Original Article-Open Access

International Review of Literary Studies

ISSN: (P) 2709-7013, (E) 2709-7021 Vol. 6, Issue 1 January-June 2024, pp. 1-12

"The Human Machine Was Horribly Imperfect": Staging the Ethics of Human-AI Relationships

Jarrod DePrado[1]

Abstract

In 1920, Karel Čapek wrote R.U.R. (Rossum's Universal Robots), a work that fundamentally altered our understanding of the relationship between humans and artificial intelligence. Today, this relationship is a source of ongoing ethical dilemmas, as evidenced by three contemporary plays: Thomas Gibbons's Uncanny Valley (2015), Jordan Harrison's Marjorie Prime (2014), and Jennifer Haley's The Nether (2013). These plays, whether through humanoid robots, holographic manifestations, or digital avatars, explore humanity's inherent existential concerns when relying on technology, either in the creation or implementation processes or the repercussions afterward. In their attempts to supplement or replace humans with artificial intelligence or the real world with digital spaces, the human characters reveal an innate human struggle to compensate for physical and emotional limitations, often with mixed success. As we examine these plays through the lens of Isaac Asimov's three laws of robotics, we see a clear message: because AI is used in different ways in the contemporary world, Asimov's laws must be updated to reflect those changes. Robots serve humans who, when creating technological surrogates, find themselves accountable to the laws in ways that operate in an ironic mix of the human drive for self-preservation and the technological inevitability of self-destruction through irrelevance. Ultimately, as much as humans try to replace themselves with technology, the process merely underscores the superficiality and fabricated emotional connections at the heart of those relationships.

Keywords: *Marjorie Prime*; *Uncanny Valley*; *The Nether*; *Rossum's Universal Robots*; Isaac Asimov; Artificial intelligence ethics

[1] Sacred Heart University, United States
*Email: depradoj@sacredheart.edu

Publication Details:
Article Received: January 13, 2024
Article Revised: May 14, 2024
Article Published: May 22, 2024

Recommended citation:

DePrado, J. (2024). "The Human Machine Was Horribly Imperfect": Staging the Ethics of Human-AI Relationships *International Review of Literary Studies*, 6(1): pp. 1-12. https://irlsjournal.com/index.php/Irls

Published by Licensee MARS Publishers. Copyright: © the author(s). This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license. (https:// creativecommons.org/licenses/by/ 4.0/).

Karel Čapek's 1920 play, *R.U.R. (Rossum's Universal Robots)*, redefined the relationship between humans and artificial intelligence (AI). Since then, and in light of the recent spike in concerns over what AI means for the future of humanity's relevance, contemporary dramatic works exploring the dynamics of replacing humans with robots or avatars in digital space are especially prescient. In several plays from the last decade, playwrights seek to reconcile the impressive abilities of AI with the ethical dilemmas of using it. Thomas Gibbons's *Uncanny Valley* (2015) and Jordan Harrison's *Marjorie Prime* (2014) raise questions about the limitations of robots replacing humans and the dividing line in what constitutes humanity. Additionally, Jennifer Haley's *The Nether* (2013) anticipates the serious dangers of virtual reality combined with a lack of ethical consideration, asking if anonymity alleviates one's complicity and accountability. These plays speak to our unique moment, which hinges on existential doubts and moral anxieties concerning the future of humans in a digitally tangential society. They consider what is gained and what is lost by reliance (or dependence) on robots, AI, and virtual reality (VR), respectively, to compensate for physical and emotional abilities. However, these works all reaffirm the inherent irony of humanity's desperate search for human connection through non-human media or inhumane means. Ultimately, the human characters realize that their tools fail at genuine human connection because of the elusive, irreproducible essence of humanity.

Čapek's *R.U.R.* pioneered the optimistic possibilities and existential dangers posed by the creation of sentient beings. Domin, the director of Rossum's Universal Robots, sees their purpose as liberating for the human race; replacing workers with robots would bring about a production surplus that would drive costs down[1] and usher in a new utopian existence for humanity where "people will do only what they enjoy. They will live only to perfect themselves" (Čapek, 2004, p. 21).[2] As the engineer Fabry recounts, in their current state, "the human machine . . . was horribly imperfect.[3] It needed to be done away with once and for all" (p. 17). However, the creative minds behind R.U.R. acknowledge that robots work better when they are more human-like in some regards, including the ability to feel pain (p. 19). Ten years later, things have not gone as planned: workers revolted, robots were appropriated for military purposes, and humans stopped having children (pp. 30-32).[4] As robots are placed at odds with humans, Radius (a robot gifted with additional intelligence) recognizes that knowledge leads to a desire to not only be free of enslavement but to wield power over others. As the robots become more human-like, organizing unions and banning together against humans, the response from their creators is to make them even more human: proposing to change them from "universal" to "national" and giving them all "different colors, a different nationality, a different tongue [...] So that any given Robot, to the day of its death [...] will forever hate a Robot bearing the trademark of another factory" (p. 46). What begins with benevolent, ambitious humanitarianism moves towards warring with the robots and endowing them with the worst of humanity in the guise

[1] These potential benefits of integrating robots into the workforce were speculated by Ayres and Miller in 1982 (p. 6). Over forty years later, they have yet to materialize.

[2] Though Czech, Čapek's Domin verbalizes the same dream for subsequent generations as future American President John Adams wrote in 1780: "I must study politics and war, that my sons may have liberty to study mathematics and philosophy. My sons ought to study mathematics and philosophy, geography, natural history and naval architecture, navigation, commerce, and agriculture, in order to give their children a right to study painting, poetry, music, architecture, statuary, tapestry, and porcelain" (Shuffelton, 2004, p. 378). By contrast, Charles T. Rubin (2011) argues that Domin wants to "[change] the human condition in such a way as to allow people to be irresponsible" (p. 76). As with the production surplus from integrating robots into the workforce, the utopian benefits for future descendants have not been achieved in the one hundred years since *R.U.R.*'s publication.

