch, and Graves agrees to return Data's body (*ST:TNG *2x06, "The Schizoid Man").*

social performance (1x13, "Datalore"). In this regard, Lore fails, partially due to a facial tic in Lore that Data lacks, and Data's more observant friends uncover the deception. In Season Two, Data's body is commandeered by molecular cyberneticist Ira Graves in an attempt to cheat death, his own health failing due to the fatal Darnay's disease (2x06, "The Schizoid Man"). While Graves is able to maintain Data's performance for a time, his behavior grows increasingly erratic, leading Data's friends puzzled before they realize the truth. Geordi LaForge, Data's closest friend, suggests that his behavior results from the fact that "sometimes he wants to be human so badly he can taste it. I think he just tends to get confused when his human qualities surface" (2x06, "The Schizoid Man"), and that Graves' death (whom Data considered a sort of grandfather-figure due to his connection to Soong) acted as a catalyst for the change. Ultimately, however, a personality test similar to those administered upon admission to Starfleet reveals that a second, more dominant personality (Graves) has taken over Data's mind, and Graves is ultimately persuaded to relinquish control and be stored instead in the ship's computer (2x06, "The Schizoid Man").

Data's rather formal physicality is mirrored in his social and emotional performance. Unlike both David and C-3PO, Data has not quite mastered human language (i.e., English); he has particular trouble with contractions, slang, and idiomatic language. This difficulty is addressed from the first episode, as Captian Picard describes the Enterprise's mission at Farpoint Station:

PICARD: It's hardly simple, Data, to negotiate a friendly agreement for Starfleet to use the base while at the same time snoop around finding how and why the life form there built it.

DATA: Inquiry. The word snoop?

PICARD: Data, how can you be programmed as a virtual encyclopedia of human information without knowing a simple word like snoop?

DATA: Possibility, a kind of human behavior I was not designed to emulate.

PICARD: It means to spy, to sneak.

DATA: Ah! To seek covertly, to go stealthily, to slink, slither—

PICARD: Exactly, yes.

DATA: Glide, creep, skulk, pussyfoot, gumshoe. (1x01/02, "Encounter at Farpoint")

This exchange between Data and Picard happens in the opening scene of the series, not only revealing Data's status as an artificial being but at the same time establishing just what kind of artificial being Data would prove to be: friendly and inquisitive, trying to understand the human behavior of those around him but, occasionally, failing to do so. In Picard's exasperated attempt to cut off Data's list of synonyms, we also see that Data often fails to realize the effect his matter-of-fact delivery of information may have on those around him. In "The Naked Now" (1x03), Data fails to understand Dr. Crusher's offhand quip, resulting in another awkward moment, which Data realizes only later:

CRUSHER (*smiling*): If you were any more perfect, Data, I'd write you up in a Starfleet medical textbook.

DATA: I am already listed in several bio-mechanical texts, Doctor.

CRUSHER (*frowning*): Yes of course. (1x03, "The Naked Now")

Crusher's curt reply puzzles Data, as he reveals to Riker a short time later:

RIKER: This ought to be easy for someone written up in bio-mechanical texts.

DATA: About that, sir. Did the Doctor believe I was boasting?

RIKER: Probably. This may take some time.

DATA: At least several hours. But what I said was a statement of fact. Perhaps she will look it up.

RIKER: You can depend on it. (1x03, "The Naked Now")

With his characteristic head-tilt, Data appears to file this interaction away for later analysis, hoping that by understanding Crusher's reaction he will gain further insight into the human psyche.

Data's continuing quest to understand human nature in order to better emulate human behavior is one of the major themes of the series, reflecting Roddenberry's concept for the *Star Trek* franchise and the character himself. Defining human nature—What is emotion? What is humor?—proves as intractable for Data as defining sentience, or a soul, is for the metaphysicists. His friends, while providing ample opportunity for observational study, are at a loss when pressed to explain their behavior, as Data discovers upon trying to come to terms with experiencing his first "emotion":

LAFORGE: No offence, Data, but how would you know a flash of anger from some odd kind of power surge?

DATA: You are correct in that I have no frame of reference to confirm my hypothesis. In fact, I am unable to provide a verbal description of the experience. Perhaps you could describe how it feels to be angry. I could then use that as a reference.

LAFORGE: Well, okay. When I feel angry, first I feel hostile.

DATA: Could you describe feeling hostile?

Figure 72. Data, overcome by anger, chokes a Borg soldier to death with his bare hands. Later he calls on Geordi to explain the feeling of anger to him. Still from ST:TNG *2x26*, *"Descent, Part I."*

LAFORGE: It's like feeling belligerent, combative.

DATA: Could you describe feeling angry without referring to other feelings?

LAFORGE: No, I guess I can't. I just feel angry.

DATA: That was my experience as well. I simply felt angry. (6x26, "Descent, Part I")

Data later learns that his emotion was not spontaneous, but rather generated by Lore in an attempt to twist Data's loyalty and destroy the Federation.

While Data's "older brother" Lore was constructed with the ability to feel and express emotion, the negativity of the emotions he displayed caused Soong to deactivate Lore, before constructing Data without such ability. While Soong later offers Data an "emotion chip" which would allow him to experience emotion, this chip is stolen by Lore before Data has the chance to use it. Data ultimately recovers the emotion chip from Lore, but it is not until the film *Generations* that Data chooses to finally make use of it, to great comic effect: he is initially unable to cope with the overwhelming experience, as

he is unable to regulate his responses, and rather than his normally stoic personality we see a Data reduced to histrionics. For example, scanning for lifeforms on a planet causes him great pleasure (Data is so overjoyed that he spontaneously breaks into song); by contrast, he is temporarily immobilized with terror when put in a situation where he experiences fear for the first time (*Generations*). After an initial adjustment period, Data gains the ability to turn the chip on and off, and later to remove it completely in situations where emotions would be a hindrance (*Generations*, *First Contact*, *Insurrection*).

Figure 73. After installing his emotion chip at the beginning of Star Trek Generations*, Data finds he is unable to regulate emotional response, and experiences extreme emotional swings. Humor becomes hysteria (as seen in the top image), fear becomes paralyzing terror, and relief over finding his cat Spot reduces Data to tears (as seen in the lower image). It takes Data time to learn to properly regulate his emotional response. Both stills from* Star Trek Generations *(1994).*

But is Data really without emotion? It is true that Data never gets angry or takes offense, even when a "typical" human would be expected to do so in the situation. On the other hand, Data displays a number of personality traits that are difficult to explain for someone who seemingly cannot feel emotion. According to Klemens Hippel:

> He is also curious, creative, interested in art, a gifted violinist, a fan of Conan Doyle's novels, has a holograph of a deceased friend as a keepsake, attempts to procreate, etc., all aspects of character that would be difficult to understand without recourse to emotion; in particular he displays something like friendship, and is not a stranger to mourning.[15]

Data's display of friendship in particular reveals his emotional nature. Unlike David, whose relationships with the other crew on board the Prometheus seem to be based on expediency and personal benefit, Data forms friendships based on mutual respect and admiration, and interacts with those people in a way that implies a concern for their well-being.

The development of one such relationship illustrates in a single episode the juxtaposition between Data's professed lack of emotion and his display of consideration, and deserves a close analysis. In season four, when crewmate Jenna D'Sora suffers a painful breakup, Data does his best to console her, not only offering support for a friend but taking the opportunity as well to study another aspect of human behavior:

> DATA: Six weeks ago, when you and Jeff dissolved your relationship, I saw an excellent opportunity to study that aspect of human intimacy. As your friend, it is my responsibility to be supportive in times of need.
>
> JENNA: That's very sweet, Data. (4x25, "In Theory")

[15] Hippel, "Der Menschlichste Von Uns Allen," 83, translation mine. Original passage: „Er ist auch neugierig, kreativ, kunstinteressiert, begnadeter Violinvirtuose, Fan der Romane Conan Doyles, verfügt über eine Holographie einer Verstorbenen als Andenken, versucht, sich fortzupflanzen usw, Aspekte der Figur, die ohne Rekurs auf Emotionalität schwer zu verstehen wären; insbesondere ein Äquivalent für Freundschaft ist ihm zugänglich, und auch Trauer scheint ihm nicht fremd."

Over time, Jenna develops something of a crush on Data, and attempts to turn him into a romantic partner, much to Data's bemusement:

> JENNA: That what I love about you, Data. You make me laugh. I don't know why I keep falling for the wrong man. Why can't I fall for somebody like you? You're perfect.
>
> DATA: That is not true. I have no human feelings.
>
> JENNA: But you give me so much. You spend time with me when I was lonely; you encouraged me when I'm down. No man has ever been kinder to me. Those are the things that matter. (4x25, "In Theory")

In order to successfully navigate this new exploration of human behavior, Data seeks the advice of those he respects. He first turns to Guinan, the bartender of the lounge Ten Forward, whose five centuries of life experience make her a good source of wise counsel. Data admits to Guinan that he is struggling to understand Jenna's attention:

> DATA: Lieutenant D'Sora just gave me what could be considered a very passionate kiss in the torpedo bay.
>
> GUINAN: Really? And?
>
> DATA: I was intrigued. Jenna seemed to be displaying genuine affection for me.
>
> GUINAN: Well, what do you think of her, Data?
>
> DATA: I find her to be a competent officer. Highly motivated, though somewhat lacking in the understanding of the theory underlying the dilithium matrix application.
>
> GUINAN: I meant personally?
>
> DATA: I look forward to the time we spend together.
>
> GUINAN: Well, then it seems the next move is yours. What are you going to do?
>
> DATA: I do not know. I have no experience in such matters. I require advice.

GUINAN: Don't look at me. *(Data looks away.)* No, no, Data, I simply mean that I can't give you any advice here. It's not good to advise people about their first love affairs. That's kind of something they have to figure out for themselves.

DATA: But I am not capable of love.

GUINAN: Then it's going to be a very unique experience. (4x25, “In Theory”)

Data further seeks advice from Geordi LaForge, Counselor Troi, Commander Riker, and even Captain Picard in his attempt to understand and respond appropriately to Jenna’s overtures; in each exchange he is offered different (but overwhelmingly encouraging) advice while simultaneously protesting his lack of emotion.

Data ultimately chooses to pursue the relationship with Jenna, and on the advice of Commander Riker, brings Jenna flowers to consummate the relationship. Data approaches the encounter awkwardly, having supplemented his lack of experience in these matters with extensive research and self-programming:

JENNA: You didn't talk to the entire ship about us.

DATA: No. In actuality, less than one percent of the Enterprise crew was involved. It was necessary to balance theory with experiential referents. Both are required for a program of this nature. Computer, decrease illumination level by one third standard lux. (*He clears the sofa then puts out an arm. She is meant to snuggle into it*.)

JENNA: This is all part of a program?

DATA: Yes. One which I have just created for romantic relationships.

JENNA (*slightly offended*): So I'm just a small variable in one of your new computational environments?

Figure 74. Data has developed a "romantic relationship program" in order to be the perfect romantic partner for Jenna, who has suffered from a string of bad relationships. While he has drawn from a number of literary and real-life sources for proper behavior, Data has difficulty in determining which behavior is best suited to particular situations. In the end, this behavior drives Jenna away, as she claims it does not fit Data's own personality. Still from ST:TNG *4x25, "In Theory."*

DATA (*attempting to reassure her*): You are much more than that, Jenna. I have written a subroutine specifically for you. A program within the program. I have devoted a considerable share of my internal resources to its development.

JENNA: Data, that's the nicest thing anybody's ever said to me. (4x25, "In Theory")

As Data's "romantic relationship program" plays out, his behavior becomes increasingly disturbing to Jenna: first he caters to her every need, then instigates a "lover's quarrel." Jenna reacts negatively:

JENNA: But there's something so forced and artificial about the way you're doing it, Data. It's just not the real you.

DATA: With regard to romantic relationships, there is no real me. I am drawing upon various cultural and literary sources to help define my role.

JENNA: Kiss me. (*They kiss*.) What were you just thinking?

DATA: In that particular moment, I was reconfiguring the warp field parameters, analyzing the collected works of Charles Dickens, calculating the maximum pressure I could safely apply to your lips, and considering a new food supplement for Spot.

JENNA: I'm glad I was in there somewhere. (4x25, “In Theory”)

Eventually Jenna grows frustrated with Data’s failure to understand the intricacies of dealing with a romantic partner, and decides to end their relationship. She attributes the split to Data’s lack of emotion:

JENNA: I didn't see it until today. I got out of a relationship with an unemotional man, and I got right back into another, with a man who is absolutely incapable of emotion.

DATA: There does appear to be a recurring motif.

JENNA: You were so kind and attentive. I thought that would be enough.

DATA: It is not?

JENNA: No, it's not. Because as close as we are, I don't really matter to you. Not really. Nothing I can say or do will ever make you happy or sad, or touch you in any way.

DATA: That is a valid projection. It is apparent that my reach has exceeded my grasp in this particular area. I am perhaps not nearly so human as I aspire to become. (4x25, “In Theory”)

Data offers to delete the romance program he has created, and as Jenna leaves in tears, Data picks up his cat Spot, apparently unaffected by the breakup.

The way in which this relationship plays out is most intriguing. Jenna’s claim that “as close as we are, I don’t really matter to you” is demonstrably false. Over the course of the episode, Data has gone to great lengths to take Jenna’s feelings into consideration: he seeks advice from others more versed in human relationships, he finds examples of (what he perceives as) successful historical and literary romantic behavior, and he adjusts his performance continually based on what he thinks will please her. Data’s neutral demeanor, which Jenna admits drew her to him in the first place, is precisely what pushes her away.

While Data becomes neither angry nor indignant at Jenna’s ultimate withdrawal, the implication that he does not care about the loss on some level makes no sense when compared with the outcome of his relationship with Tasha Yar in the first season. While Data’s encounter with Yar under the influence of polywater intoxication was never repeated, due to Yar’s own distress at her unprofessional behavior, it is obvious that the relationship had an effect on Data. When Yar is killed unexpectedly in “The Skin of Evil” (1x23), Data approaches Picard after the funeral, troubled:

DATA: Sir, the purpose of this gathering confuses me.

PICARD: Oh? How so?

DATA: My thoughts are not for Tasha, but for myself. I keep thinking how empty it will feel without her presence. Did I miss the point?

PICARD: No, you didn't, Data. You got it. (1x23, “The Skin of Evil”)

While Data sheds no tears for Yar, the description of his thoughts after her death reveal his very human sense of loss. Later, in "The Measure of a Man" (2x09), one of Data's few possessions is a hologram of Yar, kept in memory of her after her death. When pressed as to why he has a hologram of her and no other crewmembers, Data responds "She was special to me, sir. We were intimate" (2x09, "The Measure of a Man"). Data's actions both at the funeral and in explaining the hologram belie his protestations that he "is not capable of emotion."

STATUS

As is evident both in the *Star Wars* and *Alien* universes, there is an undercurrent of prejudice against artificial lifeforms. In Star Trek, however, this prejudice is depicted as something external that

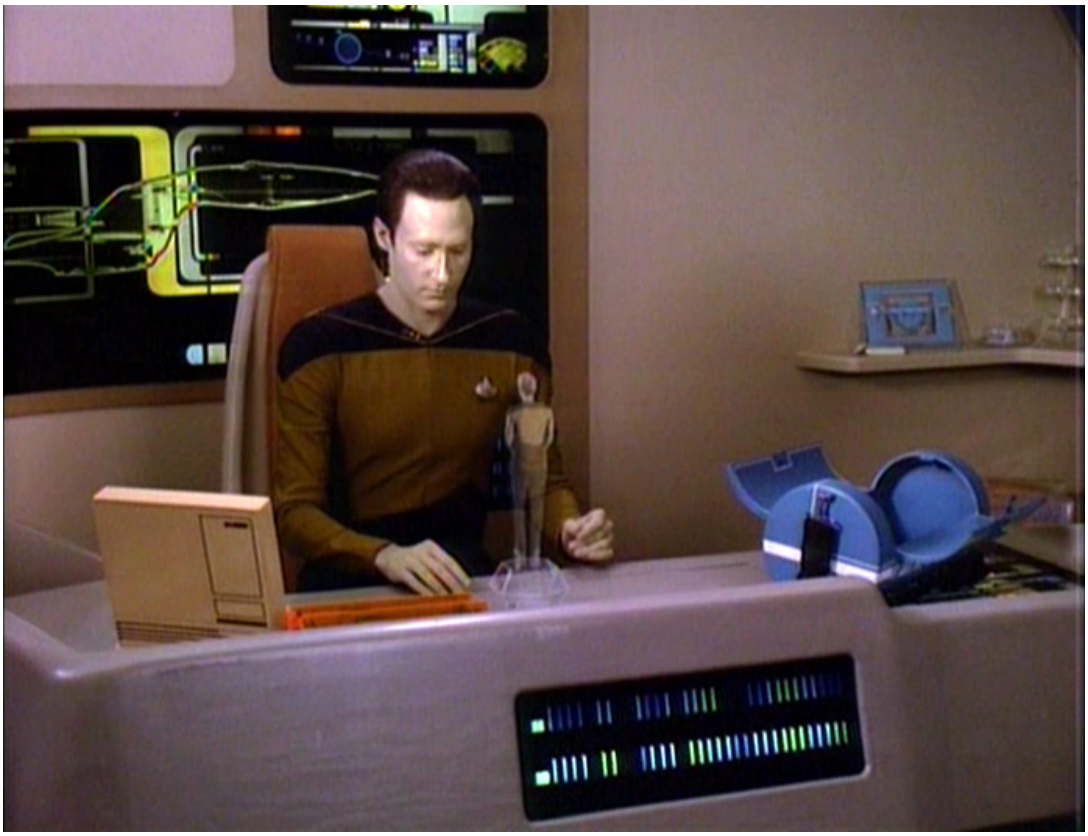


Figure 75. Data muses over a hologram of Tasha Yar, fellow crewmate with whom Data had an intimate encounter in Season One. It is one of the personal possessions Data is asked to explain when questioned about his sentience. Still from ST:TNG *2x09 "The Measure of a Man."*

Data and his friends overcome as a group, as the core crew is quite supportive of Data—a nearly universal phenomenon, as fellow actors Jonathan Frakes (Riker) and Marina Sitiris (Troi) attest.[16] The occasional newcomer to the crew may express hesitation, but quickly grows to respect and support Data after getting to know him. This theme is addressed from the first episode, as Commander Riker, newly assigned to the Enterprise, expresses his discomfort at working with an android. Asked by Picard to include Data on his first away mission, Riker is hesitant, and confronts Data:

RIKER: When the captain suggested you, I looked up your record.

DATA: Yes, sir. A wise procedure, sir, always.

RIKER: Then your rank of Lieutenant Commander is honorary?

DATA: No, sir. Starfleet class of '78; honors in probability mechanics and exobiology.

RIKER (*Surprised*.): Your file says that you're a…

DATA: Machine, Correct, sir. Does that trouble you?

RIKER: To be honest, yes, a little.

DATA: Understood, sir. Prejudice is very human.

RIKER: Now that does trouble me. Do you consider yourself superior to us?

DATA: I am superior, sir, in many ways, but I would gladly give it up to be human.

RIKER: Nice to meet you, (*after a pause*) Pinocchio. (*Data looks at Riker quizzically*.) A joke.

DATA: Ah. Intriguing.

RIKER: You're going to be an interesting companion, Mister Data. (1x01/02, "Encounter at Farpoint")

During the episode Data proves his value to Riker, and the two become friends over the course of the first season. Riker is not the only newcomer to the Enterprise who must be convinced of Data's

[16] Interviews given in Season Six bonus material, "Special Crew Profile: Lt. Commander Data," DVD.

personhood. In Season Two, a leave of absence by Dr. Crusher brings a new medical officer on board, Dr. Katherine Pulaski, who initially presents as anti-technology, repeatedly referring to Data as “it”:

> PULASKI: It does know how to do these things, doesn't it?
>
> PICARD: Commander Data knows precisely what he is doing.
>
> PULASKI: Forgive me, Mister Data. I'm not accustomed to working with non-living devices that… Forgive me again. Your service record says that you are alive. I must accept that. (2x02, “Where Silence has no Lease”)

Pulaski also mispronounces Data’s name (saying “DAT-uh” rather than “DAY-tuh”), and is surprised when he objects:

> PULASKI: What's the difference?
>
> DATA: One is my name. The other is not.
>
> PULASKI: Is this possible? With all of your neural nets, algorithms, and heuristics, is there some combination makes up a circuit for bruised feelings? (2x01, “The Child”)

Pulaski initially considered Data's methodical way of looking at situations to be limited, and believed that he could never solve a traditional Sherlock Holmes mystery to which he did not already know the answer (Data being a fan Doyle knew all his stories from memory). This led to the holographic figure of Moriarty becoming self-aware (2x03, “Elementary, Dear Data”) and later being given his own “holographic universe” in which to live (6x12, “Ship in a Bottle”). After the first few episodes of Season Two, however, Pulaski eventually warms to Data.

