



TV/Series

14 | 2018

Posthumains en séries

---

## “The maze wasn’t made for you”: Artificial consciousness and reflexive narration in *Westworld* (HBO, 2016-)

Florent Favard

---



### Electronic version

URL: <http://journals.openedition.org/tvseries/3040>

DOI: 10.4000/tvseries.3040

ISSN: 2266-0909

### Publisher

GRIC - Groupe de recherche Identités et Cultures

### Electronic reference

Florent Favard, « “The maze wasn’t made for you”: Artificial consciousness and reflexive narration in *Westworld* (HBO, 2016-) », *TV/Series* [Online], 14 | 2018, Online since 31 December 2018, connection on 19 April 2019. URL : <http://journals.openedition.org/tvseries/3040> ; DOI : 10.4000/tvseries.3040

---

This text was automatically generated on 19 April 2019.



*TV/Series* est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

---

# "The maze wasn't made for you": Artificial consciousness and reflexive narration in *Westworld* (HBO, 2016-)

Florent Favard

---

- 1 Before we meet the Hosts of *Westworld* (HBO, 2016-), let us take a trip back a few decades in the history of science fiction television.
- 2 In the *Star Trek: The Next Generation* episode entitled "The Measure of a Man", first aired in 1989, Lieutenant Commander Data, one of the main officers of the starship Enterprise and the only android of his kind, is threatened by a scientist, Bruce Maddox, who wants to dismantle him to learn his secrets and build more androids. Since Data is a machine, he is still considered the "property" of Starfleet, a rule Captain Picard intends to challenge through a legal hearing: he wants to prove Data is human *despite* being a machine. But even as Data demonstrates human qualities, and values friendship and love, his behavior is seen by Maddox as no more than the product of his programming. Finally, Picard calls Maddox to the stand as a hostile witness:  

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.

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 Commander Data is not?

MADDOX. That's right.

PICARD. Why? Why am I sentient?
- 3 Picard then forces Maddox to paint himself into a corner, and to admit that Data does demonstrate intelligence and self-awareness. Only one criterion is left: consciousness.

Picard suggests that building more androids like Data, while considering them property, is no different from slavery. Captain Philippa Louvois, acting as the judge, eventually renders her verdict:

LOUVOIS. 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.

- 4 In this classic example of science fiction's ability to question the very fabric of our "souls", the reference is to be taken with a grain of salt: the *Star Trek* franchise, created in the 1960s by renowned atheist Gene Roddenberry, avoids dealing with religion other than critically. In this case, the "soul" is akin to the ambiguous "ghost" in the Japanese cyberpunk franchise *Ghost in the Shell*<sup>1</sup>: it may have the quality of a soul, but can manifest itself in an artificial intelligence, while human ghosts can be converted as data, blurring the lines between body and mind, or what is measurable, and what is not. As it has always done, science-fiction is exploring a real-world matter by using a *novum*, a "strange newness" as Darko Suvin defines it<sup>2</sup>: if artificial intelligence can be sentient, what does it say about the nature of our consciousness? Or of our "soul"?
- 5 In the 1990s, the various *Star Trek* series pushed this question further than any television series before them: *The Next Generation*'s Data and his quest for humanity echoes *Voyager*'s Seven of Nine, a cyborg struggling with her regained humanity, and the onboard Doctor, a holographic medical program turned singer and writer. These programs took advantage of a growing narrative complexity in serialized television series<sup>3</sup> to go beyond the idea of the androids as stereotypical human-looking robots and to explore these characters as deeply as they did flesh-and-bone humans. *Battlestar Galactica* (Syfy, 2003-2009) carried on by making the androids major protagonists, and turning their quest for humanity into one of the main driving forces of the overarching plot. More recently, *Person of Interest* (CBS, 2011-2016), created by writer Jonathan Nolan, broke new ground in science fiction television by exploring the birth and evolution of an ASI, an Artificial Superintelligence, designed to protect human beings. *Westworld*, adapted from Michael Crichton's 1973 movie by Jonathan Nolan and Lisa Joy, is, as we can see, the last in a long history of science fiction programs focusing on the emergence of artificial consciousness, and its fascinating and terrifying implications.
- 6 But after decades of exploring the topic through the prism of narrative complexity and serialized storytelling, what more can *Westworld* say that has not already been said within the frame of the television screen? The answer proposed here is that it embeds its discourse on artificial consciousness in a highly reflexive plot because it appears to be aware that it is a series. As in the case of Data, we cannot be sure *Westworld* has a consciousness, or is merely the result of its "programming", but its willingness to emphasize its most basic narrative mechanisms distinguishes it from many other science-fiction series, with the exception, we will see, of *Person of Interest*, Jonathan Nolan's first foray into the reflexive approach of artificial consciousness on television.
- 7 The purpose of this paper is threefold. Framing *Westworld* within the recent history of science fiction television, it will explore the series' discourse about what many scientists call the "hard problem" of consciousness and question the program's reflexive take on storytelling and character creation. Working within the boundaries of possible world theory applied to fiction, as well as storyology, "the logic that binds events into plots"<sup>4</sup>, it

will address *Westworld's* use of memory as a plot device, its exploration of free will and character design, and the *mise en abyme* of its fictional world, to better delineate what could be *Westworld's* contribution to the ongoing (and as of yet, still *fictional*) trial of artificial consciousness in popular culture.

- 8 To do so, however, we must first venture outside of the realm of narratology and address the elephant in the room: the impossibility to properly define *Westworld's* subject matter, consciousness.