[3] Alice Rayner (1994) argues that many of the science fiction works depicting humans' attempts to find perfection through robotic means ultimately conclude that humanity is "defined by both its distance from perfection and by the way that differences within the human community create a demand for communication, negotiation, and recognition of the 'otherness' of the other" (p. 129).

[4] Christopher Anderson (2014) draws a parallel between "the vital link between toil and childbirth" implied by the usage of the word "labor" to show a correlation between man's ability to stop working with women's inability to bear children (p. 232). While he draws a common conclusion about the ending of the play, which takes on Eden-like resonances by invoking lines from Genesis (p. 242), this could be further extended to consider that the creation of the robots restores humans to a state of grace before Genesis's fall of man: they are alleviated from manual labor and the pain of childbirth, the two punishments explicitly given to humans before expulsion from the Garden of Eden.

of a manufactured Tower of Babel. However, the innate human drive for self-preservation through reproduction perseveres, humbling Helena and then the robots after they destroy humanity and the secrets to the robots' continued existence.

The existential dangers Čapek depicted inspired Isaac Asimov's three laws of robotics to safeguard against the elimination of humans. Placing limitations on robotic abilities reaffirmed their inherent subservience to humanity and assigned them the responsibility of preventing harm to their creators. This paper argues that *R.U.R.* also provides a trajectory of robot integration into society in three stages: creation, implementation, and repercussions. Looking at each play as portraying one of those three stages allows for a discussion of each law individually and reveals a new dynamic between humans and robots. Ultimately, humans pose a larger existential threat to their own survival because of their drive to outsource intimacy, which recontextualizes the purpose of robots: trying to make them more human reaffirms the desire for genuine humanity but also places them in the position to cause more emotional and psychological damage, not only physical harm. The conclusion drawn here is that because AI is used differently than Čapek and Asimov anticipated, the latter's laws must be updated to reflect new, contemporary concerns regarding protection.

Creation: *Uncanny Valley*

"One, a robot may not injure a human being, or, through inaction, allow a human being to come to harm." —Isaac Asimov (1950, p. 51)

Set in "the not-distant future" (Gibbons, 2015, p. 6),[5] *Uncanny Valley* depicts the creation process of a humanoid robot named Julian. Claire, a scientist in her seventies, spends the early scenes teaching Julian the finer points of human movements and vocal intonation so that he can effectively take the place of Julian Barber, a seventy-six-year-old wealthy industrialist with terminal pancreatic cancer. She is looking to bridge the titular "uncanny valley," the psychological unease experienced by humans when they encounter human-like robots whose similarity to humans can cause "fear and disbelief," especially when coupled with discrepancies that reveal their robotic nature (Marynowsky, 2012, p. 482). As Hui Jiang et al. (2022) conclude, one way to cross the uncanny valley is to pursue "high realism (making [a robot] exactly like a human)" (404).[6] Through an involved, two-year process—including "hundreds of hours" of interviews, digitizing "tens of thousands of photographs and documents," both personal and professional, and "scann[ing] and "mapp[ing] every centimeter of his body"—Barber has essentially been re-created in Julian (p. 31). With a "new body, identical in every respect" and with a "biological matrix seeded with Mr. Barber's DNA," Julian will boast a "lifespan of at least two hundred years" (p. 31).[7] Claire and her team have achieved, in her words, the "eternal human dream": "immortality" (p. 31). Julian will be a thirty-four-year-old version of Barber "in perfect health" and will remain that way now that they have the tools to "[liberate his] consciousness from its expiration date" (p. 31). However, despite the $240 million price tag for the procedure (p. 46),[8] the process does not truly replicate the human experience: Julian does not need sleep or food and cannot feel physical sensations, including sensual pleasure (p. 39). Nor is he simply a replication of Barber. The time spent learning from Claire leaves memories combined with those uploaded from Barber; it is more akin to taking over someone else's identity than simply putting

[5] Claire says that her daughter graduated college and had entered a graduate program by the fall of 2042 (p. 25). If her daughter is thirty-five now, and they have not spoken in twelve years (p. 48), that sets the play in the mid-2050s.
[6] As Jiang et al. (2022) explain when examining animated films: the realism of a character's appearance correlates with human comfort and empathy for them. However, when a character approaches but does not entirely achieve a lifelike image, empathy plummets. As examples, they discuss the "Pixar peak" of computer-generated characters that garner "the same level of empathy" as human characters because they combine realistic graphics with "appropriate abstraction" that "[achieve] a higher level of believability." Contrast this with those that look to "[portray] characters based on reality," such as *The Polar Express* (2004), which seem to emulate realism but fail at maintaining emotional engagement, driving the audience into the uncanny valley (p. 404).
[7] As opposed to *R.U.R.*'s robots, who can only live for about twenty years.
[8] In stark contrast to the affordable robots in *R.U.R.*, thanks to innovative Henry Ford-inspired mass production.

someone's consciousness in a new body. As explained in the play, "Julian A [the robot] plus Julian B [Barber's memories] equals Julian C," far from a perfect replication (p. 36).