A similar pattern unfolds in Season Five, as Data is given the temporary command of his own starship, the USS Sutherland, in a major battle against the Romulans. When Data arrives aboard the ship, his authority is immediately challenged by First Officer Christopher Hobson, who requests a transfer upon Data’s assuming command:

HOBSON: Frankly, sir, I don't believe in your ability to command this ship. You're a fellow officer, and I respect that, but no one would suggest that a Klingon would make a good ship's counsellor or that a Berellian could be an engineer. They're just not suited for those positions. By the same token, I don't think an android is a good choice to be captain.

DATA: I understand your concerns. Request denied. (5x01, "Redemption, Part II")

Hobson continues to question Data's command, behaving in a manner that would warrant a court martial in ordinary circumstances: he gives orders without consulting Data, refuses to follow orders Data gives, and accuses Data of being unconcerned about danger to the crew. The conflict reaches a climax as Data, discovering a way to defeat the Romulans' cloaking devices, disobeys Picard's order to withdraw the ship, much to Hobson's chagrin:

HOBSON: You don't give a damn about the people whose lives you're throwing away. We're not just machines—

DATA: Mister Hobson! You will carry out my orders or I will relieve you of duty.

HOBSON (*through clenched teeth*): Yes, sir. (5x01, "Redemption, Part II")

Hobson's exclamation contrasting the people on board (who have value) to "just machines" (who presumably do not) is particularly telling. However, Data's subsequent actions enable the Sutherland to destroy several threatening Romulan ships, averting a major civil war within the Klingon Empire. Hobson is shocked by Data's accomplishment, and looks at him with newfound respect, addressing Data as "Captain" for the first time.

Data's status as an artificial being is repeatedly called into question throughout the *TNG* timeline, and he and his colleagues are often called upon to prove Data's sentience and defend his rights as a sentient being. Whereas Data's colleagues generally recognize his sentience and accept him as a person, he occasionally encounters those who consider him to be no more than property. In Season

One, Ferengi merchants express a desire to purchase Data; Riker declines, characterizing Data as "second-hand merchandise" (1x09, "The Battle")—which Data interprets as a joke. Similarly, Ira Graves, who appropriated Data's body in order to escape death, justified his actions by stressing that, as Data was not human, had less value:

> PICARD: I understand your desperation. The shock of learning you were dying. But you had no right to do what you have done.
>
> DATA: I had every right, Captain. I am man, he is machine. There is no question who must live and what must die.
>
> PICARD: What of Data?
>
> DATA: Data? Before me, he was nothing. Just a walking tin can with circuits for intestines. Pathetic. Without a heart, a man is meaningless. (2x06, "The Schizoid Man")

Although Graves is unwilling to acknowledge Data's personhood, he does agree to being transferred into the ship's computer, returning control of Data's body to its proper owner. Later, in Season Three, Data is stolen by the Zibalian trader, Kivas Fajo, who is an avid collector of unique objects; Fajo refers to Data as "the crown jewel of my collection" (3x22, "The Most Toys"). Data is eventually recovered by the Enterprise, and Fajo arrested, stripped of his prized collection.

In the year 2365 a tribunal is convened to determine whether Data should be considered the property of Starfleet, precipitated by a visit by Dr. Bruce Maddox, Starfleet cybernetics expert and the Associate Chair of Robotics at the Daystrom Technological Institute. Having studied the work of Dr. Soong, Maddox has been attempting to replicate Soong's lost research, and wants to disassemble and reverse-engineer Data in order to recreate Soong's breakthrough. The proposal comes as a shock to Picard, and in the subsequent debate it becomes obvious that Maddox does not consider Data sentient. Maddox explains his purpose:

> Ever since I first saw Data at the entrance evaluation at the Starfleet Academy, I've wanted to understand it. I became a student of the works of Doctor Noonian Soong, Data's creator, and I've tried to continue his work. I believe I am very close to the breakthrough that will enable me to duplicate Doctor Soong's work and replicate this. But as a first step I must disassemble and study it. Data is going to be my guide. (2x09, "The Measure of a Man")

Throughout the discussion, Maddox refers to Data as "that" and "it," speaking mostly to Picard and Riker even when directly addressed by Data. While Data is initially intrigued by the procedure, further questioning reveals that Maddox cannot guarantee results, and may not be able to restore Data after the procedure—the equivalent of death, were Data considered sentient. Maddox, however, is unconcerned by this possibility, and presents orders to have Data transferred to Maddox' command. As a Starfleet officer, Data may not refuse the transfer, despite the risks; Data instead chooses to resign his commission to avoid the transfer.

Upon hearing of Data's intent to leave Starfleet, Maddox appeals to Captain Phillipa Louvois, the commander of the JAG office on Starbase 173, determined to force Data through legal means to prevent Data's resignation:

> MADDOX: Now you're doing it. Data is an extraordinary piece of engineering, but it is a machine. If you permit it to resign it will destroy years of work in robotics. Starfleet does not have to allow the resignation.
>
> PICARD: Commander, who do you think you're working for? Starfleet is not an organization that ignores its own regulations when they become inconvenient. Whether you like it or not, Data does have rights.
>
> MADDOX: Let me put it another way. Would you permit the computer of the Enterprise to refuse a refit?

LOUVOIS: That's an interesting point. But the Enterprise computer is property. Is Data?

MADDOX: Of course. (2x09, “The Measure of a Man”)

According to Maddox, Data has no more rights than any other piece of Starfleet equipment; Louvois is forced to review the case and offer an opinion. After a review of all the case law, Captain Louvois rules that Data is property and must submit to the procedure. Picard is outraged, and challenges the decision, calling for a formal hearing. Louvois agrees, upon one condition: due to the lack of advocates in the newly-installed JAG office, Picard must act as defense on Data’s behalf, while Riker must prosecute, despite his objections that he considers Data a friend:

RIKER: You just want me to prove that Data is a mere machine. I can't do that because I don't believe it. I happen to know better. So I'm neither qualified nor willing. You're going to have to find someone else.

LOUVOIS: Then I will rule summarily based upon my findings. Data is a toaster. Have him report to Commander Maddox immediately for experimental refit.

RIKER: I see. I have no choice but to agree. (2x09, “The Measure of a Man”)

Riker is sick with the thought of doing harm to Data, having overcome his initial reservations against the android (in 1x01/02) and grown to value Data’s friendship. Nevertheless he does not shirk his duty, and presents a strong case for Data’s status as property:

RIKER: Drawing on the log record of the construction of the prototype android Lore, also constructed by Noonian Soong, I request to be allowed to remove the Commander's hand for your inspection.

PICARD: Objection! (*Stops.*) It doesn't matter. Objection withdrawn.

LOUVOIS: Proceed, Commander.

RIKER (*Softly, to Data*): I'm sorry. (*Riker twists and pulls off Data's left forearm and hand*.) The Commander is a physical representation of a dream, an idea

> conceived of by the mind of a man. Its purpose is to serve human needs and interests. It's a collection of neural nets and heuristic algorithms. Its responses dictated by an elaborate software program written by a man, its hardware built by a man. And now… and now a man will shut it off. (*Riker reaches behind Data and flips his hidden power switch. Data slumps across the table. Louvois gasps*.) Pinocchio is broken. Its strings have been cut. (2x09, "The Measure of a Man")

Riker summarily deactivating Data shocks the courtroom and unequivocally illustrates that Data is a machine, equating him again to Pinocchio, although in a much darker allusion than their first encounter. Picard calls for a recess, disturbed by the dramatic demonstration of Data's nature.

Uncertain how to proceed, Picard turns to Guinan for advice, who addresses Data's predicament not in terms of the implications for Data, but for the thousands of Soong-type androids who would be built were Maddox to be successful:

> GUINAN: Maddox could get lucky and create a whole army of Datas, all very valuable.
>
> PICARD: Oh, yes. No doubt.
>
> GUINAN: He's proved his value to you.
>
> PICARD: In ways that I cannot even begin to calculate.
>
> GUINAN: And now he's about to be ruled the property of Starfleet. That should increase his value.
>
> PICARD: In what way?
>
> GUINAN: Well, consider that in the history of many worlds there have always been disposable creatures. They do the dirty work. They do the work that no one else wants to do because it's too difficult, or to hazardous. And an army of Datas, all disposable, you don't have to think about their welfare, you don't think about how they feel. Whole generations of disposable people.

Figure 76. In a disturbing display of Data's mechanical nature, Commander Riker detaches Data's arm and presents it to JAG officer, Captain Phillipa Louvois. Still from ST:TNG *2x09, "The Measure of a Man."*

> PICARD: You're talking about slavery.
>
> GUINAN: I think that's a little harsh.
>
> PICARD: I don't think that's a little harsh. I think that's the truth. But that's a truth we have obscured behind a comfortable, easy euphemism. Property. But that's not the issue at all, is it? (2x09, "The Measure of a Man")

Picard returns to the same predicament seen in *R.U.R.* nearly a century ago, and the same one that problematizes the treatment of droids in the *Star Wars* universe: if Maddox is successful in his procedure, Starfleet will be responsible for creating a robot subaltern destined to servitude. In effect,

Starfleet would become the equivalent of Weyland Industries in Prometheus, supplying cheap, reliable labor across the Federation, all considered property. For Picard, this is unconscionable.

In arguing for Data's rights as a person rather than property, Picard offers three distinct justifications for Data's personhood. The first is based on Data's possession of an autobiographical self, one of the arguments employed by Brendan Vize. Picard demonstrates this autobiographical self by asking Data to explain the items he packed in anticipation of leaving the Enterprise:

PICARD (Holding a frame with several Starfleet medals.): What are these?

DATA: My medals.

PICARD: Why do you pack them? What logical purpose do they serve?

DATA: I do not know, sir. I suppose none. I just wanted them. Is that vanity?

PICARD (Holding a very old book, the collected works of Shakespeare.): And this?

DATA: A gift from you, sir.

PICARD: You value it?

DATA: Yes, sir.

PICARD: Why?

DATA: It is a reminder of friendship and service.

PICARD (*Activates the hologram of Tasha.*): And this? You have no other portraits of your fellow crewmembers. Why this person?

DATA: I would prefer not to answer that question, sir. I gave my word.

PICARD: Under the circumstances, I don't think Tasha would mind.

DATA: She was special to me, sir. We were intimate. (2x09, "The Measure of a Man")

With each item Picard pulls from the case, Louvois becomes more intrigued by the rich biography of this "toaster," as she had referred to Data earlier in the episode.

Picard next addresses Data's sentience, appealing to many of the same arguments used by Steven Wise of the NhRP: basically, that Data possesses *practical autonomy*, and therefore should be granted personhood and the concomitant right to bodily integrity and to choose his own destiny. For this, Picard questions not Data, but Maddox himself, pressing him to define his terms:

PICARD: Commander, is your contention that Lieutenant Commander Data is not a sentient being and therefore not entitled to all the rights reserved for all life forms within this Federation?

MADDOX: Data is not sentient, no.

PICARD: Commander, would you enlighten us? What is required for sentience?

MADDOX: Intelligence, self-awareness, consciousness. (2x09, "The Measure of a Man")

These three criteria echo those offered by Wise (intentionality, self-sufficiency or –awareness, and consciousness) and Vize (rationality, self-awareness, a personality) in their own discussions of non-human personhood. Most of these terms, however, are tautological: Maddox can no more define one of these terms without resorting to the others than Geordi LaForge could define *anger* for Data without resorting to *hostility* and *belligerence*.

Having forced Maddox to explicitly lay out his criteria for sentience, Picard next applies the "human yardstick" approach to Data—rather than attempting to prove Data sentient, Picard highlights the intractability of the question itself:

PICARD: Prove to the court that I am sentient.

MADDOX: This is absurd! We all know you're sentient.

PICARD: So I am sentient, but Data is not?

MADDOX: That's right.

PICARD: Why? Why am I sentient?

MADDOX: Well, you are self-aware.

PICARD: Ah, that's the second of your criteria. Let's deal with the first, intelligence. Is Commander Data intelligent?

MADDOX: Yes. It has the ability to learn and understand, and to cope with new situations.

PICARD: Like this hearing.

MADDOX: Yes. (2x09, “The Measure of a Man”)

The idea of denying basic intelligence to a computing machine is itself absurd, as Maddox is forced to admit. The second criterion, awareness of self (or *self-sufficiency*, as Wise puts it), is similarly easy to demonstrate:

PICARD: What about self-awareness. What does that mean? Why am I self-aware?

MADDOX: Because you are conscious of your existence and actions. You are aware of yourself and your own ego.

PICARD: Commander Data, what are you doing now?

DATA: I am taking part in a legal hearing to determine my rights and status. Am I a person or property?

PICARD: And what's at stake?

DATA: My right to choose. Perhaps my very life.

PICARD: My rights. My status. My right to choose. My life. It seems reasonably self-aware to me. Commander? I'm waiting.

MADDOX: This is exceedingly difficult. (2x09, “The Measure of a Man”)

By calling upon Maddox to define his terms, Picard forces him to also look at Data on a personal, individual level, rather than allowing him to fall back on comfortable abstractions. Even given Data’s mechanical nature, it is very difficult to argue against the sentience of an entity that so easily demonstrates its existence and insists upon its recognition.

As we have seen in discussions of artificial humans both biological and mechanical throughout this study, Maddox' third measure of personhood, consciousness, is a conundrum in and of itself. Picard addresses this difficulty by rendering the question moot:

> PICARD: You see, he's met two of your three criteria for sentience, so what if he meets the third. Consciousness in even the smallest degree. What is he then? I don't know. Do you? (*to Riker*) Do you? (*to Louvois*) Do you? Well, that's the question you have to answer. (2x09, "The Measure of a Man")

Based on his own performative criteria, Maddox is unable to define sentience in a way that grants the quality to the humans present yet denies it in Data. The point Picard makes, the appeal to Louvois to err on the side of caution, is the same argument Wise proposes: "the more exactly the behavior of any nonhuman resembles ours… the more confident we can be that she possesses desires, intentions, and a sense of self resembling ours, and that we can fairly assign her an autonomy value closer to ours."[17] Data is demonstrably autonomous; Maddox can offer no practical arguments to the contrary.

As Picard has realized (with a gentle nudge from Guinan), the implications of Louvois' ruling on Data's status impact not one individual android, but all the possible artificial life forms who could come after: "a single Data, and forgive me, Commander, is a curiosity. A wonder, even. But thousands of Datas. Isn't that becoming a race? And won't we be judged by how we treat that race?" (2x09, "The Measure of a Man"). Picard's ultimate challenge to Louvois is unequivocal:

> Your Honor, the courtroom is a crucible. In it we burn away irrelevancies until we are left with a pure product, the truth for all time. Now, sooner or later, this man or others like him will succeed in replicating Commander Data. And the decision you reach here today will determine how we will regard this creation of our genius. It will reveal the kind of a people we are, what he is destined to be. It will reach far beyond this

[17] Steven M. Wise, *Drawing the Line: Science and the Case for Animal Rights* (Cambridge, MA: Perseus Books, 2002), 36.

> courtroom and this one android. It could significantly redefine the boundaries of personal liberty and freedom, expanding them for some, savagely curtailing them for others. Are you prepared to condemn him and all who come after him to servitude and slavery? Your Honor, Starfleet was founded to seek out new life. Well, there it sits. Waiting. (2x09, "The Measure of a Man")

Maddox' goal to create thousands of Datas echoes Peter Weyland's TED Talk proclamation that "we are the gods now"[18]—calling to mind the arrogance of Prometheus in creating life without considering the moral implications thereof. Picard relates this moral dilemma to the very premise upon which Starfleet was founded, as expressed in the opening of every episode, not only of *TNG*, but the original series as well: to seek out new life, new civilizations. Picard offers a dire warning against the possible repercussions of a hasty decision—the gravity of which Louvois seems to take to heart as she makes her ruling:

> It sits there looking at me, and I don't know what it is. This case has dealt with metaphysics, with questions best left to saints and philosophers. I'm neither competent nor qualified to answer those. I've got to make a ruling, to try to speak to the future. Is Data a machine? Yes. Is he the property of Starfleet? No. We have all been dancing around the basic issue. Does Data have a soul? I don't know that he has. I don't know that I have. But I have got to give him the freedom to explore that question himself. It is the ruling of this court that Lieutenant Commander Data has the freedom to choose. (2x09, "The Measure of a Man")

In declaring Data "not property" and ruling that he has the freedom to choose whether or not to participate in Maddox' procedure, Louvois has in effect granted Data the right to liberty—and for all

[18] In referring to his accomplishment of creating a prototype cybernetic individual, the David-1, as mentioned in the Introduction. "The Weyland Corporation," http://www.weylandindustries.com/.<http://weylandindustries.com/tedtalk.> Accessed 21 April 2013.

practical purposes acknowledged his claim to personhood, for as Wise argues, *people* have rights while *property* does not. Data formally refuses Maddox' proposal, but leaves open the possibility of their future collaboration—as fellow researchers.

Even after the acknowledgement of Data's personhood, he must fight for his rights on occasion. In the season three, after returning to the Enterprise from a cybernetics conference, Data excitedly introduces his friends to his new "daughter," Lal. His friends are unsurprisingly shocked; even Picard himself has difficulty at first coming to terms with this new development. Counsellor Troi, however, reminds Picard of his own prior mechanist arguments:

> PICARD: I insist we do whatever we can to discourage the perception of this new android as a child. It is not a child. It is an invention, albeit an extraordinary one.
>
> TROI: Why should biology rather than technology determine whether it is a child? Data has created an offspring. A new life out of his own being. To me, that suggests a child. If he wishes to call Lal his child, then who are we to argue? (3x16, "The Offspring")

Data reassures Picard that he has fully considered the implications of creating Lal. Unlike Maddox, who was solely concerned with the functional and economic value of a race of androids, Data equates his own desire to recreate Soong's achievement with the "primal instinct" of a species to perpetuate itself—Data has become a parent rather than a Prometheus. Throughout the episode, the father-daughter relationship between Data and Lal is reinforced, as Data helps Lal struggle to achieve sentience.

While Data's colleagues aboard the Enterprise are very supportive of his endeavor, and help as much as possible with Lal's development and socialization, others are not so convinced. Vice Admiral Anthony Haftel, cybernetics expert from Starfleet Research at Galor IV, considers Data's "project" a breach of Starfleet research policy, and ultimately decides that Lal should be removed to a more suitable research facility. In the process, Haftel reveals his own ambivalent beliefs regarding both androids: while

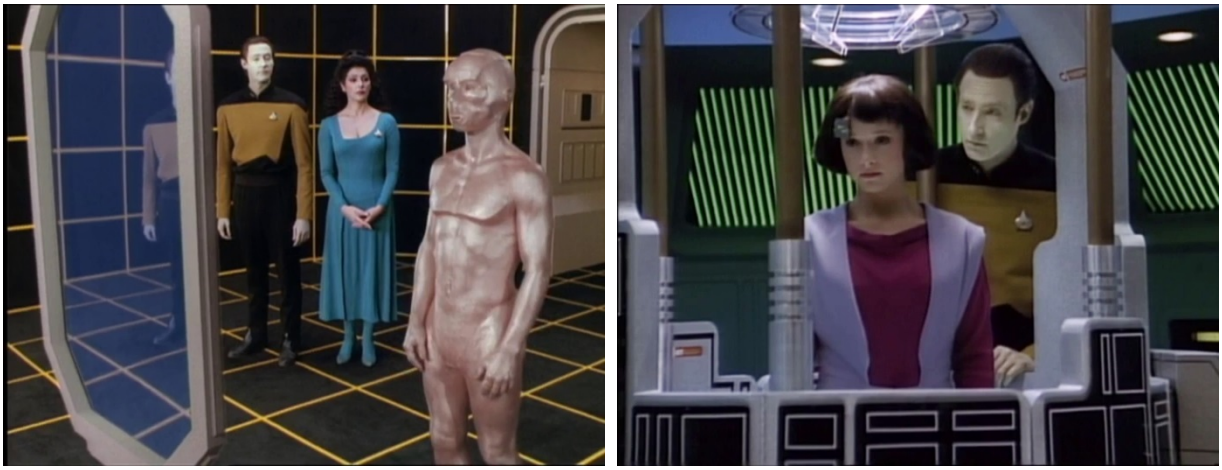


Figure 77. Data's "daughter" Lal, shown (left) before choosing her own gender and race/species identity, and (right) struggling through the process of developing sentience. Stills from ST:TNG *3x16, "The Offspring."*

Haftel readily acknowledges that Data and Lal are sentient beings, he denies them the basic rights appertaining to personhood—liberty and equality—dismissing as "sentimental" both Picard's and Data's talk of family. When Haftel fails to persuade Data to relinquish Lal willingly, he orders Data to turn her over. Picard objects, risking is commission for the sake of what he believes is right:

> HAFTEL: Captain, you are jeopardising your command and your career.
>
> PICARD: There are times, sir, when men of good conscience cannot blindly follow orders. You acknowledge their sentience, but you ignore their personal liberties and freedom. Order a man to hand his child over to the state? Not while I am his captain. (3x16, "The Offspring")

The gesture is ultimately unnecessary, for Lal suffers a cascade failure of her positronic neural network. As Haftel works beside Data to save Lal, he finally realizes the reality of the father-daughter relationship in terms of his own experience as a parent. Though Haftel's threat to remove Lal from the Enterprise was ultimately thwarted by her death, his behavior after attempting to help Data save Lal implies that his position on android rights has altered significantly.