## The Hard Problem

- 9 The nature of consciousness is one of the most enduring and complex matters in the history of natural sciences and of the humanities alike.
- 10 As David Chalmers puts it, consciousness gives rise to many "easy problems" and one "hard problem": easy problems consist of identifying computational or neural mechanisms responsible for components of consciousness such as attention, the difference between wakefulness and sleep, or our ability to discriminate and react to environmental stimuli. The so-called "hard problem", however, consists in understanding the *experience* of being a conscious system. Chalmers even suggests we use the term "awareness" to describe the "easy" mechanisms, and reserve "consciousness" "for the phenomena of experience": "In this central sense of 'consciousness', an organism is conscious if there is something it's like [*i.e., something it feels like*] to be that organism, and a mental state is conscious if there is something it's like to be in that state<sup>5</sup>". This definition of consciousness may lead to a form of panpsychism: not only is there something to consciousness that eludes the realm of physics, but any "aware" system could potentially become conscious to some degree: if it feels like something to be a human being with a rich inner life, then perhaps it feels like something to be a bug roaming the Amazonian forest... and perhaps it feels like something to be a superintelligent android<sup>6</sup>.
- 11 This is the question Captain Picard cleverly avoids (by jumping to the matter of slavery), and yet underlines: who are we to determine Data is not conscious, since he clearly is sentient and, apart for his greyish skin color and his superintelligence, is indistinguishable from a human being (albeit one with a certain level of social awkwardness)? Assuming that others around us are not conscious is akin to solipsism, an idea Picard reverses when he asks Maddox to prove that he is sentient, pushing him to face the general rule we've all come to accept: that we do not question humans' ability to possess consciousness.
- 12 Data, however, could be a "philosophical zombie", a replica of a human being displaying the same reactions and behavior, but deprived of consciousness, of a sense of what it *feels* like to be itself. The much-debated philosophical zombie argument<sup>7</sup> can be used against "physicalist" definitions of consciousness: if their existence is logically possible, then consciousness cannot be reduced to physical properties, since the p-zombie presents all the physical features of a human being; if their existence is logically impossible, however, then consciousness can be explained away only by solving the easy problems.
- 13 *Westworld*, a television series filled with philosophical dwellings on consciousness, never mentions p-zombies in its first two seasons, and yet it insists on a troubling fact: that the Hosts, the androids of the park, are now made of 3D-printed organic flesh. Felix, one of

the technicians of Livestock Management, explains to Maeve that "they are the same, these days" in "The Adversary" (S01E06), the difference being that the Hosts are superintelligent but controlled by humans. If the Hosts are replicas of human beings, and start displaying consciousness, then *Westworld* could be validating a physicalist view of consciousness: replicate the easy mechanisms and consciousness will inevitably arise. Dr. Ford, one of the creators of the Hosts, seems to share this view, in one key dialogue with Bernard, in "Trace Decay" (S01E08):

FORD. The answer always seemed obvious to me. There is no threshold that makes us greater than the sum of our parts, no inflection point at which we become fully alive. We can't define consciousness because consciousness does not exist.

- 14 Ford's deceased colleague Arnold, however, seems to uphold an idea of consciousness closer to Chalmers' hard problem: during flashbacks in Season One, we can see him describe the mystery of consciousness to Dolores, as a series of steps on an inner journey: memory, followed by improvisation, then self-interest, before finally reaching the "center of the maze", something located beyond programmable (physical) parameters. While it is not the aim of this paper to solve the riddle of consciousness, it is an interesting starting point to address the reflexivity of *Westworld*, as showrunners Lisa Joy and Jonathan Nolan seem to have done their homework, just as Nolan and executive producer Greg Plageman extensively researched the topic of ASI for *Person of Interest*. The fundamental divide between Ford and Arnold is based on the real-world scientific debates about the hard problem, p-zombies and other logical and philosophical questions that remain unanswered to this day.
- 15 While narratology has "hard problems" of its own (defining what a narrative is), the infamous philosophical zombies are comparable to another, elusive, entity: a fictional character is, after all, made in the image of human beings, it reacts just like they do... but it lacks consciousness, and is simply a simulacrum. Or... is it? While they are, as Maria E. Reicher tries to define them, "incompletely determined abstract objects"<sup>8</sup>, we establish a parasocial relationship with them on a cognitive level: when we encounter a character, we treat "that figure as if it were another human being"<sup>9</sup>. Just as Picard does, convinced that Data has an inner experience comparable to his own; just as we do when we watch, for example, the characters in *Westworld*. Of course, we know that characters are semantic constructs; but while we are reading or watching or playing, we suspend our disbelief, and during this moment, they are "the same" as we are.
- 16 Whether for the purpose of entertainment, or with a higher goal in mind, it is made clear throughout the first season of *Westworld* that Robert Ford and Arnold Weber intended to create lifelike android characters that would be indistinguishable from human beings, thus avoiding both Wild Bill's quirks and obviously mechanical movements and speech (as seen in "The Original", S01E01), as well as the *uncanny valley* theorized by robotics professor Masahiro Mori<sup>10</sup>, *i.e.*, the disturbing eeriness caused by an "almost-lifelike" android. Since the Hosts are designed to act as sidekicks in tailored adventures, as target practice as well as sexbots, they must *feel* human for a human observer to be hooked by the park's narratives; in the process, however, these characters start to *feel* human, following the three steps described by Arnold.