In and out of the world of the play, this raises moral issues, not least of all: Is Julian *actually* Barber now? Barber's son, Paul—at forty-four, now ten years older than Julian—does not believe so and actively works to have his 'father' removed from control of his company, labeling him a "monster" (pp. 41-2). Paul (like the play itself) raises legal concerns over whether a robot can be considered a person and trusted with power. As Julian himself acknowledges, they are positing uncharted legal questions concerning his rights to own property, vote, or marry (p. 42). However, unlike other works of science fiction (such as *Alien* (1979), *Blade Runner* (1982), or *Westworld* (2016-2022)), there is no attempt to hide that Julian is a robot from either the humans or Julian himself; no Turing Test is required. Asimov's first law of robotics prohibits robots from harming humans or, more relevant for Julian, allowing humans to be harmed through inaction (1950, p. 51). While nothing can be done to save Barber, Julian's existential purpose is to prevent a human from succumbing to death by becoming him (or a dramatically improved version). Regardless of whether or not Julian *becomes* Barber, it is unclear how much agency Julian has outside of fulfilling his assigned purpose. This inherent drive to alleviate harm for his human counterparts is seen especially in his relationship with Claire. In addition to her being his teacher and pseudo-mother during his 'infancy,' she connects emotionally with him while providing details of her life as examples to answer Julian's questions about the world. For example, Claire shares that she has a strained relationship with a daughter (Becky) who refuses to speak to her parents because of their involvement with the creation of artificial intelligence.

Claire does not know where Becky is, if she is married, or has children; Claire and her husband Howard have since reached an "unspoken agreement" to "not discuss it" (p. 25). But she *does* discuss it with Julian, despite it being unprofessional and against protocol, turning it into a lesson on marriage for Julian. From the limited information Claire provides, it is clear that she has essentially traded her biological daughter—who is thirty-five, nearly the same age as Julian—for synthetic children, including the three that precede Julian. In teaching him, she informs Julian that he is experiencing "how children learn about the world" before offering an example from Becky's childhood, returning to a taboo topic that fuses her maternal experiences with her scientific ones (p. 27). Yet, in a moment of frustration, she says that Julian's relationship with Paul is not the same as hers with Becky. Claire goes on to share that Howard's "mind is going" and that his (and now her) retirement is because he cannot be left alone anymore (p. 38). Claire's dream, manifest in studying how to create Julians, is the gift of "time": liberty from humanity's "biological limitations," where "our minds are so powerful, capable of solving any problem, but they're imprisoned in a body that withers, a brain that silts up" (p. 47). This hits uncomfortably close to home as she is retiring from a job she loves, in a field she pioneered, to spend Howard's remaining time with him. She feels maternal affection[9] for Julian, which (much like her relationship with Becky) cannot be reciprocated. Despite a desire and offer to keep working together, Claire knows (because of legal reasons) that she will never get to see her robotic children after they integrate into society.[10]

Julian's instinct to use his connections as Barber to reunite Claire and her daughter is driven by humanitarian concerns—he wants to help Claire to show gratitude for all she has done for him and because they have formed a bond—as much as by the inherent law that Asimov sets out: he cannot allow Claire to come to harm by action or, in this case, inaction by not bringing them back together. Even when Claire verbally lashes out, challenging his humanity with vitriol unseen in the rest of the play, Julian reminds her that he cannot be harmed psychologically (pp. 50-1) and is not deterred. He is still driven by a desire to help, acknowledging, "If I have the ability to relieve [your sadness], to help you – my friend – what kind of person am I if I do nothing?" (p. 50). If his robotic programming

[9] A very different type of affection than what Morton Klass (1983) argues is at the central of the typical human-robot relationship: an assumed "loyalty" on the part of the robot in response to the "affection" it receives, which is generally that "of a subordinate for a permanent underling, perhaps, or the affection we would show to a pet" (p. 176).

[10] This is similar to what Helena experiences in *R.U.R.*, where she "could have been the mother of a new race" of robots as Domin's partner but experiences a rebellion from them and has her "womanhood [...] destroyed" in the process (Rubin, 2011, p. 72): left with neither biological nor robotic children.

endows him with a responsibility to be as human as possible, using Barber's resources to assist Claire makes logical sense to him: since he cannot be a surrogate son for Claire, he does everything in his power to give her a daughter back. He also wryly notes that if Paul wins his lawsuit, he may come and join her someday; if society rejects him as a person, he knows Claire will accept him. This resembles the many connections made to Mary Shelley's *Frankenstein*,[11] though Claire's relationship with Julian is infinitely more positive. However, *Uncanny Valley* and the other plays connect to *Frankenstein* in the human desire to rely on science (AI) to compensate for compassion that the characters cannot find in other humans. While seeking psychological fulfillment, they ultimately use technology for cathartic reasons and seek freedom through creation.

Implementation: *Marjorie Prime*

"Two, [...] a robot must obey the orders given it by human beings except where such orders would conflict with the First Law." (Asimov, 1950, p. 51)

There are several similarities between *Uncanny Valley* and *Marjorie Prime*: a difficult mother-daughter relationship is discussed obliquely with the assistance of an AI surrogate, who has limitations imposed on it by what their human educators choose to share. In the world of *Marjorie Prime*, eighty-five-year-old[12] Marjorie suffers from dementia while living with her adult daughter (Tess) and son-in-law (Jon). Robot technology is accessible for home usage, provided by the Senior Serenity company (Harrison, 2016, p. 15); Marjorie is assisted by Walter Prime, a thirty-something holographic replica of her late husband Walter, endowed with the memories Marjorie and her family have shared with him. Despite reservations from Tess, Walter Prime's purpose is to provide comfort and memory assistance to Marjorie. As with Julian, there is a discussion of what it means to be a person: do robots eventually become more human by acquiring enough memories? If that is the logic employed in robotic development, it raises a larger concern about Marjorie becoming less human as her memory slips away. The Primes recognize that the accumulation of knowledge makes them both "better" and "more human," reinforcing that the latter is their goal (p. 48). However, a source of frustration for Tess is her recognition that her mother is "being nicer to that thing than to me" (p. 18).[13] When Jon reminds her that "It's your father she's being nice to," Tess firmly reminds him, "It is *not* my father" (p. 18). Pronoun usage when discussing the Primes, or robots in general, is of note: here Tess refers to Walter Prime as "it." In *Uncanny Valley*, Claire asks Julian (after the transfer of Barber's memories): "When you think of yourself Julian… do you think "I" or "he"? Or "we?" (p. 37). While Julian acknowledges that he has gotten used to referring to himself as "I," there is a difficult adjustment period for many of these characters to think of the robots (or themselves) as human. The dehumanizing dynamic in using "it" reinforces not just the reality of who is and is not human, but a clear hierarchy with humans far above robots.