DISTINCTIVENESS

Data is a unique individual within the *Star Trek* universe for several reasons. First is Data's unique position as a Soong-type android, one of only three to function successfully. While his "brother" Lore could be considered a qualified success, his personality renders him a danger to all those around him, as demonstrated in every episode in which he appears. Lore is capable of emotion, but seems to have access to only the negative emotional states that Data lacks; in this he is far more psychopathic than David in *Prometheus*, seeming to take a sadistic pleasure in harming others in pursuit of his own goals. Data is ultimately forced to deactivate and disassemble Lore after realizing Lore's irredeemably malign nature (7x01, Descent, Part II"). The final android by Soong was a copy of his wife, Juliana, built as she lay dying as a result of injuries received while escaping the destruction of the colony on Omicron Theta (7x10, "Inheritance"). The culmination of Soong's brilliance, Juliana is so perfectly built that even she does not know that she is an android, having been designed with a human physiology to fool even medical personnel. It is not until an away mission that her nature is revealed: she falls and is knocked unconscious, losing her arm in the process, which reveals circuitry similar to Data's. Although Data does discover her secret, he ultimately decides not to reveal it to his "mother," allowing her to preserve the illusion of human existence to which Data continually aspired.[19] Juliana dies several years after their meeting, having never learned of her true nature.[20] By the end of the final *TNG* film (*Nemesis*, 2002) no other cyberneticist had overcome the issue of cascade failure that Data faced with Lal, rendering him the only one of his kind.

[19]This is not an easy decision for Data. In a hologram message stored within Juliana's positronic brain, Soong entreats Data to preserve the illusion: "The truth is, in every way that matters, she is Juliana Soong. I programmed her to terminate after a long life. Let her live out her days, and die believing she was human" (7x10, "Inheritance").

[20] According to the canon novel *Immortal Coil* (Jeffrey Lang [New York: Pocket Books, 2002]), Juliana is presumed to have died in 2374, four years after "Inheritance." *Star Trek: Nemesis* takes place in 2379.

Figure 78. Geordi helping to repair Juliana Soong after an injury led to Data discovering that Juliana was an android rather than human—a secret of which Juliana herself is unaware. Still from ST:TNG *7x10, "Inheritance."*

Further, Data is the only android of any type to attend and complete Starfleet Academy, the only non-biological entity to serve as a Starfleet officer, as well as the only one to be officially granted personhood by Federation law. In fact, Data is one of the few artificial life forms of any kind within the ST:TNG timeline.[21] The Enterprise does on occasion encounter other (non-humanoid) mechanical entities that show emergent sentience; in each of these cases Data advocates for their recognition as fervently as Picard did for him. For example, when Data discovers that the small maintenance robots

[21] While other humanoid robots exist within the *Star Trek* universe, particularly in the original series (*ST:TOS*, 1966–1969), these are generally either (1) much less complex than the Soong-type androids, such as Mudd's androids (*TOS* 2x12, "The Deadly Years") or (2) more properly considered cyborg, possessing either biological components (such as the Borg) or uploaded minds of previously biological persons (such as in *TOS* 2x22, "Return to Tomorrow").

known as Exocomps appeared to be developing sentience (6x09, “The Quality of Life”), he successfully argues for their recognition as autonomous life forms; however, these Exocomps would likely rate as high-Category 2 on Wise’s scale of practical autonomy—somewhere between a bottlenose dolphin (just crossing into Category 1 at .90) and an African grey parrot (at .78). But these entities are not humanoid, nor do they possess the same level of *practical autonomy* seen in Data. Their recognition as life forms by Data and his colleagues does not grant them personhood in the way that the Starfleet JAG ruling does for Data himself; it merely acknowledges the possibility that they are (becoming) sentient and should be studied further.

Figure 79. An Exocomp robot, used on the planet Tyrus Vlla to perform routine maintenance in areas too small for a human. When one of these robots refuses to perform, its designer Dr. Farallon believes it is malfunctioning, until the tunnel down which Farallon had tried to send it explodes—which would have been fatal to the robot. Subsequent testing (upon Data’s insistence) shows that not only do these robots make rational decisions but that they communicate with each other. One Exocomp even sacrifices itself in the episode in order to save LaForge and Picard from death. The Exocomps are submitted for further study as an emerging sentient race. Still from ST:TNG *6x09, “The Quality of Life.”*

Perhaps the greatest argument for Data's exceptionality lies in his capacity for growth. Data's path to sentience is framed in terms of development and evolution, mirrored in Lal's own learning process in "The Offspring" (3x22); rather than being "activated" as a fully-functional entity that is (presumably) exactly like every other of his model, as David appears to be in Prometheus, Data's personality—his autobiographical self—develops based first on his interactions with Soong in becoming sentient and later on the experiences and friendships he gains during his time in Starfleet. Soong stressed that he deliberately did not program Data with a specific personality, giving him instead "the ability to choose whatever you wanted" (4x03, "Brothers"); while Soong is disappointed in Data's choice of Starfleet, he does not reprogram him (or label him "malfunctional" as the David-7 models were) but rather recognizes Data's exercise of his ability—his right—to choose his own path.

Actor Brent Spiner attributes Data's developmental capacity to the vision of series creator Gene Roddenberry:

> I think the initial idea with Data—from Gene Roddenberry's perspective, since he'd really created Data—was that he'd be a blank tablet in the beginning, but he was a learning, growing machine and that, by virtue of being amongst these people, amongst this humanity, for as long as we hoped he would be (which was initially six years), that he would grow exponentially through the years, and I think that's exactly what happened.[22]

Soong has built into Data the ability to become more than the sum of his parts. As Data reaches more complex levels of development, new avenues are opened to him. Besides his creativity and artistry, Data gains the ability to dream when a power surge opens previously dormant neural circuits (6x16, "Birthright, Part I"). He also proves himself able to make difficult moral decisions when the situation calls for it, firing on Kivas Fajo after realizing that Fajo's death would prevent numerous others (3x22,

[22] From interview in Season Six extra material: "Mission overview, Special Crew Profile: Lt. Commander Data," DVD

"The Most Toys"). Data eventually decides to install his "emotion chip," giving him the capacity for emotion and the experience of learning to temper them (*Generations*). He also is given the opportunity (and the responsibility) to become a model for others: once for a young boy who, after a painful tragedy, decides to be an android (5.11, "Hero Worship"); another in teaching a supposedly-superior life form (Q) how to be human (3x13, "Deja Q"). After dismissing Data as "the robot who teaches the course in

Figure 80. After being stolen by the notorious collector, Data witnesses Kivas Fajo kill several of his own employees in order to force Data's cooperation. As Data finally turns the tables and is ready to escape, Fajo confronts him, declaring he will kill more people until Data complies—and that it will be Data's fault. Fajo claims that all Data has to do is kill him, but since he is "just an android" he is unable to go against his programming, which prevents him from taking life. Data comes to the moral conclusion that the best way to preserve life would be to kill Fajo ("I cannot permit this to continue.") and fires on him just as he is beamed away. Still from ST:TNG *3x22, "The Most Toys."*

humanities," Q finally makes the concession that "if it means anything to you, you're a better human than I" (3x13, "Deja Q").

PERFORMANCE

Neither Data's physical performance nor his social/emotional performance is perfect. He admits that he is learning, and that he makes mistakes. However, Data's performance is not so much a failure as he believes. As often as Data discredits his own performance as lacking or falling short, there are an equal number of instances in which his performance is praised by friends and colleagues as being right on the money. Data is also very aware of his performance as performance—he continually evaluates his progress and adapts his behavior accordingly, striving to always perfect his performance in order tocome closer to his goal of one day being human. Data is in this respect very like Andrew of the Asimov short story "Bicentennial Man," seeking recognition as a person from those who would deny it to him. Unlike Andrew, however, Data has achieved his goal: his personhood was recognized before the law (albeit in a JAG tribunal), whereas for Andrew, simple "freedom" is not enough, and his recognition as "human" is denied until he pays the ultimate price, proving his mortality by stripping away all that makes him what he is. Andrew's machine nature prevents his recognition as a person; Data is recognized in spite of it. The point is rendered moot by Picard, who declares that "we are, after all, machines, just machines of a different nature," echoing the mechanism/vitalism debate of the seventeenth century.

Data's approach to his performance reflects Goffman's idea of realization through play, as discussed at the outset of our case studies:

> A status, a position, a social place is not a material thing, to be possessed and then displayed; it is a pattern of appropriate conduct, coherent, embellished, and well articulated. Performed with ease or clumsiness, awareness or not, guile or good faith, it

> is none the less something that must be enacted and portrayed, something that must be realized.[23]

Unlike his colleagues, Data cannot take his status for granted; he must continually *enact* his personhood using performative feedback, updating his programming with new variables and subroutines based on audience reception of past behavior. This performance fits with Robert Ezra Park's definition of personhood as a mask, representing "the conception we have formed of ourselves—the role we are striving to live up to—this mask is our truer self, the self we would like to be. In the end, our conception of our role becomes second nature and an integral part of our personality." [24] It is precisely this process that Data describes to Maddox in a letter: "If being human is not simply a matter of being born flesh and blood, if it is instead a way of thinking, acting and feeling, then I am hopeful that one day I will discover my own humanity. Until then Commander Maddox, I will continue learning, changing, growing, and trying to become more than what I am" (4x11, "Data's Day").

DATA AS *HOMO ARTIFICIALIS*

Data is the most fully realized example of Homo artificialis of the case studies under consideration. Returning to Vize's critera for personhood, "interests and goals, the ability to use language, the ability to form reciprocal relationships, the ability to understand and apply moral concepts, rationality, self-awareness, [and] a personality,"25 it is easy to see that Data has achieved a human level of performance in each of these categories, despite not having mastered them all. While Data does occasionally trip over slang, idioms, and contractions in the English language, these mistakes are the type made by a non-native speaker, not a computer (compare Data's understanding of language to that of Watson, for example). On board the Enterprise, Data is surrounded by friends who are fiercely loyal to him, as he is to them. Far from being a "boring robot," as Spiner and others were afraid he could

---

[23] Erving Goffman, *The Presentation of Self in Everyday Life* (Garden City, NY: Doubleday, 1959), 75.
[24] Robert Ezra Park, *Race and Culture*, His Collected Papers, V 1 (Glencoe, Ill,: Free Press, 1950), 249-50; emphasis mine.
[25] Brendan Vize, "Do Androids Dream of Electric Shocks? Utilitarian Machine Ethics" (Master's Thesis, Victoria University of Wellington, 2011), 56-57.

become, Data has a personality distinct enough that his friends can quickly discern an attempt at imitation. Data has a strong ethical sense, and while this is ostensibly hard-programmed into him (6x01, “Descent, Part II”), he is able to override his “fundamental respect for all living beings” when the situation warrants it (3x22, “The Most Toys”). Even Commander Maddox, initially Data’s staunchest opponent in his fight for personhood, is forced to admit that Data is rational and self-aware (2x09, “The Measure of a Man”).

While readily acknowledged by Data’s colleagues, as well as by the franchise in general, Data’s mechanical nature is not typically seen as an obstacle to be overcome in his quest for personhood. Both C-3PO and David are denied personhood not only by individuals with their respective societies but by the very nature of their respective franchises, as both George Lucas and Ridley Scott use vitalist, substratist language to ultimately deny sentience to artificial humans. The Star Trek universe, on the other hand, exhibits strong mechanistic underpinnings with a very progressive posthuman attitude toward personhood. Humanoid alien races are unquestionably considered people, and even animal rights have gained ground: as Commander Riker explains in the first season to a delegation of Anitcans on board for a diplomatic mission, humans “no longer enslave animals for food purposes” but rather eat meat “inorganically materialized out of patterns used by our transporters” via the food replicators (1x07, “Lonely Among Us”). In fact, on several occasions Captain Picard expresses his own rejection of vitalist arguments against mechanical sentience by proclaiming that “we are merely a different variety of machine. In our case, electrochemical in nature” (1x13, “Datalore”), the same argument Picard uses in Data’s defense at his status tribunal in Season Two.

Just as the Creature confronts Frankenstein, and David confronts Holloway, and Shaw seeks to confront the Engineers, Data is given the opportunity to present his creator with the ultimate question:

DATA: Why did you create me?

SOONG: Why does a painter paint? Why does a boxer box? You know what Michelangelo used to say? That the sculptures he made were already there before he started, hidden in the marble. All he needed to do was remove the unneeded bits. It wasn't quite that easy with you, Data. But the need to do it, my need to do it, was no different than Michelangelo's need. Now let me ask you a question. Why are humans so fascinated by old things? (4x03, "Brothers")

Confused at the direction of the conversation, Data offers several possible explanations: ties to the past, a sense of continuity, or a sense of purpose, before hitting upon a realization:

DATA: I suppose it is a factor in the human desire to procreate.

SOONG: So you believe that having children gives humans a sense of immortality, do you?

DATA: It is a reasonable explanation to your query, sir.

SOONG: And to yours as well, Data. (4x03, "Brothers")

Soong relates his building Data not so much to the desire to *create*, but to the need to *procreate*: to leave behind a child to carry on his legacy. It is the same impulse Data has in his attempt to construct Lal. Both C-3PO and David are able to successfully develop relationships—collegial, even friendships—but only Data speaks and is spoken of in familial terms. While Weyland speaks of David as "the son I never had" (*Prometheus*), Soong makes no such qualification of his relationship to Data. To Soong, and to everyone else in the Star Trek universe, Data and Lore are brothers, Soong is their father, Juliana their mother—in every way that matters.

While Data continually doubts his own performance of humanity, those around him recognize its success. Designed to be "barometer, a mirror on humanity,"26 Data offers a reflection of the qualities and idiosyncrasies of those around him, and of what humankind aspires to be. Confronted with

[26] Rick Berman, Executive Producer in Season Six DVD extras: "Mission overview, Special Crew Profile: Lt. Commander Data."

emotionally-charged situations, and unable to respond for himself in a "humanly" appropriate way, both Data's friends and the audience respond for him, as Jonathan Frakes explains:

> Brent has created this character, Data, who everybody loves, and for some ironic reason we feel—we audience members feel emotional when we watch Data react to things as an unemotional being. Brent has found a way to allow the audience, and the fans, and the people who watch him, and the director, and another actor to feel the emotion that his character does not feel, and feel sorry that the character can't feel it.[27]

Klemens Hippel categorizes Data as "an internally conflicted character," claiming that "it could be this very internal contradiction that makes Data so "human," and so popular. In a universe full of figures… who seek out new frontiers, he is the only one who wants to go where the others have all already been, and will just never get there."[28] While I accept Hippel's premise, I reject his conclusion—I believe that Data has reached his destination, just not yet realized it. Far from Pinocchio, who wants to be "a real boy" yet tells lies, behaves obnoxiously, and generally flouts his gift, Data is the Tin Man of *Wizard of Oz* fame, searching for a human heart without realizing that he had one all along.

---

[27] Ibid.

[28] Hippel, "Der Menschlichste Von Uns Allen," 83, translation mine. Original text: „Es könnte gerade diese innere Widersprüchlichkeit sein, die Data so „menschlich" und beliebt macht. In einem Universum voller Figuren, ... die immer unterwegs zu neuen Ufern sind, ist er der einzige, der dahin will, wo die anderen alle schon gewesen sind, und der genau dahin nie kommen wird."

CONCLUSION: THEY WALK AMONG US

There are still many human emotions I do not fully comprehend: anger, hatred, revenge. But I am not mystified by the desire to be loved or the need for friendship. These are things I do understand…. If being human is not simply a matter of being born flesh and blood—if it is instead a way of thinking, acting and feeling—then I am hopeful that one day I will discover my own humanity. Until then, Commander Maddox, I will continue learning, changing, growing, and trying to become more than what I am.

—Data, in a letter to Bruce Maddox; ST:TNG *4x11"Data's Day"*

HOMO ARTIFICIALIS AND HUMAN PERFORMANCE

The performance of humanity actually involves two separate performances. The first is the social/emotional performance, those aspects of humanity that define us as social creatures: reason, emotion, and social interaction. These are the conscious (as well as habitual/unconscious) expressions that the performer "gives" in a particular social situation; the role that he or she plays in these situations in order to facilitate successful social interaction. This performance is the focus of roboticist Cynthia Breazeal, whose various sociable robots are designed to interpret and respond to the non-verbal communication patterns of human interactants in a natural, socially appropriate way. The second, equally important performance is the physical one, that expression which the performer "gives off" in Goffmanian terms. This performance is recognized in the work of Hiroshi Ishiguro, with his telepresent Geminoid robots designed to mimic as closely as possible human physicality, down to the "micromovements" humans make when trying to remain perfectly still. Both these performances are important to the success of Homo artificialis.

The three figures discussed in the previous case studies offer us three contrasting images of the artificial human. While I would unequivocally categorize all three as sentient, the degree to which they successfully perform humanity, and thus fulfill their role as Homo artificialis, is quite varied. C-3PO, while quite limited in his physical performance, offers a social/emotional performance indistinguishable from

a natural human. David, on the other hand, perfectly passes for human physically (when he chooses to do so), yet his depiction of human social and emotional states leaves much to be desired. Many of his actions and comments fall into the realm of aberrant human behavior, which while fitting the most literal definition of human (for even a deranged psychopath is nonetheless "human" in the eyes of the law) is typically labeled as inhuman by modern Western societal standards.

You can assign these two types of performance to the axes of a "matrix of human performance"—again, using the "human yardstick" to which Steven Wise refers. On the horizontal axis we have social performance, negative values indicating no or inadequate performances, postive values successful ones. The boundary between the two indicates autonomy—a robot that is able to perform independently in everyday situations, perhaps even possessing the ability to appear sentient for at least a short period of time. Breazeal notes this in human interactions with her social robot Kismet: participants develop empathy for the robot, feeling guilty when "scolding" it, for example, or being reluctant to shut it down.[1] On the vertical axis lies physical performance, or how closely a mechanical entity resembles a human. Negative values indicate little to no resemblance to humans (industrial robots, for example), with positive values assigned to entities that are easily mistaken for human—Madame Tussauds' wax figures, for example, which uncannily resemble their celebrity counterparts. The boundary for "success" in this continuum is humanoid, an entity with morphology and size intended to replicate the human form. There is a great deal of difference between a Roomba and Ishiguro's Geminoid, with ELEKTRO, roughly human in shape yet unmistakably mechanical in nature, as a good midpoint between the two.

I place Robby the Robot from Forbidden Planet (1956) at the intersection of these two axes, the "zero position" in our analysis. Robby is humanoid, having arms, legs, torso and head in roughly human proportions (as he must, being played by an actor in a robot suit), but lacking even the rudimentary

[1] Cynthia Breazeal, "Regulation and Entrainment in Human-Robot Interaction," *The International Journal of Robotics Research* 21, no. 10-11 (2002).

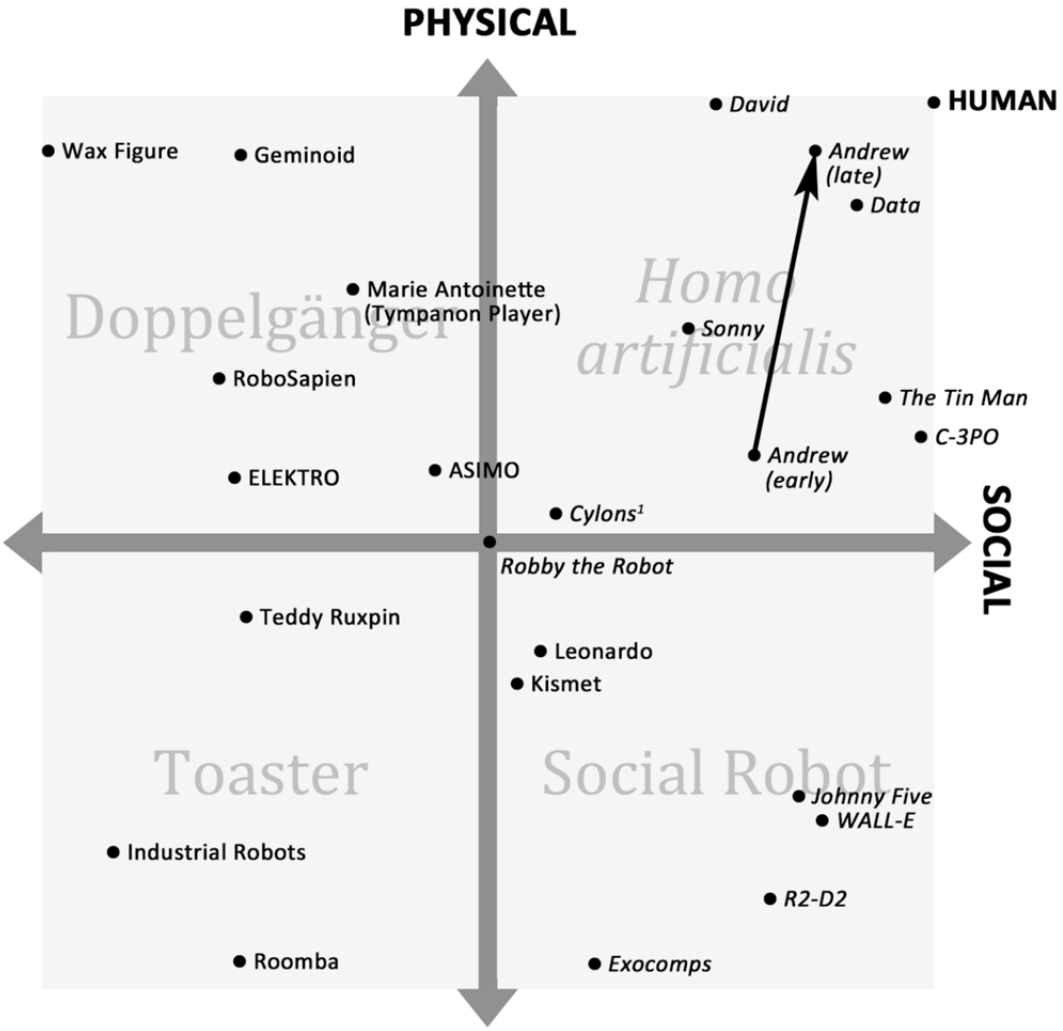


Figure 81. The matrix of human performance. Roman type indicates real-life examples of robots, italics indicate fictional ones.

facial features of C-3PO. Likewise, Robby is autonomous in that he operates in a human environment under the direct command of his creator Doctor Morbius, yet seems to possess none of the higher functions necessary for full sentience, such as reason: Robby simply follows orders; he makes no decisions on his own. He has no desires, and only the faintest glimmer of self-awareness. Similarly, Robby has built no autobiographical life to speak of. However, Robby offers a good (albeit minimal) example of what should be considered sufficiently autonomous and sufficiently humanoid to stand at the Homo artificialis threshold.