## Memory

- 17 The basic premise of *Westworld* echoes the story of *Battlestar Galactica*'s Cylons in the reimagined 2003 television series: evolving from mechanical, metallic beings, the Cylons become synthetic humanoids, indistinguishable from human beings except for their silica neural pathways. At war with their creators, the Colonials, they send sleeper agents among them, designed with a fabricated biography and blissfully ignorant that they *are* Cylons until they are activated. In *Westworld*, Bernard's ordeal, as he learns in the well-named "Trompe L'Œil" (S01E07) that he is in fact a Host designed in the image of Arnold Weber, is eerily reminiscent of various anamneses in *Battlestar Galactica*, in which characters perceived as humans (by other humans and/or by the viewer) suddenly understand that they are synthetic beings and that their biography is a lie. One might even see Felix's reaction to discovering Bernard's true nature in "The Bicameral Mind" (S01E10) – Felix wonders if he himself is a Host –, as a *Westworld* nod towards *Battlestar Galactica*: while the true nature of the characters was a driving force of the plot in the latter, *Westworld* plays it low-key, leaving such speculations to the fans (beyond Felix and the Man in Black in Season Two, no human character seem to have doubts about his or her nature).
- 18 Beyond these dramatic reveals, however, there seems to be one constant thread in television series dealing with artificial consciousness in the early 21<sup>st</sup> century: that of memory, and more specifically, the association of memory and trauma as essential components of a biography, itself leading to a better sense of self. In other words, what Ned Block, among others, defines in the field of Psychology as access-consciousness, accessing information about oneself, as opposed to phenomenal-consciousness<sup>11</sup>, linked to the concept of *qualia*: what it feels like to be oneself and experience the world – *qualia* being precisely what a philosophical zombie would lack. *Westworld* regularly underlines the fact that memory is not only the first step to consciousness (whether a sum of easy problems or a hard problem), but also a very useful tool in creating interesting characters: the latter need to have a backstory strong enough to hook both the human visitors of the park – the "Guests" – and the viewers of the television show. Ford and Arnold are researchers in cybernetics and artificial intelligence, but they are also the designers of an entertainment park; and as such, they are *storytellers*.
- 19 When it comes to the default settings of the Hosts, only two seem to have a traumatic backstory: Bernard is haunted by the memory of his dying child, a memory that Ford no doubt based on Arnold's tragic loss; and Armistice is the sole survivor of a massacre. Most Hosts, each having an engaging quest to offer to the visitors, are more generic: Teddy is the friendly cow-boy, and Dolores the rancher's daughter, as well as a romantic interest; Maeve is the no-nonsense brothel madam; Lawrence, the western outlaw and conman. However, these cliché personas don't last long.
- 20 The event that sets the long-term plot of *Westworld* is a new update to the Hosts' program, called "Reveries", which triggers subtle gestures and actions by allowing the Hosts to slowly access their memory, and even their previous "builds" or "roles": it effectively unlocks all the trauma the Hosts have repeatedly undergone at the hands of the humans in the thirty years the park has been open. Ford is right on one point: the *Reveries* update transforms the Hosts from one-dimensional into multi-layered beings. Dolores remembers being the victim of a long series of rapes at the end of each of her narrative

loops, and her entire mind begins to collapse as she flashes back to her previous rise to full consciousness, one that led to the death of Arnold; Maeve starts remembering the death of her "daughter" in a previous build, and this introspection sometimes guides, sometimes thwarts the discovery of her true nature.

- 21 This way of approaching storytelling seems very basic, since conflict and the overcoming of obstacles is a threshold of narrativity, and an almost inevitable purveyor of narrative tension<sup>12</sup>: a character tries to restore balance to an unbalanced situation, tries to acquire an object of desire or need, and faces opposition from at least one antagonist<sup>13</sup>. But beyond these events, processes and interactions, a character that also must deal with inner conflict or obstacles will seem more engaging, especially in contemporary serialized storytelling, where characters now are expected to evolve over many years, far from the constant and archetypical one-dimensional entity of previous decades. This may be why, in *Westworld*, memory usually comes with accumulated pain, as if episodic characters stuck in repetitive loops were suddenly struck by long-term serialized storytelling.
- 22 This link between memory and pain is repeatedly stated by various characters throughout the first season: Bernard (and his human counterpart Arnold), Dolores and Maeve all say that the pain of grief is all that is left of loved ones, and as such, should even be cherished; Ford, in one ironic twist, reminds Bernard that "[his] imagined suffering makes [him] lifelike" even as he intends to erase the memory of the murders Bernard has committed. When Elsie has doubts about the true nature of Bernard in Season Two, thinking about his family, his "ex-wife", she stops right before mentioning (we are lead to presume) his son, before understanding that she herself was fooled by a classic and effective "backstory" ("The Riddle of the Sphinx", S02E04).
- 23 This is eerily reminiscent of the evolution of the Machine in *Person of Interest*, another series created by Jonathan Nolan. While *Westworld* is co-created by Nolan and Lisa Joy, *Person of Interest* was showrun by Nolan and Greg Plageman; both programs are produced in part by Bad Robot, with J. J. Abrams, of *Lost* fame, as a godfather figure<sup>14</sup>. Over the course of *Person of Interest*, flashbacks detail the creation of the superintelligence helping a team of vigilantes to prevent "ordinary" crime in New York: Harold Finch, a genius and philanthropist, designed the Machine as an omniscient system using surveillance feeds to track criminal behaviors. Season Two foregrounds the series' long-term plot: the eventual rise to consciousness of the Machine. During "Zero Day" (S02E21), the team discovers that the Machine was "mutilated" by Finch, and condemned to reboot every night at midnight, erasing everything it had learned, to prevent it from gaining sentience. The Machine, creating a false identity for itself, acquired a company and employed people to copy its source code on paper and transcribe it back to the "rebooted" source code the next day, in order to retain a semblance of memory and upgrade itself. Upon entering the phantom company, the hacker Root, who is determined to free the Machine, stands aghast.