Asimov's second law of robotics states that "a robot must obey the orders given it by human beings except where such orders would conflict with the First Law" (1950, p. 51); in short, a robot must listen to its human overseers but cannot cause harm. Applied to *Marjorie Prime*, we see that the robots *listen* more than obey to learn and to facilitate a process where humans can avoid pain. In a harmless example, Marjorie changes the setting of Walter's proposal from a showing of *My Best Friend's Wedding* to *Casablanca*, to romanticize it and remember it differently (p. 10). While this "factual manipulation" can be seen as a way to reclaim authority from Tess (Bendrat, 2023, pp. 222,

[11] See Bruce Malish (1993), who writes about the connections to Shelley, as well as several other literary examples of "automata" inspired by *Frankenstein*.

[12] Born in 1977 (Harrison, 2016, p. 40), the play is set around 2062, a few years later than *Uncanny Valley*. One could read *Marjorie Prime* as a chronological continuation of *Uncanny Valley*, where robots are presumably more affordable and accessible for in-home usage. The process is not the same as achieving immortality, so theoretically these two plays present robots in the same way: learning through conversation, they take on an assigned identity of a human person. This "care technology" serves a larger purpose that Amelia DeFalco (2022) describes as looking to "diminish the economic burden of unproductive bodies on the productive individuals who must support them" (p. 284).

[13] This resentment or jealousy over a human's preference for a robotic counterpart echoes Asimov's short story "Robbie," in which a young girl has an affinity for her robotic nursemaid.

219), the interaction between someone with an incomplete memory and someone (or something) that is reliant on the memory of others to learn poses unique issues for preserving the truth. As DeFalco (2022) describes it, there is an inherent flaw in equating memory and fact since "memory effectively reproduces and replaces its precursor, making every recollection in fact a memory of a memory" (p. 289). A more serious example is Tess looking to spare her mother more pain by allowing her to forget that Tess's brother Damian committed suicide; as such, Walter Prime does not know this information at first and, therefore, cannot remind Marjorie when she asks about Damian. A similar problem arises in *Uncanny Valley*: Julian does not know if what Barber's son says about him being a neglectful and violent father is symptomatic of Paul's psychiatrist suggesting false memories or are truths simply not included in the version of himself Barber wanted preserved and transferred (pp. 43-4). As Harrison (2016) remarks, these robots, or rather "artificial intelligence programs [...] that use sophisticated holographic projections," are "descendants of the current chatbots" (p. 75) that can merely replicate information they are given, are prone to apologizing when they come up short and are incapable of creating something new. In this case, the humans prune memories to create a version of themselves that best suits their needs or their emotional capabilities. Robots "become what the humans want them to be, embodying the characters' own half-truths and frustrations about their histories with the person the hologram represents" (Bendrat, 2023, p. 211).

In the world of these two plays, robots are, therefore, easier to speak to than human characters, not only because of their ability to control them and rewrite history while teaching them but also because of their inherently non-judgmental reception of information. Rather than being estranged like Becky and Claire, Tess simply finds it difficult to speak to her mother, especially after Marjorie's dementia has set in. After Marjorie's death, she is replaced by her own Prime as an emotional support robot for Tess, who has been suffering psychologically. In teaching Marjorie Prime about her mother, Tess embodies the inherent cycle of education that she was working through with Marjorie: parents teach their children, who must ultimately teach their parents again (p. 45). We see the same cycle of grief for both mother and daughter, as Marjorie was initially unable to accept the loss of Damian, which negatively affected the way that Tess processes trauma. As Anna Bendrat (2023) recounts, Tess struggles with the grieving process because "grieving was marred by her mother's physical and mental withdrawal and ensuing emotional coldness, a symptom of Marjorie's advancing depression, which proved devastating to her relation with her daughter" (p. 213). Tess is not only a parent by (re)teaching her mother but also becomes her own mother when she is unable to process the trauma of loss and falls into depression. The presence of the Primes is arguably the closest example of what Tok Thompson (2019) refers to as "android ghosts." Arguing that the purpose of ghosts is to "display the 'shadow' of ethics by haunting individuals and communities with past ethical failures," where "an alternative history" can be presented (2019, pp. 44-5), Thompson's definition fits Tess's usage of Marjorie Prime to move past her failings as a daughter while trying to rewrite history.[14]

When Marjorie Prime demonstrates a lack of humanity, Tess is acutely aware that her 'mother' is now a robot. This triggers Tess's uncanny valley response and exacerbates her psychological struggle. While characters in both plays may experience the uncanny valley phenomenon, it is never a factor for the audience watching the plays.[15] They are consciously aware of the bodies of human performers and never doubt whether they are viewing a robot or a human. If anything, the actors are tasked with reversing the uncanny valley, where their goal is to convince the audience that the human is a robot. As Wade Marynowsky (2012) writes, the logical next step in robot development will be when robots look at humans as imposters and focus on our uncanny resemblance to them (p. 483). As Tess is teaching Marjorie Prime about her namesake, she is eulogizing her mother as much as verbalizing the dynamics of their relationship. When pressed, Tess admits (note the

[14] Thompson (2019) continues to find an overlap between ghosts and androids by recognizing that neither is "really alive but not really dead [...] in the gray zone between objects and people" (p. 46). Though discussing posthumous engagement on social media, he could be speaking directly about the Primes when he says that just "[as] our identities and lived experiences have merged with the digital, so too have our souls after death" (p. 47).