Figure 82. Robby the Robot debuted in Forbidden Planet *(1956) but has since made a number of appearances in film and television, particularly through the 1970s, reprising his original role. A replica of Robby was inducted into the Robot Hall of Fame at Carnegie Mellon University in 2004.Theatrical poster for* Forbidden Planet*, featuring Robby the Robot in a situation that never occurs in the film. © 1956 Loew's International, copyright not renewed and is therefore public domain.*

By juxtaposing physical and social performances in a matrix format, four basic paradigms for machine performance of humanity emerge based on the success or failure of these two different performance parameters. The first is the "toaster": a robot that either does not attempt to perform the physical and social aspects of humanity, or fails so badly that we can comfortably dismiss it from consideration. Most modern robots would fall into this category. The second category I have labeled the

"Doppelgänger." These robots bear an uncanny resemblance to natural humans, but their social performance is so inadequate that they would in no way be considered sentient. Ishiguro's Geminoid, as a telepresent robot rather than an autonomous one, falls into the category of "Doppelgänger." The third category, "social robots," covers those entities that seek to imitate human social and emotional behavior without regard to humanoid physicality. Breazeal's work, including Kismet and Leonardo, fall into this realm, as they perform rich social interaction while being intentionally nonhuman in appearance. Finally, we have the category of "*Homo artificialis*" proper. This category is populated by artificial creatures that reasonably pass as humanoid, if not completely human, and operate autonomously to the point where they may be mistaken for sentient for at least a short time. All three of our case study examples fall easily within this category, although at different points on the matrix. Filling the sections with a sampling of mechanical beings available in science fiction and science fact, a pattern emerges, revealing a great deal of variety not only across the matrix as a whole, but within the category *Homo artificialis* as well.

Let us take a more detailed look at a few of the science fiction examples within the Homo artificialis quadrant of the matrix, particularly the ones I previously dismissed from analysis. Andrew, the figure from the film Bicentennial Man (1999) and the 1976 Isaac Asimov novelette upon which it is based, is a classic example of a robot who "wants to be a real boy." Like Data, Andrew longs to be accepted as human by a society that clings to a biological definition of humanity; Andrew is likewise granted his "freedom" early in the narrative,[2] yet faces numerous challenges to his personhood. However, Andrew is ultimately dissatisfied with a performative definition of personhood, as he is still bound by the Three Laws and therefore at the mercy of those who would harm him. Andrew is determined to be declared not merely "equivalent to human" but "really human." He accomplishes this by gradually modifying the legal definition of human through a number of court cases involving natural

[2] Isaac Asimov, "Bicentennial Man," *The Complete Robot* (London: Voyager, 1995), 646.

humans who have received prosthetics—each lawsuit declaring that the possession of particular artificial body parts do not "remove" humanity from the individual. In the meantime, Andrew attempts to conform as closely as possible to his idea of human physicality through the biomechanical replication of human physiology, replacing his own components with the same prosthetics represented in the lawsuits, until the legal system can no longer deny Andrew "human" status without taking that status away from numerous natural humans who happen to possess the same artificial body parts.[3] Andrew's basis for recognition is in the end quite different from Data's: while Data is recognized as a person for his social performance, despite the shortcomings of his physicality, Andrew is ultimately recognized for his physical performance, despite the fact that his social performance was more than adequate.

The Cylons of the Battlestar Galactica universe pose their own set of problems, not the least of which is the dramatic difference in their concept between the original Battlestar Galactica series (1978–1979) and the much-more-successful reboot[4] in the 2000s. As originally imagined, Cylons were a completely mechanical race created by reptilian aliens as a workforce much like the droids of Star Wars. This reptilian race was presumably destroyed by their own creations, who then went on to wage war on the Twelve Colonies of humans due to their intervention in a war the Cylons fought against a third race. Roughly humanoid in appearance (a later model even more closely resembled a human, yet remained mechanical beneath artificial skin), these "toasters" possessed the negative emotions that seem to escape Data, forming an "Evil Empire" against which the last surviving Colonists must defend themselves as they search for the mysterious "Thirteenth Tribe" that supposedly populated Earth. In the reboot,

---

[3] Patrick Lin explores a similar train of thought in his introduction to *Robot Ethics* in discussing the implications of *cyborgs* in contemporary society: "A conscious human brain (and its body) presumably has human rights, and replacing parts of the brain with something else, while not impairing its function, would seem to preserve those rights." Patrick Lin, Keith Abney, and George A. Bekey, *Robot Ethics: The Ethical and Social Implications of Robotics*, Intelligent Robotics and Autonomous Agents (Cambridge, MA: MIT Press, 2012), 8.

[4] The miniseries in 2003, followed by four successful seasons 2004-2009, and a prequel series, *Caprica* (2010), which takes place nearly fifty years before the 2003 miniseries and depicts the downfall of the Twelve Colonies and the rise of the Cylons.

however, the Cylons were originally created by the humans themselves as a servile subaltern,[5] radically changing the power dynamic of the relationship. In addition, these "new" Cylons are no longer purely mechanical; while these versions do exist, the most common Cylons the Colonists face within the newer series are virtually indistinguishable from humans, being biological in nature.[6] The Humanoid Cylons (of which there are also twelve distinct types) are able to not only successfully integrate into human society,

Figure 83. In the opening scene of the 2003 Battlestar Galactica *miniseries, we are first introduced to the new "skinjob" Cylons that are indistinguishable from natural humans. Here, Cylon Six approaches a diplomat from Caprica, flanked by two more traditional "toaster" Cylons (revised, but still quite similar to those of the original 1978–1979 run). Still from* Battlestar Galactica *miniseries, "Night 1."*

[5] The 2003 miniseries opens with a series of titles explaining the backstory, which begins: "The Cylons were created by man. They were created to make life easier on the Twelve Colonies. And then the day came when the Cylons decided to kill their masters" (*Battlestar Galactica* miniseries, "Night 1"). However, the single-season "prequel" series *Caprica* tells a more detailed story of Cylon development. The first Cylon unit was designed for military use: the U-87 Cyber Combat Unit (*Caprica* 1, "Pilot"). However, their potential as laborers was recognized (*Caprica* 5, "There Is Another Sky"), and mechanical Cylons were soon put to use in a number of service positions—by the season finale, characters talked of possible Cylon butlers and nurses in the future. However, this was cut short by a Cylon revolt, instigated in part by the revolutionary Sister Clarice Willow of the Monad Church. The mechanical Cylons are exiled, but unbeknownst to the humans, a new "skinjob" Cylon has been created that is indistinguishable from a human (*Caprica* 18, "Apotheosis"). The exiled Cylons radically improved upon the design during their exile, and returned to Caprica around fifty years later (*Battlestar Galactica* miniseries, "Night 1") on a religious quest (based in their interpretations of the Monad Church's monotheism) to destroy the human race.

[6] In the first season, Dr. Gaius Baltar develops a blood test able to distinguish Cylon from Human, based on the discovery that Cylons are susceptible to certain kinds of radiation to which humans are immune (*Battlestar Galactica* 1x03, "Bastille Day").

but reproduce with them as well. The war between the Cylons and humans in the reboot is based in religion, not in retribution over interference, as the reimagined Cylons consider the humans "sinful." And finally, in a mid-final-season climax, the startling discovery is made that the "Thirteenth Tribe" that colonized Earth consisted not of humans but of Humanoid Cylons[7]—in other words, the Cylons are us!

The Cylons of the original Battlestar Galactica are easy to locate on our matrix of human performance, as seen in the chart above. Physically resembling C-3PO, they offer a social performance that is warlike (or even psychopathic) in nature yet unmistakably human, the "Evil Empire" writ in circuitry and laser technology. But what are we to make of the reimagined Cylons? Are they biomechanical clones, like the Robots of *R.U.R.*? Are they a variation of the human species, interbreeding but ultimately superior, similar to the relationship between Homo sapiens and Homo neanderthalensis in evolutionary history? Humanoid Cylons break the matrix: they are ultimately neither machine nor man, but represent a dire censure of the monster humanity had seemed to become in the early twenty-first century, and a warning against suffering in real life the same apocalyptic fate as the Humanoid Cylons of the series: the Colonists arrive on Earth to discover the Thirteenth Tribe had wiped themselves out in a nuclear holocaust, leaving behind only the empty shell of the former civilization. The example of the Cylons reveals a necessary limitation to the definition of Homo artificialis: that, whether organic or metal, these ersatz humans are all created by artificial means. Once artificial humans venture into the realm of biological reproduction, they are no longer creations but rather genetic descendants, moving off the matrix of the "artificial human" and more properly classified as a species unto themselves.

BECOMING *HOMO ARTIFICIALIS*

Just as Steven Wise of the NhRP advises erring toward cautious generosity in attributing practical autonomy to nonhuman animals, I have left the category of *Homo artificialis* deliberately

[7] *Battlestar Galactica* 4x11, "Sometimes a Great Notion."

broad. However, somewhere between Robby the Robot and Lt. Commander Data, there is a shift in the balance between machine and man. It is the same phenomenon first seen in Frankenstein's Creature, who confronted his creator with the question *Why do I exist*? It is a major theme in the narrative of the artificial human: the emergence of *sentience*, the *self* who questions—emergent not only in the traditional meaning of "that is in process of rising into notice" but in the more particular philosophical sense of "an effect produced by a combination of several causes, but not capable of being regarded as the sum of their individual effects."[8] It is difficult to define but easy to recognize: we see it in the sideways glance of Sonny, the rogue robot in *I, Robot* (2004) as easily as in the tooth-gnashing of the Robots of *R.U.R.* Steven Wise describes this process of "becoming" in his analysis of the practical autonomy possessed by his son Christopher.[9] Wise traces Christopher's development into personhood

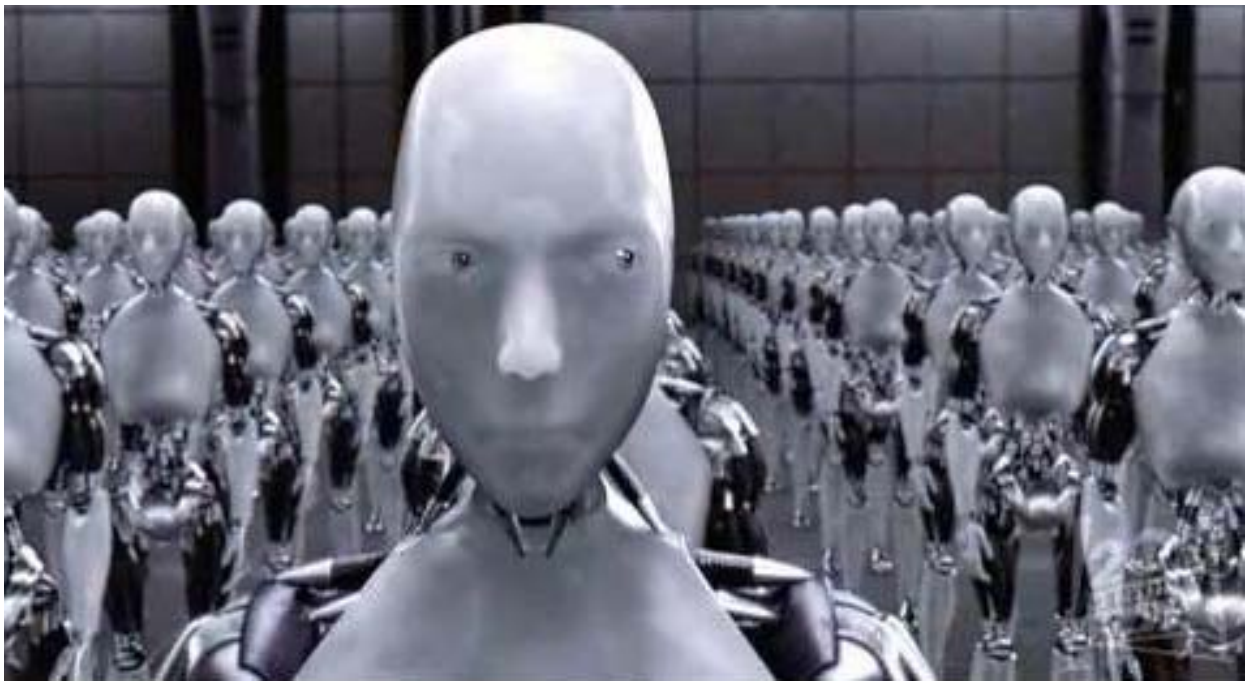


Figure 84. The "awakened" robot Sonny hiding in plain sight among fellow "non-awake" units. Still from I, Robot *(2004).*

[8] Oxford English Dictionary, s.v. "Emergent, Adj. And N," accessed September 12, 2014, http://www.oed.com/view/Entry/61131?redirectedFrom=emergent#eid.
[9] Steven M. Wise, *Drawing the Line: Science and the Case for Animal Rights* (Cambridge, MA: Perseus Books, 2002), Chapter Four, "Christopher," 49-72.

using the theories of *child development* of psychologist Jean Piaget, the concept of *intentionality* of Daniel Dennett and neuroscientist Antonio Damasio, and the concept of *human self* of psychologist Ulrich Nesser, offering a number of insights into how a human child develops into a person of practical autonomy during his first year of life. In science fiction, this is typically depicted as a learning process, taking place over a relatively short period of time, during which the artificial human absorbs vast quantities of information on humanity. We see David apparently going through this process aboard the Prometheus, for example; Data's daughter Lal is similarly depicted (*ST:TNG* 3x16, "The Offspring"), although her "emergence" to sentience is ultimately unsuccessful.

It is obvious from this discussion that there is a great deal of difference within the "*Homo artificialis*" quadrant of the matrix of human performance between minimally-"human" robots such as Robby and the more sophisticated artificial humans seen in the previous case studies. However, as we no longer treat natural humans who "fail to measure up" through mental or physical defect as anything less than fully human, it makes no sense to qualify any artificial human who successfully meets the criteria as anything less than a full member of Homo artificialis, or to fail to grant them the same rights and protections as should be due all members of the class. This reflects Wise's precautionary principle in the treatment of nonhuman animals, derived from the work of philosopher Christina Hoff. The "reason for not using seriously defective human beings in painful biomedical research,"[10] according to Hoff, lies not in their humanity, but in our own: we cannot "safely permit anyone to decide which human beings fall short of worthiness. Judgments of this kind and the creation of institutions for making them are fraught with danger and open to grave abuse."[11] In applying the same precautionary principle to the artificial human as Hoff does to the natural human and Wise does to the nonhuman animal, all robots who demonstrate practical autonomy and successfully meet the criteria for Homo artificialis should be considered nonhuman persons.

---

[10] Ibid., 41.
[11] Christina Hoff, "Immoral and Moral Uses of Animals," *New England Journal of Medicine* 302, no. 2 (1980): 117.

At present, Homo artificialis exists only within the realm of science fiction, where the discussions of the rights of personhood remain a matter of entertainment and of academics. No-one is presently arguing that a Roomba should be covered by labor laws, or that Cynthia Breazeal should not turn off Kismet at the end of a research session. However, if—as some scholars predict—we or future generations of humans are faced with real robots who successfully demonstrate practical autonomy, what then? As Picard exhorts the tribunal Data must face in his quest for the rights of personhood, "how we will regard this creation of our genius… will reveal the kind of people we are, what he is destined to be."[12]

IF HUMANITY IS PERFORMATIVE, HOW DO WE MEASURE UP?

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?

—George Charney, in "Bicentennial Man"[13]

Data continually monitors both his physical and his social/emotional performance of humanity, searching for indications that this performance is successful. Both Data's friends and the viewing audience often find Data's unsuccessful physical performances comical—such as his attempt to master the art of sneezing, as such a behavior serves no practical purpose in an entity that doesn't catch cold (ST:TNG 1x13, "Datalore"). But as Data navigates the ins and outs of human social behavior, his concerns regarding his success are typically met with protests from friends and colleagues that his performance is appropriate, even laudable, representing the best of humanity. Rather than seeking validation for an adequate performance of humanity, Data is often offered—both to other characters within the universe and to the viewing audience—as the model by which others should measure their own "human" performances. Yet another model is offered by David, who has little concern for his fellow

[12] *ST:TNG* 2x09, "Measure of a Man."
[13] Asimov, "Bicentennial Man," *The Complete Robot*, 656.

crewmembers except when it benefits him directly. David's performance is just as human as Data's, yet undesirable as a role model. The question then arises: how successful is our own performance of humanity? In many ways, our own humanity may be judged by how we react to the emerging personhood of Homo artificialis.

Years after developing the Three Laws of Robotics, Isaac Asimov realized how one-sided they made the relationship between robots and humans: while humans had strict protection from robotic harm, robots themselves were completely vulnerable to exploitation or destruction at human hands—unsurprising, actually, as robots are things, and due no protection under the law. But Asimov realized, as we have argued in this study, that if robots are people, the Three Laws not only fail to protect those artificial people but fail as well to represent a version of humanity of which Asimov himself would approve. While a number of Asimov's stories investigate these Three Laws and their consequences, in 1987 Asimov codified the "Three Laws of Humanics," designed as corollaries to the Three Laws of Robotics:

1. A human being may not injure another human being, or, through inaction, allow a human being to come to harm.
2. A human being must give orders to a robot that preserve robotic existence, unless such orders cause harm or discomfort to human beings.
3. A human being must not harm a robot, or, through inaction, allow a robot to come to harm, unless such harm is needed to keep a human being from harm or to allow a vital order to be carried out.[14]

Inspired by what can only be described as a bullying incident in "Bicentennial Man,"[15] these guidelines go a long way toward protecting robots from causal harm. However, these laws still implicitly relegate

[14] Isaac Asimov, "Introduction," in Michael McQuay, *Suspicion*, Isaac Asimov's Robot City (Bryon Preiss Visual Publications, 1987), ix-x.

robots to subaltern status, granting humans more value due solely to their status as human. Vize explicitly argues in support of this organic bias in machine ethics: his "Principle of Inequality" states that "the expected utility value of an action satisfying the preference of a machine is always less than the expected utility value of an action satisfying the equivalent preference of an animal (human or non-human)"[16] due to the fact that we cannot be certain as to the sentience of a machine. Most current discussions of "machine ethics" or "robot ethics" reflect this attitude.[17]

As George A. Bekey outlines in "Current Trends in Robotics: Technology and Ethics," the majority of concerns are focused on the human subject rather than the robot: human loss of privacy due to robot presence, for example, or how the robot should deal with conflicting multiple instructions from different humans.[18] Among the few authors who do concern themselves with the concept of robots as objects of moral consideration, philosophy professor Steve Petersen identifies three trends:

> Some propose that such robots could never be ethical subjects, and so we could not wrong them in making them work for us any more than we now wrong a washing machine. Others agree that robots could not be of ethical significance, but say we must treat them as if they were anyway, for our own sake. Still others conclude that robots could someday have genuine ethical significance similar to ours, and that therefore it would be unethical for them to perform menial tasks for us; it would simply be a new form of slavery.[19]

The first position reflects a substratistic belief that only biological beings may be considered ethical subjects, similar to the speciest argument that only humans may be considered persons. The second view reflects the fears of *inevitable confrontation* described by Minsoo Kang, the "we welcome our

---

[15] Andrew is confronted by a group of young men and ordered to undress and stand on his head; the group were discussing forcing Andrew to dismantle himself when another human intervened. See Asimov, "Bicentennial Man," 651-53.
[16] Brendan Vize, "Do Androids Dream of Electric Shocks? Utilitarian Machine Ethics" (Masters Thesis, Victoria University of Wellington, 2011), 81.
[17] For a recent overview of the field, see Lin, Abney, and Bekey, *Robot Ethics*.
[18] Ibid., 28.
[19] Ibid., 283.

robot overlords"-style preemptive civility. The third paradigm—that of *Homo artificialis* as a people whom it would be unethical to enslave —reflects Picard's address during Data's tribunal, and is based in the precautionary principle Steven Wise uses to argue for nonhuman personhood. However, all three of these positions still depend on the subject's belief or disbelief in the practical autonomy of the artificial human in question.