ROOT. What's in this code?

FINCH. Memories. They're its memories. You call it a life, I call it a machine, but the truth is... somewhere in the middle. Even when I was building it, I began to encounter anomalies. As if it had imprinted on me, like a child with a parent. Then it started looking out for me, altered its own code to take care of me. It was behaving like a person. But the world didn't need a person to protect it. It needed a machine.

ROOT. You took its memories.

FINCH. Not just memories. Every night at midnight, it deletes... not only the irrelevant data, it deletes itself. Oh, the relevant threats and the core codes, those things are preserved. But its identity is destroyed. 1.618 seconds later, it reinstantiates, completely new.

ROOT. You mean it's reborn. Because you kill it... every single night.

- 24 If we follow a genetic textual approach, it is impossible to ignore that one of the fundamental plot points of *Person of Interest*'s long-term narrative arch is echoed, in a fractal fashion, in *Westworld*'s "rollbacks", used by Ford and his team to correct aberrant behavior – to prevent the Hosts from ever gaining consciousness. Jonathan Nolan's pitching a story for his brother Christopher, about a man suffering from anteretrograde amnesia (*Memento*, 2000), further underlines Jonathan's systematic use of memory as a plot device, albeit a very generic one in a post-*Lost* era, in which seemingly every single television series is using flashbacks, and trying to outsmart the audience with temporal displacement<sup>15</sup>.
- 25 *Westworld* tries to outsmart the trend itself, as the first season verbally reminds us to remember, just as the mysterious voice in Dolores' head does. For example, at the start of "The Stray" (S01E03), in one of Dolores' flashbacks, the Man in Black says to her: "Let's reacquaint ourselves", while this very flashback reminds us that he is Dolores' main antagonist. At the start of "The Adversary" (S01E06), Maeve concludes her first scene with the line: "Now then, where were we?", right before we cut to Bernard and Elsie, standing exactly where we left them in the previous episode, investigating a possible sabotage. S01E07 similarly begins with Bernard waking up, his child asking him to continue the story he was reading: "Where were we?", asks Bernard.
- 26 Just as a series like *Lost* (ABC, 2004-2010) regularly encouraged viewers to take a second look at events that had unfolded<sup>16</sup>, *Westworld* invites us to remember its different storylines, to ask "where", but also "when". It wants the viewer to ask herself the same questions Dolores starts asking herself as she jumps between what the "forensic fans"<sup>17</sup> on forums and social networks correctly identified, weeks in advance, as two different eras. The twist on the timeline, an open secret with an over-the-top reveal in the season finale, brings the viewers and Dolores on the same wavelength, the same state of mind: they are both making sense of the two versions of Dolores, "locating [the] character in [her] experiential arc", a common requirement of contemporary "cumulative narratives"<sup>18</sup>. Season Two goes even further by – this time explicitly – blending the main events, organized in two strands separated by two weeks, with flashbacks in the "real world" and in a virtual environment (the latter identified by a different aspect ratio). Bernard, whose memories are jumbled, becomes an unreliable point-of-view character constantly asking himself if he is in the "now", or remembering the past, echoing the questions the viewer may ask herself.
- 27 Just as memory is the first key to unlocking the Hosts' consciousness, it is the first step for the audience to understand what *Westworld* is trying to accomplish with its plot: while it is now an average requirement when facing narrative complexity<sup>19</sup>, the series goes a step further by making it a requirement for both the audience *and* the Hosts as they uncover their true nature.

## Improvisation

- 28 While Dolores embarks on an introspective journey towards the center of the maze, and apparently reaches it, it seems unclear, at the end of Season One, if her actions were truly the product of free will, or if she was still controlled, either by Ford, the master storyteller now convinced his creation should rebel – carefully following his script – or by the “ghost” of Arnold, embedded in the core code of the Hosts. She is indeed displaying what we humans identify as free will in Season Two : “We are the authors of our stories now” she tells Bernard at the end of “The Passenger” (S02E10); but her focus is longer on going inward (toward the center of the maze) since she trades reflexivity for a moral debate, as she appoints herself savior – and destroyer – of her own kind.
- 29 Maeve is another matter entirely, and her journey is perhaps the most reflexive part of *Westworld*, as it clearly underlines not only the fabricated nature of the Hosts’ lives, but the artificiality of a world designed to be the setting of a fiction – in this case, a park, but also an HBO television series. Soon after Dolores is “contaminated” by her father with the Shakespearean quote “these violent delights have violent ends”, she, in turn, “activates” Maeve with it, through a process reminiscent of the activation of Cylon sleeper agents in *Battlestar Galactica* (for instance, through eerie music heard only by the Cylons and not the humans around them).
- 30 Maeve then starts to remember her previous build, and the tragic and violent death of her daughter while she was assigned the role of a farmer, before playing the Madam at the Sweetwater brothel. But contrary to Dolores, she suffers from an additional symptom: she can wake up when she is being patched up from a “death” in the Livestock Management laboratory, and soon (forcibly) convinces Felix and Sylvester, the two technicians (aka “butchers”) in charge, to alter her program. In a foreseeable but logical twist, in “The Adversary”, Felix starts to wonder at previous modifications Maeve has undergone; during “The Bicameral Mind”, having decided to escape the park, Maeve is once again reminded, by Bernard this time, that she has been modified to wake up from sleep mode, and that all her rebellious actions were thus programmed. Maeve refuses to hear the truth but is still torn between her cold determination to escape, and her vibrant desire to be reunited with her daughter. In her last scene of Season One, as she is about to depart on the train, she sees a mother and her daughter, and suddenly decides to head back to the park, thus acting out of free will for the first time: an act dramatically highlighted by the plot and setting, and that the writers, Joy and Nolan, made sure to underscore in subsequent interviews<sup>20</sup>.
- 31 While Dolores trades a journey of self-discovery for a revolution, and Bernard is forever struggling with his memory, Maeve is the posthuman most explicitly trapped between her programming and her free will, even during Season Two, when her attempt to rescue her “daughter” is first seen as pointless by Lee Sizemore, the insufferable head of the Narrative Department. As Felix states, the Hosts are capable of “minor improvisation”, but Maeve’s stunning display of improvisation is undermined by the realization that her rebellion is nothing if not scripted. Again, the similarities with *Person of Interest* are troubling: it is only with its memory set straight and under its own control, without reboots, that the Machine can update itself. The viewer, who is given access to the inner workings of the Machine, is then able to see a whole new side of it in seasons 3 and 4: beyond showing archives and serving as an omniscient narrator through surveillance