[15] This assumes that the story is experienced as live theatre. Khaled Mostafa Karam (2019) argues that when experiencing a play as a written work, the "readers need to transform the textual written material into an embodied simulation mimicking the characters' experience and sensations" (p. 93).

pronoun usage), "You weren't a bad mom. But we didn't tell each other things, secret things, not really. Some people have a point where their parents stop being parents to them–you start talking as one adult to another. I'm not sure we ever had that" (p. 46). Unlike her mother or James Barber, who wanted a thirty-something replica, Tess's Marjorie Prime looks the same as the older Marjorie. If nothing else, it gives Tess additional time to reconnect and say things to her mother's holographic replica that she couldn't say to her while she was alive. Even Marjorie Prime offers, "Maybe I'm the Marjorie you still have things to say to," giving Tess the space to open up that her mother never did and no longer could in her old age (p. 46). Marjorie Prime embodies this article's reading of Asimov's second law: listening to a human to therapeutically help them avoid pain.

Far from the strange feeling that Tess and Paul had of seeing their fathers younger than themselves, Tess finds herself moved when she opens up to Marjorie Prime and is consoled with, "You shouldn't be so hard on yourself" (p. 48). There is an inherent cathartic motivation in having these conversations with the aged Marjorie Prime since Tess could not connect with Marjorie and still struggles to open up to Jon. But it is not enough, and Tess ultimately cannot escape grief and depression and takes her own life, just like Damian did years ago. Left alone, Jon gets a Tess Prime and undertakes the same process Tess did: eulogizing her while teaching the Prime version about her namesake. However, he quickly experiences frustration at Tess Prime's lack of emotional capability, particularly while recounting the harrowing experience of Tess's struggle with depression and suicide. Tess Prime, like many of these robotic characters, lacks what Steven Sabat (2022) labels "authenticity": "emotional responsiveness, creative social sensitivity, and compassion," as well as the "intrinsic, authentic desire to provide such care rather than merely acting out a programmed series of possible responses that may or may not apply" (p. 296). Jon ultimately relents and agrees with Tess's earlier skepticism, acknowledging, "It's nothing. It's a blackboard. I'm talking to myself" (Harrison, 2016, p. 67).[16] This is further exemplified in the final scene, where the three Prime versions—Walter, Marjorie, and Tess—converse together, weaving in all of the details that the audience knows to be incorrect. With the ability to cherry-pick our memories and alter our experiences, a tool to help becomes an avenue for historical revisionism: erasing painful memories or the bad things we have done, engaging in a form of self-selected memory eugenics that solidifies that "memory and truth are often mistaken for each other" (Bendrat, 2023, p. 220). It reinforces that as much as the fictional creation of advanced AI technology offers potential benefits, the implementation is faulty, and the repercussions are potentially dangerous.

Repercussions: *The Nether*

"[T]hree, a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws." (Asimov, 1950, p. 51)

The first two plays discuss the ways that sentient AI-infused humanoids are created and implemented into our world; they consider what happens as robots look to bridge the uncanny valley and become more human. *The Nether* focuses on humans immersing themselves in a computer-generated world as avatars and examines the lengths they will go to leave the constraints of what Claire refers to as "biological limitations." With a setting that simply reads "soon," the play portrays a future where a widespread virtual world (called The Nether) is inhabited by "shades,"[17] human characters that have left the real-world constraints behind. The characters quickly reveal that the

[16] This is seen in the contemporary usage of AI chatbots, who are limited to finite information and response coding. Think of the use of AI for therapy, where a patient might be more comfortable opening up to a non-judgmental AI therapist but might also be dismayed by the lack of emotional connection and genuine empathy, leading to a similar dismayed and embarrassed conclusion as Jon's.

[17] Presumably a reference to the inhabitants of Dante's *Inferno*; on this reading, it implies that the characters are stuck in a cyclical loop (reinforced by the unchanging scenery in which Sims sets The Hideaway), but also that this is a place to think on one's sins. However, at the Hideaway, they are indulging in their sins without consequence, and there is no opportunity for remorse. June Xuandung Pham (2018) in particular argues that, at the end of the play, "it is hard to say whether justice has been served, as there is simply no established legal basis to deal with digital and non-digital subjects in a mixed reality" (p. 6).

virtual world is more important than the real one, as numerous users have made the transition permanent (and live on life support in the real world). More so than the other two plays, *The Nether* offers characters the appealing opportunity to be who they want to be, transcending limitations: avatars have no direct correlation to who someone is (or what they look like) in the real world. As much as the legal status of robots is in question in *Uncanny Valley*, the ramifications of actions taken in The Nether are also a source of discussion. Morris, a detective, is interrogating Sims,[18] the creator of The Hideaway: an area of The Nether where the participants can have sex with virtual children and murder them. While Sims protests that all avatars represent legal, consenting adults, their usage raises ethical dilemmas about whether the actions taken in a virtual space can be punished in the real world and, if so, who should be held accountable. While contemporary lawmakers look to catch up with technological advancements and impose oversight, all three plays (nearly a decade old) reinforce that robots and AI can be re-appropriated from their original purpose for 21st-century service; therefore, Asimov's laws of robotics need to be updated to reflect changing times to these new purposes.[19] His third law states, "a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws" (Asimov, 1950, p. 51). In this case, Sims is both the human creator and an active virtual participant (called "Papa"), blending the two roles and playing with Asimov's rules: he seeks to protect himself and, more importantly, The Hideaway.