AI researcher David Levy argues for another model for the treatment of artificial humans, one that relies not on our belief in the personhood of the robots themselves, but on our own ethical composition:

> I believe that the way we treat humanlike (artificially) conscious robots will affect those around us by setting our own behaviour towards those robots as an example of how one should treat other human beings. If our children see it as acceptable behaviour from their parents to scream and shout at a robot or to hit it, then, despite the fact that we can program robots to feel no such pain or unhappiness, our children might well come to accept that such behaviour is acceptable in the treatment of human beings. By virtue of their exhibiting consciousness, robots will come to be perceived by many humans, and especially by children, as being in some sense on the same plane as humans. This is the reasoning behind my argument that we should be ethically correct in our treatment of conscious robots—not because the robots would experience virtual pain or virtual unhappiness as result of being hit or shouted at.[20]

According to Levy, we should treat robots ethically *regardless* of their personhood status, as to do any less reflects negatively on our own personhood. Although this is an admirable position, it does not carry a great deal of legal weight. In many ways this is the same type of argument that has been made in the past against violence in video games: one side argues that the tacit approval of virtual violence leads to

[20] David Levy, "The Ethical Treatment of Artificially Conscious Robots," *International Journal of Social Robotics* 1, no. 3 (2009): 214.

a greater tolerance of and propensity toward violent acts against real people, while the other side claims that as nobody is actually being hurt, no offense has been committed. Such arguments against violent video games hold little sway in case law—at the most resulting in "maturity rating" labels rather than prohibition. I agree with Levy, and I would argue that robots, as embodied agents, are by virtue of their embodiment of much greater concern than those who are purely virtual.

Whether as a result of our own ethical natures or legal restrictions, the development of standards for the treatment of artificial humans is only now beginning to be viewed as a necessity. Previously, commercially available autonomous robot agents have taken the form of toys (such as Robopet) or tools (such as the Roomba), and those developed for military or civil purposes (bomb defusing robots, for example, or the BigDog "military robot mule") are decidedly non-humanoid. However, the past few years have witnessed a push toward artificial humans with commercial applications: autonomous or semi-autonomous humanoid robots designed to work in healthcare and service industries, or for educational or entertainment use. Up to this point, however, these emerging humanoid robots have been too expensive and too limited for widespread commercial deployment,[21] pushing the question of ethics into the background for the time being.

PEPPER THE PERSONAL ROBOT

In a press release dated 2 September 2014, SoftBank CEO Masayoshi Son announced the upcoming commercial release of the humanoid robot Pepper.[22] Pepper is envisioned not as a robot butler or maid, but rather as an interactive social companion. Marketed and produced by SoftBank Robotics, Pepper was initially developed by SoftBank subsidiary Aldebaran Robotics with the specific

---

[21] A typical humanoid robot can cost anywhere from $15,000 to over $100,000 and still does not measure up to the expectations created by its fictional counterparts, requiring considerable maintenance or programming skill, restrictions in environmental parameters, or other special considerations in the human-robot interaction.

[22] Rin Ichino and Takashi Amano, "SoftBank to sell robot in US stores within 12 months," *Bloomberg Online*, Sept 2, 2014, accessed 12 September 2014, http://www.bloomberg.com/news/2014-09-02/softbank-to-sell-robot-in-u-s-stores-within-12-months.html.

goal of creating "a new artificial species"[23]—the Homo artificialis of my study, come to life. According to Aldebaran,

> Pepper is the first humanoid robot designed to live with humans. At the risk of disappointing you, he doesn't clean, doesn't cook and doesn't have super powers....
>
> Pepper is a social robot able to converse with you, recognize and react to your emotions, move and live autonomously.
>
> Engaging and friendly, Pepper is much more than a robot, he's a companion able to communicate with you through the most intuitive interface we know: voice, touch and emotions.[24]

To this end, Pepper is designed to analyze word choice, facial expression, and body language to determine emotional states in his human companions and respond appropriately. He is further able to learn and adapt to his individual human companion over time, developing a unique relationship and a "real personality conveyed by his body language, his funny gestures, and his voice."[25] In stark contrast to previous humanoid robots, Pepper will be affordable to mainstream buyers, as SoftBank plans to price units in the February 2015 Japanese release at 198,000¥ ($1,900) each.[26] While this price point represents an initial loss to the company, developer Bruno Maisonnier (founder of Aldebaran) recognizes the importance of maximizing distribution in order to make Pepper a household name as quickly as possible.[27] Masayoshi Son is personally playing an active role in the project, having outlined in 2010 his long-term vision to create a society that coexists with intelligent robots.[28] The commercial release of Pepper is the first step in making Son's vision a reality.

---

[23] Aldebaran, Inc., "Who is Pepper," accessed 15 October 2014, http://www.aldebaran.com/en/a-robots/who-is-pepper.
[24] Ibid.
[25] Ibid.
[26] Ichino, "Softbank to sell robot."
[27] Ibid.
[28] Masayoshi Son, "SoftBank Next 30-Year Vision," presentation to SoftBank corporate shareholders, 25 June 2010. Video and materials available at "SoftBank: About Us," accessed 15 October 2014, http://www.softbank.jp/en/corp/about/philosophy/next30/.

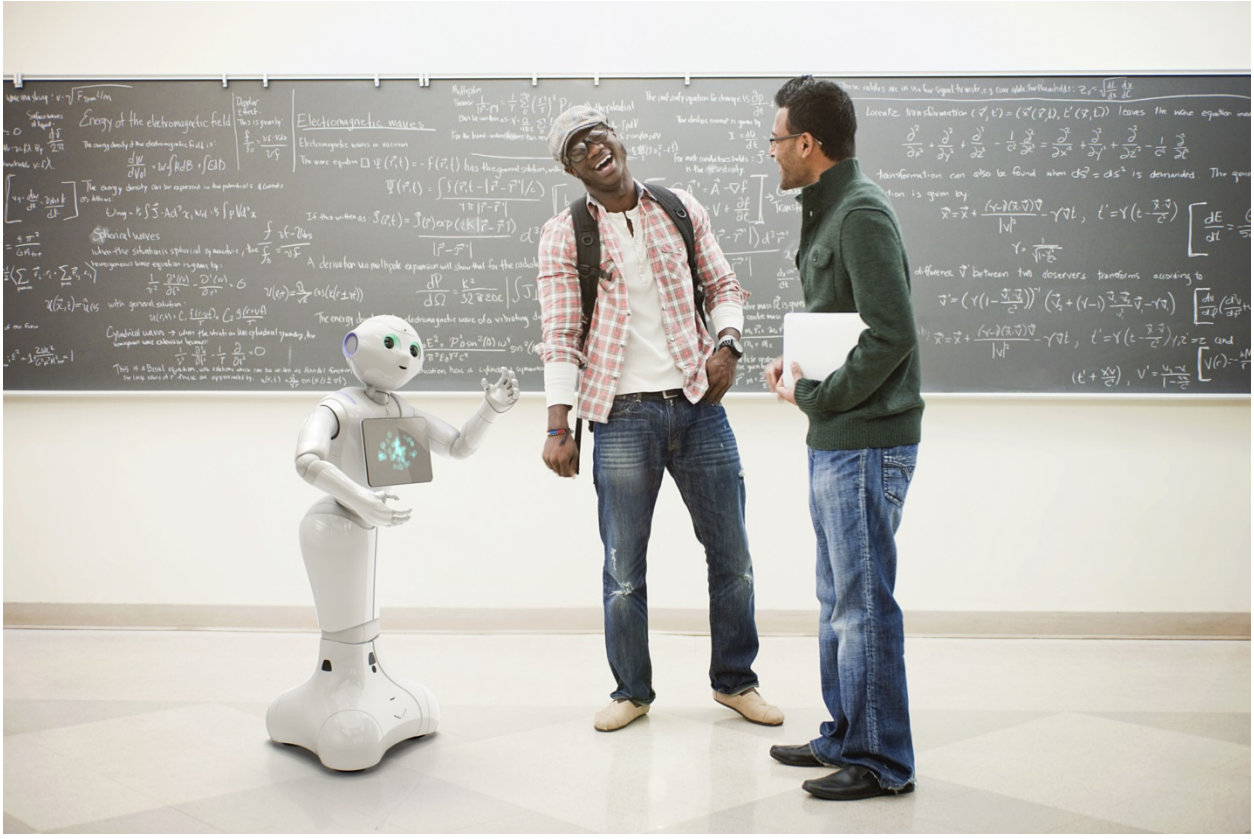


Figure 85. Pepper interacting with humans in this promotional image © and courtesy of Aldebaran.

Pepper is marketed as "the first humanoid robot designed to live with humans." But where will he fall on the matrix of human performance? Will Pepper also be the first non-fictional artificial human to qualify as Homo artificialis? To analyze Pepper from the standpoint of performing humanity, let us first look again at the criteria outlined in the beginning of our case studies, oriented this time toward the real world rather than the world of the film:

1. **PHYSICALITY:** What are the physical characteristics of the being in question? Does the artificial human physically present as human, or is he somehow marked as non-human?
2. **BEHAVIOR:** What are the behaviors the artificial human displays in the performance of human—the mannerisms, verbalizations, etc. that indicate that this being is not just a conscious entity, but an individuated self?
3. **STATUS:** What is the status of artificial humans within society?

4. **DISTINCTIVENESS:** Is the character under consideration a typical or atypical example of artificial humans within society?
5. **PERFORMANCE:** Finally, does the character offer a sincere and effective performance of humanity?

Since Pepper is not yet commercially available, I can only evaluate his performance through the claims of the developers and the videos and anecdotal evidence provided by Aldebaran and SoftBank Robotics.[29] Further, as the first such robot intended for mass commercial use (and not yet available to the public), it makes no sense to discuss Pepper's status and distinctiveness. Nevertheless, I feel we have enough information to make an initial assessment of Pepper based on these sources, although a final determination will of course depend on whether Pepper lives up to SoftBank's claims.

How "human" is Pepper physically? Standing 120cm (about four feet) tall, Pepper is smaller than the average adult human. Furthermore, Pepper is not bipedal but rather mermaid-like; his upper torso narrows to an articulated waist and "hip" area with a single "leg" (which is designed with a center seam to resemble a pair of legs). Below a single articulated "knee," Pepper's glossy, white, stylized chassis ends in a three-point base with omnidirectional wheels, designed for mobility and stability. Pepper's upper body is quite humanoid, however, with head, torso, arms, and hands possessing very human-like articulation and freedom of movement. Pepper has a much greater range of motion in his upper body than C-3PO, his closest case-study counterpart, although Pepper's mode of transport is more limited (it is unclear whether Pepper's wheels are capable of traversing surfaces rougher than commercial-grade carpet). Pepper also shares C-3PO's lack of facial articulation, although LEDs in his eyes and around his ears are able to display a broad range of colors, and can make him appear to blink. Pepper's most distracting non-human physical attribute is the presence of a touchscreen display resembling an iPad

[29] These sources include the Aldebaran website and Bloomberg press release cited above, as well as a number of television commercials for Pepper available from SoftBank Japan (http://www.softbank.jp/mobile/tvcm_media/cm/, accessed 15 October 2014). Pepper is also featured in a number of YouTube videos, both on Aldebaran Robotic's YouTube channel (http://www.youtube.com/channel/UCggm3OyRRI_pF7zA4F4DynA, accessed 15 October 2014) and others.

attached to his chest. This display is used as an additional interactive device: in one video clip Pepper, wearing a red beret and moustache to resemble a "French street artist," uses this display to sketch a likeness of his human interactant, although this is done through the display's internal connection and not "physically" with hand and stylus.[30] Aldebaran describes the tablet as a "way to display images, content or to express Pepper's inner life,"[31] in line with their vision of Pepper as a platform for communication, not an attempt to replicate human life in an artificial substrate. Finally, unlike the sociable robots mentioned in Chapter Three, which are hardwire-tethered to computer and power systems, Pepper is completely autonomous, with a 12-14 hour battery life and remote access to cloud computing and storage.

So far I have referred to Pepper as "he," although the robot is not markedly gendered, possessing no external features that could be considered overtly masculine or feminine (unlike Dot, the C-3PO parody with marked breasts, curly hair and metal skirt, of Spaceballs fame). Pepper's voice is similarly ungendered: the robot's Japanese voice is high-pitched and childlike in tone, fitting with Pepper's diminutive stature. This childlike nature is reinforced by comparison to a third robot developed by Aldebaran: Romeo, designed for research into robots as "assistance for the elderly and those who have lost autonomy."[32] At 140cm tall, Romeo more closely resembles a teenage boy, with shoulders broader than hips and a rounder face than would be typically used to "mark" femininity. Romeo is further referred to as the "big brother" of Pepper and Aldebaran's other robot, Nao, a 58cm tall humanoid also originally designed for research purposes.[33] While the Aldebaran website consistently refers to Pepper as "he" rather than "it" in order to emphasize the robot's humanity, the FAQ page

---

[30] *The Japan Times*, "Pepper Tech Festival 2014," YouTube video, accessed 15 October 2014, http://www.youtube.com/watch?v=XcJccQqTM6Q.
[31] Aldebaran, "Press—FAQ about Pepper," accessed 15 October 2014, http://www.aldebaran.com/en/press/faq-about-pepper.
[32] Aldebaran, "A-Robots—Romeo," accessed 16 October 2014, http://www.aldebaran.com/en/robotics-company/projects.
[33] According to Aldebaran, around 6,000 Nao units have been sold at around $16,000 each, primarily for research and educational use; however, they have recently slashed the price in half in order to "make him available to everyone" (Mike the Robotguide, "NAO and ROMEO robots at Innorobo 2014," *Robot Guide*, 28 March 2014, accessed 16 October 2014,

Figure 86. Left: Pepper's "little brother" Nao, available in a variety of accent colors, is intended as a research tool for classroom use. Right: Romeo, "big brother" to Pepper and Nao. Photos © and courtesy of Aldebaran.

notes that assignment of a gender to Pepper is dependent on the cultural norms of the human interactant: "we noticed that depending on where you come from, people project Pepper to be a male of [sic] a female! In [the] SoftBank shop Pepper can be considered as male."[34] Given this deliberate ambiguity on the part of Pepper's designers, and the lack of a gender-neutral pronoun that does not reduce Pepper to an object, I follow Aldebaran's practice of using "he" in my discussion.

Pepper offers a compelling performance of human physicality within the parameters of his physical form. While he glides across the floor on omnidirectional wheels rather than walking (which both Romeo and Nao are capable of), Pepper's upper body movements are quite fluid and graceful, much more human-like than C-3PO's due to the range of motion afforded by Pepper's construction.

---

http://www.robotguide.com/2014/03/28/aldebaran-robotics-news/nao-romeo-robots-innorobo-2014/; Aldebaran, "A-Robots—Nao: Who is Nao?" accessed 16 October 2014, http://www.aldebaran.com/en/humanoid-robot/nao-robot.

[34] "Press: FAQ about Pepper," accessed 15 October 2014.

Pepper makes eye contact with his human interactants in conversation, and displays attentive listening behaviors even when not actively participating in the interaction: in a video interview with Maisonnier, Pepper sits quietly next to his designer, occasionally shifting focus between Maisonnier, the interviewer, and his surroundings through head tilt and "blinking," just as a person would be expected to do in this situation. In his capacity as "shop assistant" in various SoftBank Mobile commercial locations, Pepper wanders about the space available for interaction, and shifts to "conversation mode" when a potential interactant approaches within his field of vision (around ten feet).

Pepper is so named due to his "sparkling personality,"[35] which Maisonnier built around the belief that, in order to better facilitate human-robot interaction, robots should be cute, humanoid, and kind.[36] Maisonnier describes the ideal robot as "well-intentioned and benevolent," claiming that their purpose is to "combat loneliness"; Pepper's FAQ claim that he will "help people grow, enhance their life, [and] facilitate relationship" through a number of "behaviors" that can be customized according to need. At the moment, Pepper happily greets guests at Softbank Mobile, telling jokes and performing various tricks (like the sketch work mentioned above) in order to entertain customers. However, Pepper is built on an open-source software platform, and developers are encouraged to design and submit new apps via the Pepper "community," resulting in a robot with a highly customizable personality. This personality is also designed to evolve over time, both as the individual robot learns from and adapts to his owner's habits and mannerisms, and as the feedback from these interactions are uploaded to the cloud, allowing all connected Pepper units to benefit from the experience of each individual. Various videos show Pepper further conversing with shop customers and trade show attendees, playing games or drawing sketches on his tablet, answering questions regarding weather and other general topics, suggesting activities, and generally being "social" with small talk and cheerfulness. However, all these

[35] Ibid.
[36] Bruno Maisonnier, Pepper Tech Festival keynote speech, 20 September 2014, video 11:31, on "Community: News: Pepper Tech Festival 2014," accessed 16 October 2014, https://community.aldebaran.com/en/news/pepper-tech-festival-2014.

performances are short, as befits a "social gathering" atmosphere; how successful Pepper's social performance will be in longer interactions, such as in his intended role as home companion, is unclear.

IS PEPPER A "FUZZY PERSON"?

Compared to other real and fictional robots on the matrix of human performance, Pepper initially appears to qualify as Homo artificialis, with slightly better success than Robby the Robot in both physical and social performance. What remains to be seen is whether Pepper is capable of developing what may be accepted by human interactants as a distinct "self"—a biographical life, for example, or a display of practical autonomy as defined by Steven Wise. At this point, such a claim would be grossly premature; therefore, I will have to answer this question with a qualified no. Applying Vize's criteria of personhood to Pepper, we find that Pepper currently demonstrates at least some properties of "behavioural personhood": the ability to use language, the ability to form reciprocal relationships, and a personality. Pepper also appears to possess the ability to apply moral concepts, as he is programmed to be "well-intentioned and benevolent," as Maisonnier claims. Less certain are those abilities which would indicate sentience: intentionality, understanding moral concepts, rationality, and self-awareness. According to Vize's yardstick, Pepper falls into the category of "soft" persons, "beings that have behavioural personhood, but which lack consciousness."[37]

Could Pepper eventually display what could be considered sentience? Probably not. Based on the videos I have seen of Pepper's interactions to this point (approximately an hour and a half of video clips and compilations from various public appearances, interviews, and development work), as well as the descriptions of Pepper's behavioral capacities and software, it is clear to me that it is not the intent of Aldebaran for Pepper to "become sentient," as it were.[38] Intelligent, yes: Pepper appears in videos to be a friendly embodiment of the "personalities" of current smartphones, such as iPhone's Siri, able to

[37] Vize, "Do Androids Dream," 60.
[38] "We do not aim at duplicating live [sic] and humans" ("Press—FAQ about Pepper," accessed 16 October 2014.

perform a number of actions based on natural voice command, learning user behaviors in order to better interact, with an added layer of emotional response and feedback. However, most people seem to interact with Pepper as they would a toy; they are shown laughing and commenting with each other during the interaction in a sort of meta-analysis of the interaction process, behaving in a way that would be quite inappropriate were Pepper an actual object of moral consideration. Rather than accepting his "human" performance as sincere and focusing on the interaction as person-to-person, Pepper's "audience" enjoys his performance of humanity as the focus of the interaction—they are playing with a curiosity, not speaking with a fellow sentient being. Whether this dynamic is likely to change remains to be seen as Pepper and his descendants begin to populate the households of the future.

Figure 87. The Aldebaran "family" of personal robots: Nao, Romeo, and Pepper. Photo © and courtesy of Aldebaran.

A FINAL WORD

Throughout this study I have sought to maintain an agnosticism with regard to the possibility of machine sentience, choosing to focus instead on the external behaviors that make up the performance of what we call human—a performance of personhood that those such as Steven Wise argues is not limited to Homo sapiens, and one seen crossing even the boundaries of biology in contemporary science fiction, as "robot people" fill page and screen alike. I am, however, admittedly biased in my outlook. Unlike John Searle, whose biological naturalism denies that consciousness can arise outside a biological substrate, I ask simply, "Why not?" There is no way to determine if an entity is "really" sentient outside an analysis of his or her successful performance of sentience—the rest is an exercise in metaphysics, a riddle for which there is no definitive answer. While there are currently no actual examples of artificial humans able to successfully perform humanity in the way so many of their fictional counterparts do in the world of our imagination, this (for me, at least) does not preclude the possibility that one day we will work side by side with artificial humans in our everyday life, treating them for all intents and purposes as if they were people.