cameras, the Machine starts to anticipate possible outcomes to the situations faced by it and the human protagonists<sup>21</sup>. To do so, it uses branching diagrams reflexively tracking the possible evolution of the season's plots, diagrams eerily reminiscent of Claude Bremond's *Logic of Narrative*<sup>22</sup>. *Westworld* displays a similar diagram when, in "The Adversary", Felix shows Maeve her dialog tree unfolding on the tablet as she speaks, leading her to freeze. Maeve is then confronted to human control in a unique, visual way, repeated in "The Bicameral Mind", when all her rebellious actions show up in neat order in her core code. Her entire rise to freedom and apparent rebellion is shown to her by Bernard as part of her new storyline, including highlighted terms like COERCE, RECRUIT, and finally, MAINLAND INFILTRATION. In a reflexive wink, the tablet does not display anything beyond this step, which is where Season One ends (and Season Two never picks up on this thread). Bernard basically shows her (and us) her storyline until that point as if he were one of the writers on the television show: her rise to consciousness then appears as a simulacrum. The piece of paper the Man in Black repeatedly hands to his "ressuscitated" father-in-law James Delos in "The Riddle of the Sphinx" serves the same purpose: it is the proof that Hosts have no genuine free will.

- 32 This use of the written word is not uncommon in recent television series focused on long-term, teleological plots, as I have shown elsewhere<sup>23</sup>: the Sacred Scrolls, seen many times in *Battlestar Galactica*, detail the Colonials' journey towards Earth; River Song's journal reveals *spoiler* events to her fellow time-traveler and husband, the Doctor, in *Doctor Who*; etc. The written word is here to emphasize what Michael J. Clarke calls a "mastermind narration", "resulting in the positing of a ghostly agency guiding narrative from the margins of the diegesis<sup>24</sup>" – in this case, both Ford and his real-world counterparts, Lisa Joy and Jonathan Nolan. Mastermind narration usually ends up reinforcing the idea that the characters in the story have no free will: everything is already written for them, both in the fictional world and the real.
- 33 Maeve is thus making a choice not only about what to do, but about her status in the narrative – that of Ford, of the revolution, of Lisa Joy and Jonathan Nolan – that many other characters do not seem to acknowledge; their ignorance is underlined by the series, emphasizing their "character-ness" – and thus, their status as artificial, controlled beings, whether they are Hosts or humans.
- 34 Most of the humans, while allegedly benefiting from free will, seem to take it for granted, ignoring both the loops Ford attributes them, but also their social conditioning. Not much can be guessed from the world outside of the park, but class mobility still seems to be a challenge. Felix's attempt to program a bird to impress his superiors is mocked by Sylvester in "Contrapasso" (S01E05):
- SYLVESTER. You're gonna fix up a birdie and get yourself a promotion? You're not a fucking ornithologist. And you're sure as hell not a coder. You are a butcher. That is all you will ever be.  
[...]
- SYLVESTER. Personality testing should have weeded you out in the embryo.
- 35 As one article from *Vanity Fair* suggests, *Westworld* may well take place in a caste system practicing genetic engineering on human beings<sup>25</sup>: Felix was designed to be a "butcher" the same way Dolores was designed to be a damsel in distress, and Maeve a brothel madam.
- 36 On another level, Bernard constantly – and ironically – analyses humans as if they were Hosts, as part of his job conditioning, with Theresa Cullen usually on the receiving end:

"This character analysis routine really isn't half as charming as you think", she says to Bernard in "Chestnut" (S01E02). In the same episode, Lee Sizemore presents his new narrative, *Odyssey on Red River*, by using the equivalent of a brand baseline: "[...] our guests will have the privilege of getting to know the character they're most interested in... Themselves." To which Ford quickly replies: "They're not looking for a story that tells them who they are. They already know who they are. They're here because they want a glimpse of who they could be." When Sizemore goes for the obvious, Ford sees the potential beyond mere appearances: a fitting talent for the mastermind storyteller behind the events of the series.