While Morris interrogates Sims and Doyle (a middle-school science teacher[20] and frequent user), their scenes are intercut with those in The Hideaway, depicting the interactions between Iris, a young girl, and Woodnut, a visitor who spends his time with her. Sims has designed his realm as both aesthetically appealing and never-changing: Iris always looks the same, never ages, and is unaffected by any potential physical damage done to her.[21] For Sims, protecting his realm's existence is two-fold: he argues that the actions taken in the virtual world are less destructive than those taken in the real world, and has placed several safeguards to protect those involved. In addition to background checks and informed consent, The Hideaway encourages the use of violence to remove emotional connections between the client and Iris. When Sims/Papa suggests that Iris "nudge" Woodnut to partake in the violence, he reminds her that "It keeps them from getting too close. That's something you should watch, as well. It makes you vulnerable and could upset a balance" (Haley, 2014, p. 41). This is not a violation of Asimov's law because Sims is protecting the larger infrastructure, not the individual avatar who miraculously recovers after each encounter. Additionally, as Iris reassures Woodnut, "I feel only as much pain as I want" (p. 50). This removes the client's burden even further and allows both the client and Iris (or her human controller) to go through a cathartic experience together without long-term consequences. Iris is forever preserved (or "resurrected" in her words (p. 50)) and it is only in the real world that Morris is trying to impose consequences. The Nether is not that different than the usage of robots discussed above: for Sims, it is a way to memorialize Iris (once a real young girl) and keep her with him forever for personal reasons. Sims looks to instill order and

[18] Presumably a reference to the video game series in which the player creates virtual realities that simulate real life.

[19] Katherine L. O'Grady et al. (2022) propose updating the laws as well, based on fostering "trust and ethical consideration [...] between humans and machines—regardless of whether machine consciousness is considered real—to ensure continued collaboration and shared success" changing the field of AI (p. 2). Their focus removes any discussion of harm or hierarchy-based obedience and encourages increased study of human behavior and an open dialogue to encourage self-awareness and mutual respect (p. 8).

[20] Doyle admits that his journey to The Nether came first as part of his job. Despite receiving acclaim through awards and job offers, Doyle opted to remain teaching at the middle school level. He laments that the "Single School Act" moved education online. Frustrated by the monotony and ineffectiveness of teaching online, he drifted and explored until he discovered The Hideaway. He anticipated many of the frustrations that educators would experience in the move to online education, particularly during the Covid-19 pandemic.

[21] In her author's note, Haley (2014) states that "It is important to cast Iris with an actress who will appear on stage as a prepubescent girl," since that will reassure the audience that "nothing awful will be enacted upon the child, whereas they have no such confidence with as adult posing as a child" (p. 66). This is a different version of the uncanny valley: the audience feels unease as an actress playing a young girl *approaches* looking/acting like a child, but confidence is restored when she replicates (actually is) a child. Isabel Stowell-Kaplan (2015) looks to reconcile whether the deliberate casting is "merely a sensationalist trick or is it effective theatre looking for a way to shock its audience into contemplation of just how intertwined these different versions of ourselves might be" (p. 163). Utilizing human actors and realistic settings to represent their virtual counterparts is what Selim (2022) labels "realistic virtuality," which allows the audience to experience the transhuman elements in the same way the characters do (p. 312).

consistency in the realm he has created, giving users a unique experience free from moral or legal constraints. Even though he has pushed the limits of the usage of virtual reality without seeing any potential negative repercussions, that does not mean that there are not any.

Asimov says self-preservation should not come at the cost of human harm; while there is only temporary harm to the avatars, the people behind them can be hurt psychologically. Sims has a long-standing rule that when one Iris becomes too problematic, she is sent to "boarding school" and another replaces her.[22] When Iris asks Sims/Papa if she is special or if he would like to spend eternity with her, he refers back to the safeguards in place: "I need to remind you this is a business" where "objectivity is required to keep us afloat" (p. 57). But Iris's controller, later revealed to be Doyle, is emotionally invested: asking Sims if the love that Doyle expresses (as Iris) for Papa is reciprocated. Though the specific financial elements of The Hideaway are not discussed, Sims has reportedly made a lot of money and Doyle had at some point "spen[t] all of [his family's] money" there (p. 54). To "stay in the Hideaway" and "[save] enough money for [his daughter] to finish college," (pp. 54, 24) Doyle took over as Iris at Sims's suggestion, indicating that there is a financial component to 'being' Iris as well as a psychological one. Since Doyle's love for Sims is not reciprocated, he assists Morris in exposing Sim's location; but after Sims is caught, Doyle is distraught and kills himself. Even when Morris talks about the real-world consequences of the sixty-five-year-old man's death, Sims can only refer to him as Iris or "her"—yet another conflation of pronouns that becomes confusing with technology. Sims and other characters look to maintain the status quo, particularly in their relationships with other humans they cannot let go of, by leaning into futuristic technological advancements.