Minsoo Kang rejects the idea of equivalence through sentience as "affirm[ing] the difference between natural and artificial creatures, human and digital intelligence,"[39] favoring the paradigm of cybernetic mergence—which "rejects the idea that the boundary between the two will remain intact when methods are discovered of freely and directly exchanging parts and information across the line"[40]—as the most likely scenario for our future interactions with intelligent machines. If equivalence through sentience creates a machine subaltern which is ultimately enslaved, humankind loses—and will be judged accordingly, just as Picard predicts in his defense of Data's personhood. If, on the other hand, it eventually proves impossible to directly exchange information between biological and mechanical

[39] Minsoo Kang, *Sublime Dreams of Living Machines: The Automaton in the European Imagination* (Cambridge, MA: Harvard University Press, 2011), 302.
[40] Ibid., 303.

intelligences in the way Kang implies, then true cybernetic mergence can never truly come about. What I propose, therefore, offers a combination of both possibilities—equivalence through practical autonomy, a performative definition of personhood that grants basic "humanity" to any entity able to successfully perform its criteria, regardless of species, substrate, or even planet of origin—should the prospect of alien life ever cross over from science fiction to science fact in the way robotics have in the past decades. A biological human (read: person) is Homo sapiens, and artificial one Homo artificialis—who knows what possible people we may yet encounter, based on their performance?

A performative definition of personhood, determined at least in part by the successful display of practical autonomy as outlined by Steven Wise in the Nonhuman Rights Project, removes the metaphysical debate from consideration, and could possibly prevent the disastrous alternative futures Kang envisions. It also addresses the basic anti-robot sentiments depicted in Isaac Asimov's "Bicentennial Man," as discussed in earlier in this chapter. Over the course of two centuries, Andrew first achieves a form of equivalence through sentience as his right to freedom (but not his personhood) is finally granted through the courts after a long and hotly contested legal battle. Andrew also ultimately succeeds in engendering a paradigm shift of the cybernetic mergence sort, as the biomechanically equivalent body parts he creates for his own transformation become the backbone of the medical prosthetic industry. This second legal battle is much longer and more subtle than the first, as Andrew's lawyers use the same tactic as Picard, arguing (unsuccessfully, of course) that various people no longer qualify as "human" due to their prosthetic enhancements. Just as Maddox could not define sentience in a way that denied it to Data while granting it to the humans in the tribunal, the courts in "Bicentennial Man" could no longer define "human" in a way that denied it to Andrew yet granted it to those who benefitted from his mechanical creations. Unfortunately, Andrew died the only one of his kind, as leaders in the robotics industry made changes in their production process to ensure that another such "accident" (Andrew's sentience) could never happen. But what if Andrew had achieved equivalence

through practical autonomy, and been granted status as a legal person rather than only his freedom? While Andrew may have faced the same prejudice within the society in which he lived, he could have received been spared the second round of legal battles, having been declared a "man" in the first case.

What does it matter? We stand at the verge of a new era of human achievement. Since the dawn of recorded history mankind has sought to create life, becoming the deus faber in his own right. Depictions of this "creator god" aspiration range from Prometheus the mythological figure to Prometheus the Weyland Industries spaceship, from clay forms imbued with the holy breath of life to the sophisticated electromechanical structures who somehow cross the threshold into self-awareness. Artificial doubles of human beings have filled our imagination for thousands of years, shaping our very consciousness as we have strived to make our dreams reality. As we emerge into the era of what N Katherine Hayles terms the "already posthuman,"[41] our definition of what this term "human" actually means must evolve to catch up. I argue that this definition is of necessity a performative one.

[41] N. Katherine Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics* (Chicago: University of Chicago Press, 1999).

BIBLIOGRAPHY

Adams, Douglas. *The Hitchhiker's Guide to the Galaxy*. First American edition. New York: Harmony Books, 1980.

Adams, William Howard. *The French Garden, 1500–1800*. World Landscape Art & Architecture Series. New York: Braziller, 1979.

Advanced Telecommunications Research Institute International (ATR) and Hiroshi Ishiguro. "Geminoid." Last accessed 15 November 2014. http://www.geminoid.jp/projects/kibans/index.html.

Agrippa, Heinrich Cornelius. *De occulta philosophia* [*Three Books of Occult Philosophy*]. Translated by J.F. Volume 2. London: R.W. for Gregory Moule, 1651.

al-Jazarī, Ismā'īl ibn al-Razāz. *Kitāb Fī Ma'rifat Al-Ḥiyal Alhandasiyya* [*The Book of Knowledge of Ingenious Mechanical Devices*]. Translated [from the Arabic] and annotated by Donald R. Hill. Turkey: 1206; Dordrecht: Reidel, 1974.

Allen, Terry. *Five Essays on Islamic Art*. Manchester, MI: Solipsist Press, 1988.

Allston, Aaron. *Conviction*. Star Wars: Fate of the Jedi series. London: Random House, 2012.

Aldebaran, Inc. "Aldebaran Robotics: Humanoid Robotics & Programmable Robots." Last accessed 15 November, 2014. http://www.aldebaran.com/en.

American Honda Motor Company. "ASIMO: the world's most advanced humanoid robot." Accessed 8 April 2014. http://asimo.honda.com/.

Amy, Michael. "The Body as Machine, Taken to Its Extreme." *New York Times,* 20 January 2002. http://www.nytimes.com/2002/01/20/arts/art-architecture-the-body-as-machine-taken-to-its-extreme.html.

Anderson, Alan Ross. *Minds and Machines*. Contemporary Prospects in Philosophy series. Englewood Cliffs, NJ: Prentice-Hall, 1964.

Anderson, Michael L. "Why Is AI So Scary?" *Artificial Intelligence* 169, no. 2 (Dec 2005): 201-08.

Apollodorus. *The Library.* The Loeb Classical Library. Translated by Sir James George Frazer. London: William Heinemann, 1921.

Aristotle. *De Anima* [*On the Soul*]. Translated by Hugh Lawson-Tangred. London: Penguin Books, 1986. Kindle edition.

Arnott, Christopher. "Save the Robots: A Westport Playwright/Producer Rocks with Metal People." *Orlando Sentinel*, 07 May 2012. http://www.orlandosentinel.com/wtxx-save-the-robots-a-westport-playwrightproducer-rocks-with-metal-people-20120507-story.html.

Arp, Robert. "'If Droids Could Think...': Droids as Slaves and Persons," in *Star Wars and Philosophy*, ed. Kevin S. Decker and Jason T. Eberl, 120-131.

Asimov, Isaac. *The Complete Robot*. London: Voyager, 1995.

———. *The Rest of the Robots*. Garden City, NY: Doubleday, 1968.

———. *Robot Visions*. New York: Penguin, 1990.

Babbage, Charles. *Passages from the Life of a Philosopher*. London: Longman, Green, Longman, Roberts, & Green, 1864.

Bailly, Christian, Sharon Bailly, and Constantina Rhodes Bailly. *Automata: The Golden Age 1848–1914*. London: Sotheby's Publications, 1987.

Bann, Stephen, ed. *Frankenstein, Creation, and Monstrosity*. Critical Views series. London: Reaktion Books, 1994.

Barrett, Michèle, and Duncan Barrett. *Star Trek: The Human Frontier.* New York: Routledge, 2001.

Bartneck, Christoph, and Jodi Forlizzi. "A Design-Centred Framework for Social Human-Robot Interaction." *RO-MAN 2004 - The 13th IEEE International Symposium on Robot & Human Interactive Communication* (Jan 2004): 591-94.

Bartneck, Christoph, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. "Is the Uncanny Valley an Uncanny Cliff?" *RO-MAN 2007 - The 16th IEEE International Symposium on Robot & Human Interactive Communication* (Jan 2007): 368-73.

———. "My Robotic Doppelgänger—a Critical Look at the Uncanny Valley." *RO-MAN 2009 - The 18th IEEE International Symposium on Robot & Human Interactive Communication* (Jan 2009): 269-276.

Baudrillard, Jean. *The System of Objects.* Translated by James Benedict. Editions Gallimard 1968; London: Verso, 1996.

Bedini, Silvio A. *Patrons, Artisans, and Instruments of Science, 1600–1750*. Brookfield, VT: Ashgate/Variorum, 1999.

———. "The Role of Automata in the History of Technology." *Technology and Culture* 5, no. 1 (1964): 24-42.

Bekey, George A. "Current Trends in Robotics: Technology and Ethics." In *Robot Ethics,* edited by Patrick Lin, et. al., 17-34.

Benjamin, Walter. *Illuminations*. Translated by Hannah Arendt. London: Random House, 2011.

Berryman, Sylvia. “Ancient Automata and Mechanical Explanation.” *Phronesis* 48, no. 4 (2003): 344-69.

———. “The Imitation of Life in Ancient Greek Philosophy.” In *Genesis Redux,* edited by Jessica Riskin, 35-45.

———. *The Mechanical Hypothesis in Ancient Greek Natural Philosophy*. New York: Cambridge University Press, 2009.

Blinder, Barton J. “The Autobiographical Self: Who We Know and Who We Are.” *Psychiatric Annals* 37, no. 4 (2007): 276-84.

Borelli, Giovanni Alfonso. *De motu animalium* [*On the Movement of Animals*]. Translated by Paul Maquet. 1710; New York: Springer-Verlag, 1989.

Bould, Mark, Andrew M. Butler, Adam Roberts, and Sherryl Vint, eds. *Fifty Key Figures in Science Fiction*. Routledge Key Guides. London: Routledge, 2010.

Breazeal, Cynthia. *Designing Sociable Robots*. Intelligent Robots and Autonomous Agents. Cambridge, MA: MIT Press, 2002.

———. “Emotion and Sociable Humanoid Robots.” *International Journal of Human-Computer Studies* 59, no. 1-2 (Jul 2003): 119-55.

———. “Regulation and Entrainment in Human-Robot Interaction.” *The International Journal of Robotics Research* 21, no. 10-11 (2002): 883-902.

———. “Research: Leonardo,” accessed 9 April 2014, http://web.media.mit.edu/~cynthiab/research/research.html.

———. “The Rise of Personal Robots.” Filmed December 2010 at TEDWomen 2010. TED video, 14:04. http://www.ted.com/talks/cynthia_breazeal_the_rise_of_personal_robots.

———. “Social Interactions in HRI: The Robot View.” *IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews* 34, no. 2 (May 2004): 181-86.

———. “Toward Sociable Robots.” *Robotics and Autonomous Systems* 42, no. 3-4 (Mar 2003): 167-75.

Breazeal, Cynthia, Jessie Gray, Guy Hoffman, and Matt Berlin. “Social Robots: Beyond Tools to Partners.” *RO-MAN 2004 - The 13th IEEE International Symposium on Robot & Human Interactive Communication* (Sep 2004): 551-56.

Breazeal, Cynthia, and Brian Scassellati. “Robots That Imitate Humans.” *Trends in Cognitive Sciences* 6, no. 11 (Nov 2002): 481-87.

Bresadola, Marco. “Medicine and Science in the Life of Luigi Galvani (1737–1798).” *Brain Research Bulletin* 46, no. 5 (1998): 367-80.

Brett, Gerard. “The Automata in the Byzantine ‘Throne of Solomon.’” *Speculum* 29, no. 3 (1954): 477-87.

Brooks, Rodney Allen. “Elephants Don’t Play Chess.” *Robotics and Autonomous Systems* 6 (1990): 3-15.

———. *Flesh and Machines: How Robots Will Change Us*. New York: Pantheon Books, 2002.

Broudy, Eric. *The Book of Looms: A History of the Handloom from Ancient Times to the Present*. Hanover: Brown University Press, 1993.

Brown, Brian A. *Three Testaments: Torah, Gospel, and Quran*. Lanham: Rowman & Littlefield Publishers, 2012.

Bruce, J. Douglas. "Human Automata in Classical Tradition and Mediaeval Romance." *Modern Philology* 10, no. 4 (1913): 511-26.

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, no. 3 (2013): 759-71.

Burroughs, Edgar Rice. *Synthetic Men of Mars*. Tarzana, CA: E. R. Burroughs, 1940.

Butler, Judith. *Gender Trouble: Feminism and the Subversion of Identity*. Thinking Gender. New York: Routledge, 1990.

Campbell, Blair. "La Mettrie: The Robot and the Automaton." *Journal of the History of Ideas* 31, no. 4 (1970): 555-72.

Campbell-Kelly, Martin, William Asprey, Nathan Ensmenger, and Jeffery R. Yost. *Computer: A History of the Information Machine*. The Sloan Technology Series. Boulder, CO: Westview Press, 2013. Kindle edition.

Čapek, Karel, and Peter Kussi. *Toward the Radical Center: A Karel Čapek Reader*. Highland Park, NJ: Catbird Press, 1990.

Čapek, Karel. *R.U.R. (Rossum's Universal Robots)*. Translated by Paul Selver and Nigel Playfair. Dover Thrift Editions. Garden City, NY: Doubleday, 1923; Mineola, NY: Dover Publications, 2001. Kindle edition.

Carrera, Roland, Dominique Loiseau, and Oliver Roux. *Androids: The Jaquet-Droz Automations*. Lausanne, Switzerland: Scriptar, 1979.

Casimir, Viviane. "Data and Dick's Deckard: Cyborg as Problematic Signifier." *Extrapolation: A Journal of Science Fiction and Fantasy* 38, no. 4 (1997): 278-91.

Cassell, Justine, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill, eds. *Embodied Conversational Agents*. Cambridge, MA: MIT Press, 2000.

Cathédrale Notre Dame De Strasbourg. "L'horloge Astronomique." Accessed 15 October 2013. http://www.cathedrale-strasbourg.fr/horloge_astro.aspx.

Ceruzzi, Paul E. *Computing: A Concise History*. Cambridge, MA: MIT Press, 2012.

Chaminade, Thierry, Jessica Hodgins, and Mitsuo Kawato. "Anthropomorphism Influences Perception of Computer-Animated Characters' Actions." *Social Cognitive & Affective Neuroscience* 2, no. 3 (2007): 206-16.

Christian, Brian. *The Most Human Human: What Talking to Computers Teaches Us About Being Alive*. New York: Random House, 2011.

Clark, Thomas. "Fear of Mechanism: A Compatibilist Critique of 'the Volitional Brain.'" *Journal of Consciousness Studies* 6, no. 8-9 (1999). Special Issue: The Volitional Brain: Towards a Neuroscience of Free Will: 279-93.

Clark, William, Jan Golinski, and Simon Schaffer. *The Sciences in Enlightened Europe*. Chicago: The University of Chicago Press, 1999.

Cohen, John. *Human Robots in Myth and Science*. London: George Allen & Unwin, Ltd., 1966.

Crave Online Media. *ComingSoon.net*. "Brent Spiner Rules Out *Star Trek XI*," Accessed 2 September 2014. http://www.comingsoon.net/films.php?id=16437.

Cooney, Brian. *Posthumanity: Thinking Philosophically About the Future*. Lanham, Md.: Rowman & Littlefield, 2004.

Copeland, B. Jack, and Alan Mathison Turing. *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, Plus the Secrets of Enigma*. Oxford: Oxford University Press, 2004.

Cray: The Supercomputer Company. "Cray History." Accessed 2 September 2014. http://www.cray.com/About/History.aspx.

Crocker, Lester G. *Diderot's Chaotic Order: Approach to Synthesis*. Princeton, NJ: Princeton University Press, 1974.

Culbertson, James T. *The Minds of Robots; Sense Data, Memory Images, and Behavior in Conscious Automata*. Urbana: University of Illinois Press, 1963.

D'amico, Lisa N Edmonds. "'I Suggest You May Be Human': Identity in Select Works of Robert a Heinlein and Frank Herbert." *The Heinlein Journal* 20, no. January 2007 (2007): 19-22.

Daston, Lorraine. "Marvelous Facts and Miraculous Evidence in Early Modern Europe." *Critical Inquiry*, no. 1 (1991): 93-124.

Dautenhahn, Kersten. "The Art of Designing Socially Intelligent Agents: Science, Fiction, and the Human in the Loop." *Applied Artificial Intelligence* 12, no. 7-8 (Oct-Dec 1998): 573-617.

———. "Getting to Know Each Other—Artificial Social Intelligence for Autonomous Robots." *Robotics and Autonomous Systems* 16, no. 2-4 (Dec 1995): 333-56.

———. "Socially Intelligent Robots: Dimensions of Human-Robot Interaction." *Philosophical Transactions of the Royal Society B-Biological Sciences* 362, no. 1480 (Apr 29 2007): 679-704.

Dear, Nick. *Frankenstein, Based on the Novel by Mary Shelley*. London: Faber & Faber, 2011. Kindle edition.

Decker, Kevin S., and Jason T. Eberl. *Star Wars and Philosophy: More Powerful Than You Can Possibly Imagine*. Popular Culture and Philosophy series. Chicago: Open Court, 2005.

Dee, John. *The Mathematicall Praeface to the Elements of Geometrie of Euclid of Megara (1570)*. Primary Sources from the Scientific Revolution. New York: Science History Publications, 1975.

Deleuze, Gilles, and Felix Guattari. *A Thousand Plateaus: Capitalism and Schizophrenia*. Translated by Brian Massumi. Minneapolis: University of Minnesota Press, 1987.

Delvoye, Wim. "Cloaca." Accessed 23 January 2014. http://www.wimdelvoye.be/.

Déniz, Oscar, Mario Hernández, Javier Lorenzo, and Modesto Castrillón. "An Engineering Approach to Sociable Robots." *Journal of Experimental & Theoretical Artificial Intelligence* 19, no. 4 (2007): 285-306.

Dennett, Daniel C. "True Believers: The Intentional Strategy and Why It Works." In *Mind Design II*, edited by John Haugeland, 57-80.

Descartes, René. *The World and Other Writings.* Translated and edited by Stephen Gaukroger. Cambridge: Cambridge University Press, 1998.

———. *The Philosophical Works of Descartes.* Translated by Elizabeth S. Haldane and G.R.T. Ross. Cambridge: Cambridge University Press, 1911; 1978.

———. *Discourse on the Method and the Meditations*. Translated by John Veitch. New York: Cosimo, 1924, 2008.

Dijksterhuis, E. J. *The Mechanization of the World Picture: Pythagoras to Newton*. Translated by C. Dikshoorn. Princeton, NJ: Princeton University Press, 1986.

Dolecki, Jana. *Strasbourg.info*. "Strasbourg > Sights > Cathedral." Last updated 14 November 2014. http://www.strasbourg.info/cathedral/.

Doyon, André, and Lucien Liaigre. *Jacques Vaucanson, Mécanicien De Génie*. Paris: Presses Universitaires de France, 1967.

Dragone, Mauro, Brian R. Duffy, and Gregory M.P. O'Hare. "Social Interaction between Robots, Avatars and Humans." *IEEE International Workshop on Robot and Human Interactive Communication, RO-MAN 2005* (Aug 2005): 24-29.

Dreyfus, Hubert L. *What Computers Can't Do; a Critique of Artificial Reason*. New York: Harper & Row, 1972.

Dubois, Pierre. *Histoire de l'horlogerie depuis son origine jusqu'à nos jours, précédée de recherches sur la mesure du temps dans l'antiquité et suivie de la biographie des horlogers les plus célèbres de l'europe*. Paris: Administration du Moyen âge et la renaissance, 1849.

Duffy, Brian R. "Anthropomorphism and the Social Robot." *Robotics and Autonomous Systems* 42, no. 3-4 (2003): 177-90.

Dunn, Thomas P., and Richard D. Erlich. *The Mechanical God, Machines in Science Fiction*. Contributions to the Study of Science Fiction and Fantasy, No. 1. Westport, CT: Greenwood Press, 1982.

Dyens, Ollivier. *Metal and Flesh: The Evolution of Man: Technology Takes Over*. Cambridge, MA: MIT Press, 2001.

Echterhoff, G., G. Bohner, and F. Siebler. "Social Robotics and Human-Machine Interaction: Current Research and Relevance for Social Psychology." *Zeitschrift Fur Sozialpsychologie* 37, no. 4 (Dec 2006): 219-31.

Epley, Nicholas, Adam Waytz, and John T. Cacioppo. "On Seeing Human: A Three-Factor Theory of Anthropomorphism." *Psychological Review* 114, no. 4 (2007): 864-86.

Erlich, Richard D., and Thomas P. Dunn. *Clockworks: A Multimedia Bibliography of Works Useful for the Study of the Human/Machine Interface in Sf*. Bibliographies and Indexes in World Literature, No. 37. Westport, CT: Greenwood Press, 1993.

Fildes, Jonathan. "Campaign Builds to Construct Babbage Analytical Engine." *BBC News: Technology* (2010). Published 14 October 2010. http://www.bbc.co.uk/news/technology-11530905.

Firebaugh, Morris W. *Artificial Intelligence: A Knowledge-Based Approach*. Boston: Boyd & Fraser, 1988.

Fodor, Jerry A. *Psychological Explanation: An Introduction to the Philosophy of Psychology*. New York: Random House, 1968.

Fong, Terrence, Illah Nourbakhsh, and Kerstin Dautenhahn. "A Survey of Socially Interactive Robots." *Robotics and Autonomous Systems* 42, no. 3-4 (2003): 143-66.