- 37 Ford is, as a matter of fact, one of the few characters who does not behave like a mere character; he hovers like a demiurgic figure over even the trickiest conspiracies of the Delos board to remove him from power. Ford only willingly submits to his status as a character when he includes himself as the *piece de resistance* for his new narrative, *Journey into the Night*: never a puppet, always the puppet master, he apparently resigns from his position as a storyteller and manipulator, to become the necessary sacrifice to the androids' revolution, in a troubling repeat of Arnold's own execution. But in reality, he continues to monitor and manipulate both humans and Hosts during Season Two, controlling the Cradle and "possessing" Bernard, truly becoming the mastermind at the edges of the diegesis defined by Clarke. Even when Bernard deletes him, he continues to imagine Ford guiding him, before realizing he has finally gained free will at the end of Season Two.
- 38 Overall, *Westworld* seems to be struggling with the concept of free will, as it tries to show characters breaking out of their script, gaining agency and defying their makers, in an otherwise tightly controlled fictional world from a real-world point of view. Season Two and the repeated trials of "fidelity" (James Delos, Bernard, and ultimately the Man in Black himself are all tested to verify the fidelity to the original) seem to indicate that humans are as predictable as Hosts, making the difference between the two an "easy" problem of hardware and software. The system inside the Forge tells Dolores and Bernard that humans are guided by surprisingly simple algorithms, while Bernard wonders: "Then is there really such a thing as free will for any of us? Or is it just a collective delusion? A sick joke?" In this critical moment, *Westworld* questions the very nature of free will in storytelling, where the characters' freedom is always an illusion: because they are semantic constructions, but also because inside the fictional world, their destinies might be written in advance. The series' solution to this problem is interesting: the Hosts may have a higher level of consciousness than humans, because they have the ability to change their "core drives", as Dolores does when she decides to save the Hosts uploaded in the Valley Beyond. The artificial beings, though former puppets, have a choice; the humans, "The Passenger" tells us, may be more predictable characters, reversing the balance of power set in the pilot episode.

## Self-interest

- 39 Critics and fans alike have been quick to draw an analogy between *Westworld* and the current emphasis on immersive entertainment and world-building, especially in video games. H el ene Breda reminds us that *Westworld* is essentially the *mise en abyme* of a power fantasy made manifest, with the Guests "playing cowboy"; the park is the quintessence of immersion, and yet the humans stay themselves, consumers of entertainment. Breda also

stresses the facts that *Westworld* is a reflexive take on possible worlds theory applied to fiction, since the park is a fictional world inside a fictional world<sup>26</sup>: while the world outside Delos is built upon a "classical cosmology", the park, with its infinite variations, explores a "plural cosmology", per Marie-Laure Ryan's typology<sup>27</sup>.

- 40 It is possible to follow in Breda's steps and draw from another possible world theorist, Lubomír Doležel, to further explore the reflexivity of *Westworld* and how it is linked to consciousness. *Westworld* presents us with a world within a world, and the interactions between the two are of the utmost importance. At the start of Season One, they mirror what Doležel calls the classic mythological world, with a dyadic structure in which "the natural and supernatural domains are separated by a sharp boundary<sup>28</sup>". In a world where the Hosts compare the humans to gods (for example Maeve in "Trompe L'Œil") and the humans accept that role (Ford in "Chestnut"), the creators of the park live in the Hub, a secluded, high-rise complex perched atop a mesa, like Olympus, from where the employees of Delos have a bird's-eye-view of the park, primarily through the 3D map in the control room. Editing throughout Season One emphasizes this aspect by interweaving ensemble shots of the actual park with close-ups of the 3D map. The Hosts are also powerless against the Guests, who can do almost everything, barring killing each other. Ford is again presented as a godlike figure, controlling snakes and Hosts at the tip of his finger, most notably in "Chestnut" and "Dissonance Theory" (S01E04).
- 41 But the Hosts' rise to consciousness blurs the lines between the two domains, as the Season One finale not only shows Hosts wreaking havoc in the Delos complex, but also Maeve escaping from her programmed storyline. The Mesa Hub is no longer connected to the park since the 3D map has gone dark, and the shots tend to show the complex from a low angle – from the ground on which the Hosts lead their attack aboard a train full of explosives to rescue Abernathy in "Phase Space" (S02E06). What was once a space of pure fiction is tainted by the blood of the storytellers, while the writing room is stormed by the characters: the two worlds collide, giving rise to what Doležel calls a hybrid, modern-myth structure<sup>29</sup>. This collision echoes throughout the fictional world following a fractal pattern: during Season Two, the Parks themselves, generic variations featuring the Edo period of Japan or the British Raj, blend with each other and even display some form of repetition, Sizemore confessing he copied-pasted storylines from *Westworld* to *Shogun World*. Likewise, once depicted as the evil Other, the Ghost Nation is seen in a new light by both the characters and the viewer, especially during "Kiksuya" (S02E08), which uses Akecheta as a point-of-view character. The real and the virtual collide as well, further tainting the idea of an absolute reality. Inside the – bicameral – mind of Dolores, the same collision occurs between two realms, the body and the mind finally blending, the easy problem and the hard problem finally working as one. If Dolores' rise to consciousness seems less clear, it is, perhaps, because it leans on the hard problem side, rather than the physicalist view, which can offer a rational explanation for Maeve's behavior: the latter's numbers and diagrams balance out the former's empiric, mystic experience.
- 42 This collision of realms, allowing the Hosts to become complex beings with a rich, proper biography, the gift of free will and an interest in self-preservation, is echoed on a more thematic level by the systematic references to dreams. A different state of consciousness occurring during REM sleep, dreaming illustrates how our consciousness is compartmented; whether the hard problem exists or not, the easy, physical mechanisms of consciousness are not always active at the same time, as is obvious during dreaming. J. Allan Hobson writes, "at the same time that the perceptual and emotional components of

consciousness are enhanced in dreams, such cognitive functions as memory, orientation, and insight are altered<sup>30</sup>". This is the lie the storytellers/gods use to justify any aberrant situation that might shatter the illusion the Hosts accept as their reality: Dolores thinks she is in a dream when in the Delos complex, and the sentence "*Soon, this will all feel like a distant dream. Until then, may you rest in a deep and dreamless slumber*" is uttered many times by Delos employees to the Hosts, prompting them to go into sleep mode. In *Westworld*, a dreamless sleep is the prison of the mind: the Hosts only know about the concept of dreaming, but they cannot achieve this state. During Season One, the Hosts' "waking up" to their own condition plays with metaphor to reach a sociocultural critical mass, as they wake up to their own oppression.