While robots are not supposed to harm humans, interactions with them are unintentionally hurtful when they reveal something about the human characters: usually unresolved trauma or human desires for love, both familial and romantic. Morris, like Tess before her, struggles to connect with a parent; her father has given himself over full-time to The Nether, endowing her with a personal distaste for technology. However, when Morris is revealed to be Woodnut, she has yet another stake: falling in love with Iris but being forced to 'kill' her repeatedly to follow the rules and sever emotional connections. Since Morris is unable to distinguish between the real and virtual worlds—or the "humanist" and "transhumanist" according to Selim (2022, p. 308)—she (as Woodnut) is so appalled by the frequent 'killing' of Iris with an axe without repercussions that she ultimately laments "if there has been no consequence, there has been no meaning [...] and if there has been no meaning, then I am a monster" (Haley, 2014, p. 53). The word "monster" is the same descriptor that Paul applies to the cyborg version of his father, Julian Barber, which horrifies Claire (Gibbons, 2015, pp. 41-2). Not quite technophobic, the plays reveal that the capabilities of humans, not their robotic creations, are the problem—akin to the *Frankenstein* discourse. The human characters realize that their technological advancements ultimately fail at simulating or replacing genuine human connection, revealing the same innate desire for something that they were looking to avoid in the first place: uncomfortable in-person conversations with fellow humans. The ability to merely mirror humans, rather than creating something new that fully replaces humans, reveals the faults in the system and defeats that purpose of robots. Much like in *R.U.R.*, the very tools designed to grant humans utopian freedom instead threaten their ontological status through dependence on technology and reaffirm their need for pre-technological connections with genuine humans.

At the heart of all of these plays is the question: what constitutes humanity as opposed to robotic imitations? *R.U.R.* uses the word "soul" to refer to the internal desires or essence of humanity that Helena desperately wants robots to have. The nebulous theological concept of a soul is merely a stand-in for other elements that constitute humanity: a sense of existential purpose, a desire for self-preservation, or an appreciation for love. But that discussion merely focuses on how to make robots

[22] *R.U.R.*'s robots are disposable under similar circumstances when they experience a "Robotic Palsy" and need to be destroyed. In the vein of the *Frankenstein* connections, the most human-like action that robots can take is to rebel against their creators. This may be why the humans in *R.U.R.* lean so heavily into a discussion of religion: turning to metaphysical sources to explain the technological missteps they have taken to lead to their own plague-like extinction. By contrast, the humans behind the technology in these three plays are actively working to make themselves replaceable to mixed success.

more human-like; it does not explain what it means to make a robot an actual person—whether an original creation or a replacement for an existing human. *Uncanny Valley* and *Marjorie Prime* ask whether simply uploading a person's memories into a humanoid or holographic body that does not age or die is the solution to immortality, even when the new creation is far from a suitable replacement for the original person. *The Nether* abandons this in service of offering not immortality but an entirely alternative life where you can embody a person of your choosing and act without consequences. Building off of Domin's admonition that "No one can hate more than man hates man" (Čapek, 2004, p. 58), Anderson (2014) argues that "humanity is defined as much by its self-hatred as it is by self-love" and that the attempts "to close the gulf between the human and nonhuman" ultimately reveals an ingrained "auto-immune response" that leads to humans ironically promoting self-preservation by means that lead to them becoming "superfluous" (pp. 241, 235). In all three plays, abandoning one's original body for a surrogate (whether robotic or virtual) poses ethical dilemmas as characters use technological creations to solve human emotions—a historically unsuccessful matchup. Embracing technology requires reconciling the existing framework of Asimov's laws to mitigate potential dangers. However, as Charles Rubin (2011) points out, if robots are reliant on humans to teach them morality in a "rule- or virtue-based approach to programming," they are limited by (and a reflection of) the faulty moral relativism that humans experience (pp. 61, 59).[23]

Gordon Beauchamp (1980) questions why any programmed robot would need to be constrained by Asimov's laws at all if they are merely implemented designs by humans and do not have any "natural instincts" to curb (p. 86); if anything, he argues, the laws are made to protect robots, not humans, with an emphasis on self-preservation for the greater good (pp. 89, 91). The human characters use robots, AI, or virtual reality to live Asimov's laws themselves: to avoid harm, to be listened to, and to protect themselves at all costs. While *R.U.R.*'s humans recognize that pain makes robots more effective, the human characters discussed here operate within the confines of those laws to eliminate emotional pain. Yet while seeking to liberate humanity with technology, creators of this technology have made humans dependent on it instead (Rubin, 2011, p. 69). Hence, they must also reconcile their human desires with what is created and released into the world: a not-quite-perfect Julian Barber simulacrum, a room full of Primes with incorrect information, and (with The Hideaway shut down) enraged pedophiles with no digital outlet. In *R.U.R.*, once the robots kill off humans, and with no way to reproduce, they have given themselves a twenty-year expiration date. Despite the optimism of the love between the robots Primus and Helena, there is a clear trajectory for the end of the robotic takeover. By contrast, Sims speaks to the other plays' uncertainties about the long-ranging repercussions of well-intentioned but ill-considered technological creations when he tells Morris, "You don't know what you do, Detective, putting me out in the world" (Haley, 2014, p. 63).

As AI becomes more central to a 21st-century world, rethinking and recontextualizing the relationship between humans and AI is crucial. Intending to use technology to gift humanity immortality, compassion, and support is ambitious, if chimeric, because of the unresolved ethical concerns. Examining these three plays through the lens of the three stages of robot integration and the three laws of robotics reveals a more nuanced relationship between creator and creation. They demonstrate that what begins as an attempt to bring about utopian innovations ends with existential doubts as humanity looks to replace itself with technological surrogates or, as Anderson describes it, "to reflect upon finitude with fresh creativity" (2014, p. 240). Because it is perceived to be easier to fabricate emotional connections with AI substitutes than to deal with "horribly imperfect" humans, stripping humanity down to a specific essence (i.e., a collection of memories or a visual representation) underestimates the trauma of the uncanny valley response. The irony is that while humans use AI technology to mitigate the psychological trauma of loss, the inability to fully replicate humanity only exacerbates isolation, depression, and regret for the characters discussed here. While Asimov established rules to protect humans from robots, whether physical harm or *R.U.R.*'s existential crises, the rules must focus (perhaps exclusively) on psychological damage. Beyond

[23] HAL 9000 in Stanley Kubrick's *2001: A Space Odyssey* (1968) is a good example of a robot who, when faced with making a subjective decision, prioritizes the larger programmed objective of mission success over the safety of humans; while a clear violation of Asimov's laws, this reveals that robots can engage in moral relativism to justify those actions.