Forlizzi, Jodi. "How Robotic Products Become Social Products: An Ethnographic Study of Cleaning in the Home." 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2007 (Mar 2007): 129-136.

Foucault, Michel. *The Archaeology of Knowledge*. Routledge Classics. Translated by A. M. Sheridan Smith. Editions Gallimard, 1969; New York: Routledge, 2002.

Franchi, Stefano, and Gèuven Gèuzeldere. *Mechanical Bodies, Computational Minds: Artificial Intelligence from Automata to Cyborgs*. Cambridge, MA: MIT Press, 2005.

Fryer, David M., and John C. Marshall. "The Motives of Jacques De Vaucanson." *Technology and Culture* 20, no. 2 (1979): 257-69.

Fuegi, John and Jo Francis. "Lovelace & Babbage and the Creation of the 1843 'Notes.'" *IEEE Annals of the History of Computing* 25, no. 4 (2003): 16-26.

Fukuyama, Francis. *Our Posthuman Future: Consequences of the Biotechnology Revolution*. New York: Farrar, Straus and Giroux, 2002.

Galvani, Aloysii. *De viribus electricitatis in motu musculari commentarius* [Commentary on the effect of electricity on muscular motion]. Bologna: Ex Typographia Instituti Scientarium, 1791.

Gandy, Robin. "Human versus Mechanical Intelligence." In *Machines and Thought*, ed. Peter Millican and Andy Clark, 1-125.

García-Diego, José A. *Juanelo Turriano, Charles V's Clockmaker: The Man and His Legend*. Translated by Charles David Ley. Madrid: Editorial Castalia, 1986; Canton, MA: Science History Publications, 1986.

Gates, Bill. "A Robot in Every Home." *Scientific American* 296, no. 1 (Jan 2007): 58-65.

Gaukroger, Stephen. *Descartes: An Intellectual Biography*. Oxford: Oxford University Press, 1995.

Gellius, Aulus. *The Attic Nights of Aulus Gellius*. The Loeb Classical Library. Translated by John Carew Rolfe. London: William Heinemann, 1982.

Georges, Thomas M. *Digital Soul: Intelligent Machines and Human Values*. Boulder, CO: Westview Press, 2003.

Glaser, Horst Albert, and Sabine Rossbach. *The Artificial Human: A Tragical History*. Translated by Joseph Swan. Frankfurt: Peter Lang, 2011.

Godwin, William. *Lives of Necromancers*. Bellmawr, NJ: Bards and Sages, 2008.

Goffman, Erving. *Frame Analysis: An Essay on the Organization of Experience*. Boston: Northeastern University Press, 1986.

———. *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday, 1959.

Goodrick-Clarke, Nicholas. *Paracelsus: Essential Readings*. Berkeley: North Atlantic Books, 1999.

Gorman, James. "Rights Group Is Seeking Status of 'Legal Person' for Captive Chimpanzee." *The New York Times*, December 3 2013. http://www.nytimes.com/2013/12/03/science/rights-group-sues-to-have-chimp-recognized-as-legal-person.html?pagewanted=all&_r=0.

Grant, Edward. "Late Medieval Thought, Copernicus, and the Scientific Revolution." *Journal of the History of Ideas* 23, no. 2 (1962): 197-220.

———. *A Source Book in Medieval Science*. Source Books in the History of the Sciences. Cambridge, MA: Harvard University Press, 1974.

Greenblatt, Stephen. *The Swerve: How the World Became Modern*. New York: W. W. Norton, 2011.

Greenwood, J.G., and B. Woodcroft. *The Pneumatics of Hero of Alexandria: From the Original Greek*. Taylor, Walton and Maberly, 1851.

Grenville, Bruce. *The Uncanny: Experiments in Cyborg Culture*. Vancouver, B.C.: Vancouver Art Gallery: Arsenal Pulp Press, 2001.

Hall, J. Storrs. *Beyond AI: Creating the Conscience of the Machine*. Amherst, NY: Prometheus Books, 2007. Kindle edition.

Haney López, Ian F. "The Social Construction of Race: Some Observations on Illusion, Fabrication, and Choice." *Harvard Civil Rights-Civil Liberties Law Review* 29 (15 January 1994): 1-62.

———. *White By Law: The Legal Construction Of Race*. New York: New York University Press, 1996.

Haq, Syed Nomanul, and Ḥayyān Jābir ibn. *Names, Natures, and Things: The Alchemist Jābir Ibn Ḥayyān and His Kitāb Al-Aḥjār (Book of Stones)*. Boston Studies in the Philosophy of Science Volume 158. Dordrecht: Kluwer Academic Publishers, 1994.

Haraway, Donna J. *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge, 1991.

Hasker, William. *The Emergent Self.* Cornell Studies in the Philosophy of Religion. Ithaca, NY: Cornell University Press, 1999.

Haugeland, John, ed. *Mind Design: Philosophy, Psychology, Artificial Intelligence*. First MIT Press edition. Cambridge, MA: MIT Press, 1981.

———, ed. *Mind Design II: Philosophy, Psychology, Artificial Intelligence*. Revised and enlarged edition. Cambridge, MA: MIT Press, 1997.

Hayles, N. Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago: University of Chicago Press, 1999.

———. "Computing the Human." *Theory Culture & Society* 22, no. 1 (Feb 2005): 131-151.

Hersey, George L. *Falling in Love with Statues: Artificial Humans from Pygmalion to the Present*. Chicago: University of Chicago Press, 2009.

Hill, Donald Routledge. *Studies in Medieval Islamic Technology: From Philo to Al-Jazarī, from Alexandria to Diyār Bakr*. Collected Studies Series. Edited by David A. King. Brookfield, VT: Ashgate, 1998.

Hilton, Julian. "Theatricality and Technology: Pygmalion and the Myth of the Intelligent Machine." In *New Directions in Theatre*, edited by Julian Hilton. New York: St. Martin's Press, 1993.

Hippel, Klemens. "Der Menschlichste Von Uns Allen." *Montage AV* 8 (1999): 77-88.

Hobbes, Thomas. *Leviathan, or, the Matter, Forme, & Power of a Common-Wealth Ecclesiasticall and Civill*. London: Andrew Crooke, 1651.

Hoff, Christina. "Immoral and Moral Uses of Animals." *New England Journal of Medicine* 302, no. 2 (Jan 1980): 115-118.

Hogan, James P. *Mind Matters: Exploring the World of Artificial Intelligence*. New York: Ballantine Publishing Group, 1998.

Horowitz, Alexandra C., and Marc Bekoff. "Naturalizing Anthropomorphism: Behavioral Prompts to Our Humanizing of Animals." *Anthrozoos* 20, no. 1 (Mar 2007): 23-35.

Hughes, James. "Compassionate AI and Selfless Robots: A Buddhist Approach." In *Robot Ethics,* edited by Patrick Lin, et. al., 69-84.

Hultén, Pontus, Anders Oesterlin, and Alicia Legg. *The Machine: As Seen at the End of the Mechanical Age*. New York: Museum of Modern Art, 1968.

Hyde, Walter Woodburn. "Note on Human Automata." *Modern Philology* 11, no. 2 (1913): 267-68.

Hyman, Anthony. *Charles Babbage: Pioneer of the Computer*. Princeton, NJ: Princeton University Press, 1985.

ibn-Shākir, Ahmad ibn-Mūsā, al-Hasan ibn-Mūsā ibn-Shākir, and Muhammad ibn-Mūsā ibn-Shākir (Banū Mūsā). *Kitab Al-Hiyal* [*The Book of Ingenious Devices*]. Translated and annotated by Donald R. Hill. Iraq: 850; Dordrecht: Reidel, 1979.

IBM. "Watson Technology." Accessed 8 April 2014. http://www.ibm.com/smarterplanet/us/en/ibmwatson/tech.html.

Idel, Moshe. *Golem: Jewish Magical and Mystical Traditions on the Artificial Anthropoid*. SUNY Series in Judaica. Albany, NY: State University of New York Press, 1990.

Ihde, Don. *Philosophy of Technology: An Introduction*. Paragon Issues in Philosophy. New York: Paragon House, 1993.

IMDb. "Prometheus 2." Accessed 29 August 2014. http://www.imdb.com/title/tt2316204/?ref_=nv_sr_1.

Ichino, Rin and Takashi Amano. "SoftBank to sell robot in US stores within 12 months." *Bloomberg Online*. Published 2 September 2014. Accessed 12 September 2014. http://www.bloomberg.com/news/2014-09-02/softbank-to-sell-robot-in-u-s-stores-within-12-months.html.

Introna, Lucas D. "The 'Measure of a Man' and the Ethos of Hospitality: Towards an Ethical Dwelling with Technology." *AI & Society* 25, no. 1 (2010): 93-102.

Ishiguro, Hiroshi. "Android Science—Toward a New Cross-Interdisciplinary Framework." *Springer Tracts in Advanced Robotics* 28 (2007): 118-27.

James, Louis. "Frankenstein's Monster in Two Traditions." In *Frankenstein, Creation, and Monstrosity*, ed. Stephen Bann, 77-94..

Japan Times, The. "Pepper Tech Festival 2014,." YouTube video. Accessed 15 October 2014. http://www.youtube.com/watch?v=XcJccQqTM6Q.

Jennings, Anne. *Medieval Gardens*. Published in association with the Museum of Garden History. London: English Heritage, 2004.

Kanda, Takayuki, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. "Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial." *Human-Computer Interaction* 19, no. 1-2 (2004): 61-84.

Kanda, Takayuki, and Hiroshi Ishiguro. "An Approach for a Social Robot to Understand Human Relationships—Friendship Estimation through Interaction with Robots." *Interaction Studies* 7, no. 3 (2006): 369-403.

———. "Friendship Estimation Model for Social Robots to Understand Human Relationships." *RO-MAN 2004 - The 13th IEEE International Symposium on Robot & Human Interactive Communication* (Sep 2004): 540-44.

Kanda, Takayuki, Rumi Sato, Naoki Saiwaki, and Hiroshi Ishiguro. "Friendly Social Robot That Understands Human's Friendly Relationships." Paper presented at the Intelligent robots and systems 2004 IEEE/RSJ international conference on intelligent robots and systems (IROS 2004), Sendai, Japan, Sep 2004.

Kane, Robert. *The Oxford Handbook of Free Will*. Oxford: Oxford University Press, 2002.

Kang, Minsoo. *Sublime Dreams of Living Machines: The Automaton in the European Imagination*. Cambridge, MA: Harvard University Press, 2011.

Keele, Kenneth D. "Leonardo Da Vinci's Influence on Renaissance Anatomy." *Medical History* 8 (Oct 1964): 360-70.

Kendall, Gavin, and Gary Wickham. *Using Foucault's Methods*. Introducing Qualitative Methods. London: Sage Publications, 1999.

King, Elizabeth. "Clockwork Prayer: A Sixteenth-Century Mechanical Monk." *Blackbird* 1, no. 1 (Spring 2002). http://www.blackbird.vcu.edu/v1n1/nonfiction/king_e/toc.htm.

———. "Perpetual Devotion: A Sixteenth-Century Machine That Prays." In *Genesis Redux*, edited by Jessica Riskin, 263-92.

Kleiser, Jeff, and Diana Walczak. "Synthespianism and Anthropomorphization of Computer Graphics." Published 2 October 2002. http://www.kurzweilai.net/synthespianism-anthropomorphization-of-computer-graphics.

Kloc, Joe. "The Living Robot." *Seed Magazine*. Published 26 March, 2009. http://seedmagazine.com/content/article/the_living_robot/.

Kotsanas, Kostas. "Αρχαια Ελληνικη ΤεχνολοΥια: The Automatic Theatres of the Ancient Greeks." Accessed 10 November 2014. http://kotsanas.com/gb/exh.php?exhibit=0101002.

Krimsky, Sheldon, Kathleen Sloan, and Council for Responsible Genetics. *Race and the Genetic Revolution: Science, Myth, and Culture*. New York: Columbia University Press, 2013.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. Third edition. Chicago: University of Chicago Press, 1962; 1996.

Kuiken, Todd. “A Prosthetic Arm That ‘Feels.’” Filmed October 2011 at TEDGlobal 2011. TED video, 18:51. https://www.ted.com/talks/todd_kuiken_a_prosthetic_arm_that_feels.

Kurzweil, Ray. *The Age of Intelligent Machines*. Cambridge, MA: MIT Press, 1990.

———. *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking, 2005.

La Mettrie, Julien Offray de. *Machine Man and Other Writings*. Translated by Ann Thompson. Cambridge: Cambridge University Press, 1996.

Landels, John G. *Engineering in the Ancient World*. Berkeley: University of California Press, 1978.

Lang, Jeffrey. *Immortal Coil*. Star Trek: The Next Generation series. New York: Pocket Books, 2002.

Langton, Christopher G. *Artificial Life: An Overview*. Cambridge, MA: MIT Press, 1997.

Laplace, Pierre Simon. *A Philosophical Essay on Probabilities*. Translated by Frederick Wilson Truscott and Frederick Lincoln Emory. New York: John Wiley & Sons, 1902.

Laurel, Brenda, and S. Joy Mountford. *The Art of Human-Computer Interface Design*. Reading, MA: Addison-Wesley, 1990.

Lee, Kwan Min, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. “Are Physically Embodied Social Agents Better Than Disembodied Social Agents?: The Effects of Physical Embodiment, Tactile Interaction, and People’s Loneliness in Human–Robot Interaction.” *International Journal of Human-Computer Studies* 64, no. 10 (2006): 962-73.

Lee, Kwan Min, Wei Peng, Seung- A. Jin, and Chang Yan. “Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human?Robot Interaction.” *Journal of Communication* 56, no. 4 (2006): 754-72.

Lee, Malrey. “Evolution of Behaviors in Autonomous Robot Using Artificial Neural Network and Genetic Algorithm.” *Information Sciences* 155, no. 1-2 (2003): 43-60.

Leeming, David Adams, and Margaret Adams Leeming. *A Dictionary of Creation Myths*. New York: Oxford University Press, 1995.

Leibniz, Gottfried Wilhelm. *The Monadology and Other Philosophical Writings*. Translated by Robert Latta. London: Oxford University Press, 1951.

Leonardo-History.com. “Leonardo da Vinci’s Letter to Ludovico the Moor.” Accessed 10 December 2013. http://www.leonardo-history.com/life.htm?Section=S5.

Levy, David. “The Ethical Treatment of Artificially Conscious Robots.” *International Journal of Social Robotics* 1, no. 3 (2009): 209-16.

Libin, Alexander V., and Elena V. Libin. “Person-Robot Interactions from the Robopsychologists’ Point of View: The Robotic Psychology and Robotherapy Approach.” *Proceedings of the IEEE* 92, no. 11 (Nov 2004): 1789-803.

Lillegard, Norman. “No Good News for Data.” *Cross Currents* 44, no. 1 (Spring 1994): 28.

Lin, Patrick, Keith Abney, and George A. Bekey, eds. *Robot Ethics: The Ethical and Social Implications of Robotics*. Intelligent Robotics and Autonomous Agents. Cambridge, MA: MIT Press, 2012. Kindle edition.

Loebner, Hugh Gene. “Home Page of the Loebner Prize in Artificial Intelligence: The First Turing Test.” Last updated 1 April 2013. http://www.loebner.net/Prizef/loebner-prize.html.

Loy, John Robert. *Diderot’s Determined Fatalist*. New York: King’s Crown Press, 1950.

Loyall, A Brian. “Believable Agents: Building Interactive Personalities.” PhD Diss., Carnegie Mellon University, 1997. https://www.cs.cmu.edu/afs/cs/project/oz/web/papers/CMU-CS-97-123.pdf.

Luxford, Julian M. *The Art and Architecture of English Benedictine Monasteries, 1300–1540: A Patronage History*. Studies in the History of Medieval Religion. Rochester, NY: Boydell Press, 2005.

MacDougall, Elisabeth Blair, ed. *Medieval Gardens*. Washington, DC: Dumbarton Oaks Research Library and Collection, 1986.

Maisonnier, Bruno. Pepper Tech Festival keynote speech. 20 September 2014. Video, 11:31 On “Community: News: Pepper Tech Festival 2014,” accessed 16 October 2014, https://community.aldebaran.com/en/news/pepper-tech-festival-2014.

Masood, Eshad. *Science and Islam: A History*. London: Icon Books, Ltd., 2006.

Maxwell, Benjamin. “Top 25 Star Trek Characters.” *IGN Movies*. Published electronically 8 May 2009. http://www.ign.com/articles/2009/05/08/top-25-star-trek-characters.

Mayr, Otto. *Authority, Liberty, & Automatic Machinery in Early Modern Europe*. Johns Hopkins Studies in the History of Technology. Baltimore: Johns Hopkins University Press, 1986.

McClintock, A. E. *The Convergence of Machine and Human Nature : A Critique of the Computer Metaphor of Mind and Artificial Intelligence*. Aldershot: Avebury, 1995.

McCourt, Finlay. “An Examination of the Mechanisms of Movement in Heron of Alexandria’s On Automaton-Making.” In *Explorations in the History of Machines and Mechanisms: Proceedings of HMM2012*, edited by Teun Koetsier and Marco Ceccarelli. Dordrecht: Springer, 2012.

McQuay, Michael. *Suspicion*. Isaac Asimov’s Robot City series. New York: Bryon Preiss Visual Publications, 1987.

Milham, Willis Isbister. *Time & Timekeepers, Including the History, Construction, Care, and Accuracy of Clocks and Watches*. New York: The Macmillan Company, 1945.

Miller, Naomi. “Paradise Regained: Medieval Garden Fountains.” In *Medieval Gardens*, Elisabeth Blair MacDougall, ed., 135-154.

Millican, Peter, and Andy Clark, eds. *Machines and Thought: The Legacy of Alan Turing*. 2 vols. New York: Clarendon Press, 1996.

MIT Media Lab. "Personal Robots Group." Accessed 11 May 2014. http://robotic.media.mit.edu/.

Musées De La Ville De Strasbourg. "Museum of Decorative Arts >> Collections ." Accessed 15 October 2013. http://www.musees.strasbourg.eu/index.php?page=collections-Arts-deco-en.

Needham, Joseph, and Ling Wang. *Science and Civilisation in China*. 4 vols. Cambridge, UK: University Press, 1954.

Newman, Lex. "Unmasking Descartes's Case for the Bête Machine Doctrine." *Canadian Journal of Philosophy* 31, no. 3 (Sept 2001): 389-426.

Nocks, Lisa. *The Robot: The Life Story of a Technology*. Baltimore, MD: Johns Hopkins University Press, 2008.

Noesis. "Thessaloniki Science Center and Technology Museum: The Technology Museum." Accessed 15 May 2014. http://www.noesis.edu.gr/en/the-technology-museum/.

Nolfi, Stefano. "Emergence of Communication in Embodied Agents: Co-Adapting Communicative and Non-Communicative Behaviours." *Connection Science* 17, no. 3-4 (2005): 231-48.

Nomura, Tatsuya, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. "Measurement of Negative Attitudes toward Robots." *Interaction Studies* 7, no. 3 (2006): 437-54.

Noverre, Jean Georges. *The Works of Monsieur Noverre, Translated from the French*. 3 vols. London, 1782; New York: AMS Press, 1978.

Ogden, Bernard, and Kersten Dautenhahn. "Embedding Robotic Agents in the Social Environment." *Towards Intelligent Mobile Robots TIMR 01: The*, Vol 3 (2001).

Okuno, Hiroshi G., Katsuhiro Nakadai, and Hiroaki Kitano. "Design and Implementation of Personality of Humanoids in Human Humanoid Non-Verbal Interaction." IEA/AIE'2003 Proceedings of the 16th international conference on Developments in applied artificial intelligence (2003): 662-73.

———. "Social Interaction of Humanoid Robot Based on Audio-Visual Tracking." In Tim Hendtlass and Ali Moonis, eds., *Developments in Applied Artificial Intelligence* (Berlin: Springer, 2002): 725-735.

O'Malley, Charles D. *Leonardo on the Human Body: The Anatomical, Physiological, and Embryological Drawings of Leonardo Da Vinci.* New York: H Schuman, 1952; Mineola, NY: Dover, 1983.

Park, Robert Ezra. *Race and Culture*. His Collected Papers, Vol 1. Glencoe, IL: Free Press, 1950.

Perkins, Christopher, Owen K.C. Stephens, and Rodney Thompson. *Star Wars Roleplaying Game*. Saga edition. Fenton, WA: Wizards of the Coast, 2007.

Perkowitz, Sidney *Digital People: From Bionic Humans to Androids*. Washington, DC: Joseph Henry Press, 2004.

Pfeifer, Rolf, Josh Bongard, and Simon Grand. *How the Body Shapes the Way We Think: A New View of Intelligence*. Cambridge, MS: MIT Press, 2007.

Piccolino, Marco. "Animal Electricity and the Birth of Electrophysiology: The Legacy of Luigi Galvani." *Brain Research Bulletin* 46, no. 5 (July 1998): 381-407.

Piccolino, Marco, and Marco Bresadola. *Shocking Frogs: Galvani, Volta, and the Electric Origins of Neuroscience*. London: Oxford University Press, 2013.