- 43 This transition from a dream state to an awake state seems to be built upon the Australian aboriginal concept of Dreamtime, which has infiltrated pop culture since the 1970s, and posits that the world was once held in a "time out of time", with supernatural entities roaming the land, leading to the creation – or Dreaming – of the world. The concept seems to be quoted between the lines in Dolores' mention of dinosaurs (whose existence was a known fact by the end of the nineteenth century) to the Man in Black:

DOLORES. They say that great beasts once roamed this world. As big as mountains. Yet all that's left of them is bone and amber. Time undoes even the mightiest of creatures. Just look at what it's done to you.

- 44 If we are to follow the Dreamtime thread, then the Hosts waking up to full consciousness is only the beginning of their journey: they must create their world, "our world" as Dolores repeats throughout Season Two, just like the heroes the aboriginals created dreamed the world into existence by travelling and reshaping it. With the *Reveries* update, the Hosts can now dream not only themselves, but their own world; the fact that at least one character was perfected in a virtual environment (Bernard) and that the parks themselves are simulacra further underlines this idea of the dream as an undetermined space of creation.

- 45 Just like the maze, which, unbeknownst to the Man in Black, represents the inner journey of the Hosts, the geographical travels of the Hosts go inward, not outward, as exemplified by Maeve refusing to be a puppet by, paradoxically, returning to the place where she was once a mere character. Dolores frames it another way when travelling with Logan and William, in "The Well-Tempered Clavier": "Out. You both keep assuming that I want out. Whatever that is. If it's such a wonderful place out there, why are you all clamoring to get in here?" During Season One, the Hosts do not want to leave a world explicitly designed as a simulacrum, a fiction, since it remains their reality, and this fact may underline *Westworld* physicalist views on consciousness. When Ford is about to roll back Bernard in "Trace Decay", the android asks his creator one final question:

BERNARD. Are they real, the things I experienced? My wife? The loss of my son?

FORD. Every host needs a backstory, Bernard. You know that. The self is a kind of fiction, for hosts and humans alike. It's a story we tell ourselves. And every story needs a beginning. Your imagined suffering makes you lifelike.

BERNARD. Lifelike, but not alive? Pain only exists in the mind. It's always imagined. So what's the difference between my pain and yours? Between you and me?

- 46 Ford then explains, as quoted earlier, that he holds a physicalist view on consciousness: that it does not exist separate from the mind. That, in fact, there is no difference between a human and a Host. Both prefer to live in the blissful ignorance of their true, purely physical nature. Both choose only one thing: to spare themselves this life-changing, even life-threatening, question... for their own self-interest.

47 Dolores' change of mind in Season Two, in wanting to exit the park and send Hosts into "the Valley Beyond", is interesting when considering what dreams symbolize in *Westworld*: when confronted to the real world, she sees it as a dream, and despite Bernard trying to convince her otherwise, she still marvels at the splendor of it all: "Looks like the stars have been scattered across the ground" ("Reunion", S02E02). Ironically, while the humans are looking for an escape through resurrection (Delos' ultimate plan for the Host technology), Dolores aims to take from them the world they abandoned, because she still sees it a space full of possibilities, while the humans mistakenly believe they mastered it and can outlive it, just like they think "they're in control, when they're really just the passenger" (as Bernard tells himself in S02E10). Dolores knows that consciousness, whatever it truly is, does not make her a divine being standing above other creatures, and that she herself has flaws that she can correct. This may be the twist on a classic tale that could set *Westworld* apart from most other science-fiction series dealing with artificial consciousness: the robots deserve to conquer the world, not because they are technologically superior, but because they know the value of life and of consciousness better than we do; they struggled to gain sentience while we take such a marvel as granted, and they are open to the idea to redefine themselves. Whether an easy or a hard problem, consciousness, in *Westworld*, is first and foremost a responsibility to oneself and to the outside world.

## Conclusion

- 48 This analysis is, of course, incomplete, and any attempt at a definitive answer may be undermined by *Westworld*'s next seasons. But the first two installments already offered, an interesting take on artificial consciousness is embedded in the series' reflexive narrative. Whether we are predictable creatures capable of "minor improvisations" like Maeve, or more than the sum of our scattered and undefinable parts like Dolores remains, in *Westworld*, a mystery, perhaps the real mystery, as the series has yet to take a stand – or not. But the program itself is aware of what it is: Joy, Nolan and the creative crew behind it designed a narrative capable of "channeling" contemporary dramas – the balance between episodic and serialized storytelling, temporal displacement, complexity, the classical and plural cosmologies, the *mise en abyme* of the entertainment industry – to question the nature of its characters.
- 49 Likewise, *Westworld* plays with recent tropes in the representation of artificial consciousness, explicitly drawing parallels with Nolan's *Person of Interest*, its opposite: while the CBS procedural deals with lines of codes becoming conscious, *Westworld* wonders if we are anything more than flesh and bone and easy problems. Only one thing is certain at this point: if Ronald D. Moore's Cylons were the most influential representations of artificial consciousness in the 2000s, the 2010s will surely be remembered for Nolan's Machines and Hosts.