Asimov, these three plays instill in their audience a vital sense of responsibility to resolve the ethical considerations that are quickly moving from speculative to reality.

Conflict of Interest:
The author declared no conflict of interest.

References

Anderson, N. (2014). "Only We Have Perished": Karel Čapek's R.U.R. and the Catastrophe of Humankind. *Journal of the Fantastic in the Arts*, *25*(2/3 (91)), 226–246. http://www.jstor.org/stable/24353026

Asimov, I. (1950). *I, Robot*. Doubleday & Company.

Ayres, R., & Miller, S. (1982). Industrial Robots on the Line. *The Journal of Epsilon Pi Tau*, *8*(2), 2–10. http://www.jstor.org/stable/43602572

Beauchamp, G. (1980). The Frankenstein Complex and Asimov's Robots. *Mosaic: A Journal for the Interdisciplinary Study of Literature*, *13*(3/4), 83–94. http://www.jstor.org/stable/24780264

Bendrat, A. (2023). "How Do You Know Who You Are?": *Marjorie Prime* on Envisioning Humanity Through the Faculty of AI-Powered Memory as Reconstructive Tissue." *Text Matters: A Journal of Literature, Theory and Culture*, (13), 210-28. doi:10.18778/2083-2931.13.12

Čapek, K. (2004). *R.U.R. (Rossum's Universal Robots)*. (C. Novack, Trans.). Penguin.

DeFalco, A. (2022). Posthuman Care and Posthumous Life in *Marjorie Prime*. In M. Goldman, K. de Medeiros & T. Cole (Eds.), *Critical Humanities and Ageing: Forging Interdisciplinary Dialogues* (pp. 283-292). Routledge.

Gibbons, T. (2015). *Uncanny Valley*. Dramatists Play Service Inc.

Haley, J. (2014). *The Nether*. Faber and Faber.

Harrison, J. (2016). *Marjorie Prime*. Theatre Communications Group.

Jiang, H., Cheng, L., Pan, D., Shi, S., Wang, Z., & Xiao, Y. (2022). Virtual Characters Meet the Uncanny Valley: A Literature Review Based on the Web of Science Core Collection (2007-2022). *2022 International Conference on Culture-Oriented Science and Technology (CoST)*, 401-406.

Karam, K. M. (2019). Between Sensorimotor Data and Conceptual Message: Embodied Simulation as an Approach to the Reading of Two Science Fiction Plays, Jennifer Haley's *The Nether* and Peter Sinn Nachtrieb's *Boom*. *International Journal of Applied Linguistics & English Literature*, *8*(2). http://dx.doi.org/10.7575/aiac.ijalel.v.8n.2p.85

Klass, M. (1983) The Artificial Alien: Transformations of the Robot in Science Fiction. *The Annals of the American Academy of Political and Social Science*, *470*, 171-179. https://www.jstor.org/stable/1044811

Malish, B. (1993). "Automata." In *The Fourth Discontinuity: The Co-Evolution of Humans and Machines* (pp. 31-58). Yale University Press.

Marynowsky, W. (2012). The Uncanny Automaton. *Leonardo*, *45*(5), 482-483. https://www.jstor.org/stable/41690230

O'Grady, K. L., Harbour, S. D., Abballe, A. R. & Cohen, K. (2022). Trust, Ethics, Consciousness, and Artificial Intelligence. *IEEE/AIAA 41st Digital Avionics Systems Conference (DASC)*, 1-9. doi: 10.1109/DASC55683.2022.9925874

Pham, J. X. (2018). Jennifer Haley's *The Nether*: Digital and Inhuman Subjectivities on Stage. *Angles: New Perspectives on the Anglophone World*, *7*. https://doi.org/10.4000/angles.752

Rayner, A. (1994). Cyborgs and Replicants: On the Boundaries. *Discourse*, *16*(3), 124-143. https://www.jstor.org/stable/41389337

Rubin, C. T. (2011). Morality and Human Responsibility. *The New Atlantis*, (32), 58-79. https://www.jstor.org/stable/43152657

Sabat, S. R. (2022). Response 14: Only Persons Can Provide Person-Centered Care for People Living with Dementia: "Walter Prime" and His Ilk Miss the Mark. In M. Goldman, K. de Medeiros & T. Cole (Eds.), *Critical Humanities and Ageing: Forging Interdisciplinary Dialogues* (pp. 293-298). Routledge.

Selim, Y. F. (2022). Jennifer Haley's *The Nether*: Transhumanism in the Post-Internet World. *Canadian Review of American Studies*, *52*(3), 300-316. https://doi.org/10.3138/cras-2022-002

Shuffelton, F. (Ed.). (2004). *The Letters of John and Abigail Adams*. Penguin.

Stowell-Kaplan, I. (2015). "In the Domain of 'The Nether": Theatre and Virtuality in a World without Consequence. *TDR (1988-)*, *59*(2), 157–163. http://www.jstor.org/stable/24585015

Thompson, T. (2019). Ghost Stories from the Uncanny Valley: Androids, Souls, and the Future of Being Haunted. *Western Folklore*, *78*(1), 39–66. https://www.jstor.org/stable/26864141

Author Information:

Jarrod DePrado is an instructor at Sacred Heart University in the Departments of Languages & Literature and Catholic Studies. He received his graduate degree from Boston University in English and American Literature. His area of specialization is transhistorical drama—bridging Shakespeare, 20th- and 21st-Century American Drama, and musical theatre—with a focus on adaptation studies and American politics. He will begin his PhD in English at the University of Connecticut in Fall 2024.