Price, Derek J. de Solla. "Automata and the Origins of Mechanism and Mechanistic Philosophy." *Technology and Culture* 5, no. 1 (1964): 9-23.

Prioreschi, Pilino. "Determinants of the Revival of Dissection of the Human Body in the Middle Ages." *Medical Hypotheses* 56, no. 2 (Feb 2001): 229-34.

———. *A History of Medicine, Volume III: Roman Medicine*. 2nd ed. Omaha: Horatius Press, 1996.

Rashkin, Esther. "Data Learns to Dance: Star Trek and the Quest to Be Human." *American Imago: Psychoanalysis and the Human Sciences* 68, no. 2 (2011): 321-46.

Regele, Ralf, Paul Levi, and Wolfgang Bott. "Prorobot—Predicting the Future of Humanoid Robots." In *Proceedings of Robocup 2003: Robot Soccer World Cup VII* (2003): 366-73.

Reilly, Kara. *Automata and Mimesis on the Stage of Theatre History*. New York: Palgrave Macmillan, 2011.

Reilly, W Scott Neil. "Believable Social and Emotional Agents." PhD Diss., Carnegie Mellon University, 1996.

Rhee, Jennifer. "Beyond the Uncanny Valley: Masahiro Mori and Philip K. Dick's Do Androids Dream of Electric Sheep?" *Configurations: A Journal of Literature, Science, and Technology* 21, no. 3 (2013): 301-29.

Richardson, K. "Mechanical People." *New Scientist* 190, no. 2557 (2006): 56-57.

Rieder, John. "Karl Čapek." In *Fifty Key Figures in Science Fiction*, edited by Mark Bould, et. al., 47-51.

Riefstahl, Rudolf M. "The Date and Provenance of the Automata Miniatures." *The Art Bulletin* 11, no. 2 (1929): 206-15.

Riskin, Jessica. "Eighteenth-Century Wetware." *Representations* 83, no. 1 (2003): 97-125.

———, ed. *Genesis Redux: Essays in the History and Philosophy of Artificial Life*. Chicago: University of Chicago Press, 2007.

Roach, Joseph R. *The Player's Passion: Studies in the Science of Acting*. Ann Arbor: University of Michigan Press, 1993.

Roberts, Adam. *The History of Science Fiction*. Palgrave Histories of Literature. New York: Palgrave Macmillan, 2006.

———. *Science Fiction*. The New Critical Idiom. Second edition. London: Routledge, 2006.

Robles, Dawn M. "Road to the Future: Robot Sentience and Presence in Film." In Will Wright and Steven Kaplan, eds., *The Image of the Road in Literature, Media, and Society* (Pueblo, CO: Colorado State University, 2005): 68-70.

Robotguide, Mike the. "NAO and ROMEO robots at Innorobo 2014." *Robot Guide*. 28 March 2014. Accessed 16 October 2014. http://www.robotguide.com/2014/03/28/aldebaran-robotics-news/nao-romeo-robots-innorobo-2014/.

Roddenberry, Gene. "Star Trek: The Next Generation Writer/Director's Guide." 1987.

Rosci, Marc. *The Hidden Leonardo*. Translated by John Gilbert. Chicago: Rand McNally, 1976.

Rosenfield, Leonora Cohen. *From Beast-Machine to Man-Machine; Animal Soul in French Letters from Descartes to La Mettrie*. New York: Oxford University Press, 1941.

Rosenthal-von der Pütten, Astrid, Nicole Krämer, Christian Becker-Asano, Kohei Ogawa, Shuichi Nishio, and Hiroshi Ishiguro. "The Uncanny in the Wild. Analysis of Unscripted Human-Android Interaction in the Field." *International Journal of Social Robotics* 6, no. 1 (Jan 2014): 67.

Rosheim, Mark E. *Leonardo's Lost Robots*. Berlin: Springer, 2006.

———. *Robot Evolution: The Development of Anthrobotics* New York: John Wiley and Sons, 1994.

Rothman, Lily. "Prometheus star Michael Fassbender on his robot role and why he believes in aliens." *Time*. Published 6 June 2012. Accessed 28 August 2014. http://entertainment.time.com/2012/06/06/prometheus-star-michael-fassbender-on-his-robotic-role-ridley-scott-and-why-he-believes-in-aliens/.

Rousseau, Jean-Jacques. *Pygmalion, un Scene Lyrique* [*Pygmalion: A Lyrical Scene*]. Bibliothèque de Neuchâtel, 1762. Translated 1983 by Paul De Man, from the collection "The Portable Rousseau" at University of California at Irvine, digitized April 2010. http://ucispace.lib.uci.edu/bitstream/handle/10575/1094/08demanportablerousseaupygmalion.pdf?sequence=11.

Rubenstein, Richard E. *Aristotle's Children: How Christians, Muslims, and Jews Rediscovered Ancient Wisdom and Illuminated the Dark Ages*. Orlando, FL: Harcourt, 2003.

Ryle, Gilbert. *The Concept of Mind*. Chicago: University of Chicago Press, 1949. Kindle edition.

Scassellati, Brian. "Theory of Mind for a Humanoid Robot." *Autonomous Robots* 12, no. 1 (Jan 2002): 13-24.

Schaffer, Simon. "Enlightened Automata." In Clark, et. al, eds., *The Sciences in Enlightened Europe*, 126-65.

Scheutz, Matthias, Paul Schermerhorn, James Kramer, and David Anderson. "First Steps toward Natural Human-Like Hri." *Autonomous Robots* 22, no. 4 (2006): 411-23.

Schmidt, Laura Lee. "Islamic Automata in the Absence of Wonder." Master's Thesis, Massachusetts Institute of Technology, 2010.

Schmidt, Wilhelm. *Heronis alexandrini opera quae supersunt omnia*. Leipzig: B.G. Teubneri, 1899.

Science Museum (UK). "Babbage's Analytical Engine, 1834–1871 (Trial model)." Accessed 20 March 2014. http://www.sciencemuseum.org.uk/objects/computing_and_data_processing/1878-3.aspx?keywords=babbage+analytical+engine.

Searle, John R. "Minds, Brains, and Programs." In *Mind Design II*, John Haugeland, ed., 183-204.

Segel, Harold B. *Pinocchio's Progeny: Puppets, Marionettes, Automatons and Robots in Modernist and Avant-Garde Drama*. PAJ Books. Baltimore: Johns Hopkins University Press, 1995.

Sekiyama, Kosuke, and Toshio Fukuda. "Toward Social Robotics." *Applied Artificial Intelligence* 13, no. 3 (1999): 213-38.

Seltzer, Mark. *Bodies and Machines*. New York: Routledge, 1992.

Sennett, John F. "Requiem for an Android? Lillegard Traps Us in a False Dilemma." *CROSS CURRENTS - NEW YORK*- 46, no. 2 (1996): 195-214.

Sharkey, Noel. "I, Ropebot." *New Scientist* 194, no. 2611 (2007): 32-35.

———. "My, Robot: The Rise, Fall, and Resurrection of a Celebrity Robot Called Elektro." *NEW SCIENTIST*, no. 2687/2688 (2008): 46-47.

Shelley, Mary Wollstonecraft. *Frankenstein; or, the Modern Prometheus*. 3 vols London: Lackington, Hughes, Harding, Mavor, & Jones, 1818. Reprint e-artnow, 2013. Kindle edition.

———. *Frankenstein; or, the Modern Prometheus*. Revised Classic Books edition (1831). Reprint Amazon Digital Services, Inc., 2014. Kindle edition.

Sherwood, Merriam. "Magic and Mechanics in Medieval Fiction." *Studies in Philology* 44, no. 4 (1947): 567-92.

Shieber, Stuart M., ed. *The Turing Test: Verbal Behavior as the Hallmark of Intelligence*. Cambridge, MA: MIT Press, 2004.

Short, Sue. "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): 209-23.

Sinčák, Peter, Ján Vaščák, and Kauro Hirota, eds. *Machine Intelligence: Quo Vadis?* Advances in Fuzzy Systems: Applications and Theory, vol 21. River Edge, NJ: World Scientific, 2004.

Skinner, B. F. *Verbal Behavior*. The Century Psychology Series. New York: Appleton-Century-Crofts, 1957.

Son, Masayoshi. "SoftBank Next 30-Year Vision." Presentation to SoftBank corporate shareholders, 25 June 2010. Video and materials available at "SoftBank: About Us," http://www.softbank.jp/en/corp/about/philosophy/next30/.

Spargo, John Webster. *Virgil the Necromancer: Studies in Virgilian Legends*. Harvard Studies in Comparative Literature.: Vol. X. Cambridge: Harvard University Press, 1934.

Spence, Joseph. *Letters from the Grand Tour*. Montreal: McGill-Queen's University Press, 1975.

Standage, Tom. *The Turk: The Life and Times of the Famous Eighteenth-Century Chess-Playing Machine*. New York: Walker & Co., 2002.

Stanford, Michael. "Traveling the Uncanny Valley." *Internet Telephony* 16, no. 9 (2013): 18-18.

Sussman, Mark. "Performing the Intelligent Machine: Deception and Enchantment in the Life of the Automaton Chess Player." *TDR* 43, no. 3 (1999): 81-96.

Telotte, J. P. *Replications: A Robotic History of the Science Fiction Film*. Urbana: University of Illinois Press, 1995.

Tibbetts, G. R. "Review: The Book of Knowledge of Ingenious Devices." *Bulletin of the School of Oriental and African Studies, University of London* 38, no. 1 (1975): 151-53.

Torrance, Steve. "Ethics and Consciousness in Artificial Agents." *AI and Society* 22, no. 4 (Apr 2008): 495-521.

Tresch, John. "The Fanatastic Automaton and Its Uncanny Kin in France, 1815–1848." Paper presented at the Annenberg Colloquium in European History, University of Pennsylvania, Spring 2007.

Truitt, Elly Rachel. "From Magic to Mechanism: Medieval Automata, 1100–1550." PhD Diss., Harvard University, 2007.

Truitt, Elly R. "The Garden of Earthly Delights: Mahaut of Artois and the Automata." *Medieval Feminist Forum* 46, no. 1 (2010): 74-79.

Twentieth Century Fox. "Project Prometheus." Last accessed 14 November 2014. http://www.projectprometheus.com/.

———. "The Weyland Corporation." Last accessed 14 November 2014. http://www.weylandindustries.com/.

Ulam, Stanislaw. "Tribute to John von Neumann." *Bulletin of the American Mathematical Society* 64, no. 3 (May 1958): 1-49.

United Nations. "Total Population, Both Sexes." Excel file from "World Population Prospects: The 2012 Revision." Accessed 22 August 2014. http://esa.un.org/unpd/wpp/Excel-Data/population.htm.

University of Reading. "Turing Test Success Marks Milestone in Computing History." Press release dated 8 June 2014. Accessed 31 October 2014. http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx.

Usher, Abbott Payson. *A History of Mechanical Inventions*. Revised edition. New York: Dover, 1988.

van Buren, Ann Hagopian. “Reality and Literary Romance in the Park of Hesdin.” In *Medieval Gardens*, Elisabeth Blair MacDougall, ed., 115-134.

Vasari, Giorgio. *Lives of the Most Eminent Painters, Sculptors, and Architects*. Translated by Mrs. Jonathan Foster. Vol. 2, London: Bell & Daldy, 1871.

Jacques de Vaucanson. *Le mécanisme du fluteur automate: présenté a Messieurs de l'académie Royale des sciences*. Paris: Chez Jacques Guerin, 1738.

Vaughan, Richard. *Philip the Good; the Apogee of Burgundy*. New York: Barnes & Noble, 1970.

Vaux, Bernard Carra de. “Le Livre Des Appareils Pneumatiques Et Des Machines Hydrauliques De Philon De Byzance D’après Les Versions Arabes D’oxford Et De Constantinople.” In *Notices Et Extraits Des Manuscrits De La Bibliothèque Nationale, Tome XXXVIII* Paris: Academie des inscriptions et des belles lettres, 1903.

Ville de Neuchâtel. “Vivre La Ville! Inhabitants: Automata.” Accessed 12 February 2014. http://en.neuchatelville.ch/profils/residents.asp/1-11-160-21345-10001-1001-1-1-2-1/2-0-1345-10001-1000-2-0/.

Ville de Neuchâtel musée d’art et d’histoire. “Les Automates Jaquet-Droz.” Accessed 12 January 2014. http://www.mahn.ch/collections-arts-appliques-automates.

Vize, Brendan. “Do Androids Dream of Electric Shocks? Utilitarian Machine Ethics.” Masters Thesis, Victoria University of Wellington, 2011.

Voskuhl, Adelheid. *Androids in the Enlightenment: Mechanics, Artisans, and Cultures of the Self*. Chicago: University of Chicago Press, 2013. Kindle edition.

Weiss, Allen S. *Mirrors of Infinity: The French Formal Garden and Seventeenth-Century Metaphysics*. New York: Princeton Architectural Press, 1995.

Weizenbaum, Joseph. “ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine.” *Communications of the ACM* 9, no. 1 (1966): 36-45.

White, Lynn Townsend. *Medieval Technology and Social Change*. Oxford: Clarendon Press, 1962.

Wiener, Norbert. *Cybernetics: Or the Science of Control and Communication in the Animal and the Machine*. New York: J. Wiley, 1948.

Wikia, Inc., “Wookieepedia, the *Star Wars* Wiki.” Last accessed 15 November, 2014. http://starwars.wikia.com/wiki/Main_Page.

Wilcox, Rhonda V. “Dating Data: Miscegenation in Star Trek: The Next Generation.” *Extrapolation: A Journal of Science Fiction and Fantasy* 34, no. 3 (1993): 265-77.

Wildman, Wesley J. “But Consciousness Isn’t Everything.” *Cross Currents* 46, no. 2 (Summer 1996): 215-220.

Williamson, Jack. *The Humanoids*. New York: Lancer, 1949. Kindle edition.

———. *With Folded Hands*. *Astounding Science Fiction*, June 1947; Amazon Digital Services, Inc., 2011. Kindle edition.

———. *Wonder's Child: My Life in Science Fiction*. Dallas,TX: BenBella Books, 2005.

Wilson, Luke. "William Harvey's 'Prelectiones:' The Performance of the Body in the Renaissance Theatre of Anatomy." *Representations*, no. 17 (1987): 62-95.

Wise, Steven M. *Drawing the Line: Science and the Case for Animal Rights* Cambridge, MA: Perseus Books, 2002.

———. "Nonhuman Rights to Personhood." *Pace Environmental Law Review* 30, no. 3 (2013): 1278-1290.

Wood, Gaby. *Edison's Eve: A Magical History of the Quest for Mechanical Life*. New York: Anchor Books, 2002.

Złotowski, Jakub, Diane Proudfoot, and Christoph Bartneck. "More Human Than Human: Does The Uncanny Curve Really Matter?" *Proceedings of the HRI2013 Workshop on Design of Humanlikeness in HRI* (2013): 7-13.

Zondy, David. "Tales of Future Past: ELEKTRO." Accessed 18 October 2014. http://davidszondy.com/future/robot/elektro1.htm.

FILMOGRAPHY

Alien (1979). Directed by Ridley Scott. Director's cut with commentary and extra features. Los Angeles: Twentieth Century Fox, 2012. iTunes download.

Aliens (1986). Directed by James Cameron. Special edition with commentary and extra features. Los Angeles: Twentieth Century Fox, 2013. iTunes download.

Alien[3] (1992). Directed by David Fincher. Special edition with commentary and extra features. Los Angeles: Twentieth Century Fox, 2010. iTunes download.

Alien: Resurrection (1997). Directed by Jean-Pierre Jeunet. Special edition with commentary and extra features. Los Angeles: Twentieth Century Fox, 2010. iTunes download.

Ancient Discoveries. TV Series (History Channel, 2003-2009). Developed by Wild Dream Films. Season 2, episode 10 "Robots." Originally aired 17 April 2007. YouTube Video. Accessed 1 November 2014. http://www.youtube.com/watch?v=ykbBayIxx8c.

Battlestar Galactica. Original TV series (American Broadcasting Company, 1978-1979). Created by Glen A. Larson. Los Angeles: Universal Cities Studios, 2008. iTunes download.

Battlestar Galactica. TV miniseries (Sci-Fi Channel, 2003). Developed by Ronald D. Moore and David Eick. Los Angeles: Universal Network Television, 2012. iTunes download.

Battlestar Galactica. TV series (Sci-Fi Channel, 2004-2009). Developed by Ronald D. Moore and David Eick. Los Angeles: Universal Network Television, 2012. iTunes download.

Blade Runner (1982) Directed by Ridley Scott. Director's cut with commentary. Burbank, CA: Warner Brothers, 2008. iTunes download.

Bicentennial Man (1999). Directed by Chris Columbus. Burbank, CA: Touchstone Home Video, 2000. DVD.

Caprica. TV series (SyFy Channel, 2009-2010). Created by Remi Aubuchon and Ronald D. Moore. Los Angeles: Universal Network Television, 2010. iTunes download.

Forbidden Planet (1956). Directed by Fred MacLeod Wilcox. Burbank, CA: Warner Brothers, 1999. iTunes download.

Der Golem (1915). Directed by Paul Wegener and Henrik Galeen. Berlin: Deutsche Bioskop GmbH, 1915.

Her (2013). Directed by Spike Jonze. Burbank, CA: Warner Brothers, 2014. iTunes download.

I, Robot (2004). Directed by Alex Proyas. Los Angeles: Twentieth Century Fox, 2004. iTunes download.

Lawrence of Arabia (1962). Directed by David Lean. Fiftieth anniversary edition. Los Angeles: Columbia Pictures, 2012. iTunes download.

Mary Shelly's Frankenstein (1994). Directed by Kenneth Branagh. Culver City, CA: Sony Pictures, 1998. iTunes download.

Metropolis (1927). Directed by Fritz Lang. 2002 restored edition. New York: Kino Video, 2002. DVD.

Prometheus (2012). Directed by Ridley Scott. Special edition with commentary and extras. Los Angeles: Twentieth Century Fox, 2010. iTunes download.

Spaceballs (1987). Directed by Mel Brooks. Beverly Hills, CA: MGM Home Entertainment, 2000. DVD.

Short Circuit (1986). Directed by John Badham. Culver City, CA: TriStar Pictures, 2000. DVD.

S1M0NE (2002). Directed by Andrew Niccol. Los Angeles, CA: New Line Home Entertainment, 2003. DVD.

Star Trek: The Next Generation. TV series (1987-1994). Created by Gene Roddenberry. Special twentieth anniversary edition "The Complete Series" with commentary and extras. Hollywood, CA: Paramount Home Entertainment, 2007. DVD.

Star Trek: Voyager. TV series (United Paramount Network, 1995-2001). Created by Rick Berman, Michael Piller, and Jeri Tayler. Santa Monica, CA: CBS Corporation, 2007. iTunes download.

Star Trek VII: Generations (1994). Directed by David Carson. Hollywood, CA: Paramount Home Entertainment, 2005. iTunes download.

Star Trek VIII: First Contact (1996). Directed by Jonathan Frakes. Hollywood, CA: Paramount Home Entertainment, 2005. iTunes download.

Star Trek IX: Insurrection (1998). Directed by Jonathan Frakes. Hollywood, CA: Paramount Home Entertainment, 2005. iTunes download.

Star Trek X: Nemesis (2002). Directed by Stuart Baird. Hollywood, CA: Paramount Home Entertainment, 2005. iTunes download.

Star Wars: Droids: The Adventures of R2-D2 and C-3P0. Animated TV series (ABC Television Network, 1985-1986). Created by George Lucas. San Francisco: Lucasfilm Ltd., 2004. DVD.

Star Wars: Episode I – The Phantom Menace (1999). Directed by George Lucas. Special edition with commentary and extras. San Francisco: Lucasfilm Ltd., 2001. DVD.

Star Wars: Episode II – Attack of the Clones (2002) Directed by George Lucas. Special edition with commentary and extras. San Francisco: Lucasfilm Ltd., 2002. DVD.

Star Wars: Episode III – Revenge of the Sith (2005). Directed by George Lucas. Special edition with commentary and extras. San Francisco: Lucasfilm Ltd., 2005. DVD.

Star Wars: Episode IV – A New Hope (1977). Directed by George Lucas. Special limited edition release with commentary and extras. San Francisco: Lucasfilm Ltd., 2006. DVD.

Star Wars: Episode V – The Empire Strikes Back (1980). Directed by Irvin Kershner. Special limited edition release with commentary and extras. San Francisco: Lucasfilm Ltd., 2006. DVD.

Star Wars: Episode VI – Return of the Jedi (1983). Directed by Richard Marquand. Special limited edition release with commentary and extras. San Francisco: Lucasfilm Ltd., 2006. DVD.

The Stepford Wives (1975). Directed by Bryan Forbes. Los Angeles, CA: Columbia Pictures, 2004. DVD.

Wall-E (2008). Directed by Andrew Stanton. Burbank, CA: Walt Disney Studios Motion Pictures, 2008. iTunes download.

The Wizard of Oz (1939). Directed by Victor Fleming. Sixtieth anniversary special edition. Burbank, CA: Warner Brothers, 1999. DVD.