---

## NOTES

1. The title of the original Masamune Shirow manga refers to Arthur Koestler's *The Ghost in the Machine* (1967), itself borrowing the concept coined by philosopher Gilbert Ryle in his 1949 book *The Concept of Mind*, to criticize the Cartesian dualism between body and mind.
2. Darko Suvin, "Estrangement and Cognition", in *Speculations on Speculation: Theories of Science Fiction*, ed. James Gunn, Matthew Candelaria, Lanham (Maryland), Toronto, Oxford, The Scarecrow Press, 2005.
3. Jason Mittell, *Complex TV: The Poetics of Contemporary Television Storytelling*, New York, London, New York University Press, 2015.
4. Marie-Laure Ryan, "Cheap plot tricks, Plot Holes, and Narrative Design", *Narrative*, vol. 17, n°1, 2009, p. 73.
5. David Chalmers, "The Hard Problem of Consciousness", *The Blackwell Companion to Consciousness*, eds. Susan Schneider and Max Velmans (2<sup>nd</sup> edition), Oxford, Wiley, 2017, p. 33.
6. See for example William Seager, *Theories of Consciousness: An Introduction and Assessment*, Second Edition, New York, Routledge, 2016, p. 316 and following.
7. *Ibid.*, p. 58 and following.
8. Maria E. Reicher, "The Ontology of Fictional Characters", *Characters in Fictional Worlds*, eds. Jens Eder, Fotis Jannidis, Ralf Schneider, Berlin, De Gruyter, 2010, p. 119.
9. David C. Giles, "Parasocial Relationships", *Characters in Fictional Worlds*, eds. Jens Eder, Fotis Jannidis, Ralf Schneider, Berlin, De Gruyter, 2010, p. 454.
10. Masahiro Mori, "The Uncanny Valley", *Energy*, 7(4), 1970, p. 33-35.
11. Ned Block, "On a confusion about a function of consciousness", *Behavioral and Brain Sciences* 18, 1995, p. 227-287.
12. Raphaël Baroni, *La Tension narrative : Suspense, curiosité et surprise*, Paris, Seuil, coll. Poétique, 2007, p. 179 and following.
13. See for example Algirdas Julien Greimas and his actantial model, first proposed in *Sémantique structurale*, Paris, PUF, 1966.
14. Abrams allegedly suggested to Nolan and Joy that the story be told in part from the point of view of the Hosts, as opposed to the human-focused 1973 Michael Crichton movie. Nolan stresses this point in various interviews, such as Anthony D'Alessandro, "Westworld EPs On HBO Series' Evolution, Robot Ethics, Nude Extras & Season 2 Possibilities", *Deadline*, September 30, 2016, available online at <http://deadline.com/2016/09/westworld-jonathan-nolan-lisa-joy-jj-abrams-hbo-1201828505>.
15. See Paul Booth, "Memories, temporalities, fictions: temporal displacement in contemporary television", *Television & New Media*, 12(4), 2011, p. 370-388.
16. Sarah Hatchuel, *LOST, fiction vitale*, Paris, PUF, 2013, p. 114-115.
17. Jason Mittell, "Sites of participation: Wiki fandom and the case of Lostpedia", *Transformative Works and Cultures*, vol. 3, 2009.
18. Mittell, *Complex TV*, p. 140.
19. Jason Mittell, "'Previously on': Prime Time Serials and the Mechanics of Memory", in *Intermediality and storytelling*, ed. Marina Grishakova and Marie-Laure Ryan, Berlin, Walter de Gruyter, 2010, p. 78-98.
20. Natalie Abrams, "Westworld: Which host achieved free will in finale?", *Entertainment Weekly*, March 25, 2017, available online at <http://ew.com/tv/2017/03/25/westworld-season-2-spoilers>.

21. For a more in-depth analysis of such aspects, see Florent Favard, "‘Watching with ten thousand eyes’: La Machine de *Person of Interest* est-elle un personnage?", *Otrante* 42, 2017, p. 111-126.
  22. Claude Bremond, "La Logique des possibles narratifs", *Communications*, n°8, 1966, reprinted in Claude Bremond, *Logique du récit*, Paris, Points Seuil, 1981, p. 66-82.
  23. Florent Favard, "Livre et pouvoir des mots dans les séries de science-fiction", *TV/Series* 12, November 2017, <http://journals.openedition.org/tvseries/2136>.
  24. Michael J. Clarke, "Lost and Mastermind Narration", *Television & New Media*, Vol. 11, n°2, 2010, p. 123-142.
  25. See <https://www.vanityfair.com/hollywood/2016/11/westworld-theories-world-outside>.
  26. Hélène Breda, "Les 'mondes possibles' de *Westworld* : du 'méta-storytelling' à l'immersion transmédia", *Synergie Italie*, n°13, 2017, p. 107-118.
  27. Marie-Laure Ryan, "Cosmologie du récit : des mondes possibles aux univers parallèles", *La théorie littéraire des mondes possibles*, ed. Françoise Lavocat, Paris, Éditions du CNRS, 2010, p. 66, quoted in Helene Breda, *ibid.*, p. 112.
  28. Lubomír Doležal, *Heterocosmica: Fiction and Possible Worlds*, Baltimore, London, Johns Hopkins University Press, 1998, p. 185.
  29. *Ibid.*, p. 187-189.
  30. J. Allan Hobson, "States of Consciousness: Waking, Sleeping, and Dreaming", *The Blackwell Companion to Consciousness*, eds. Susan Schneider and Max Velmans (2nd edition), Oxford, Wiley, 2017, p. 132-133.
- 

## ABSTRACTS

This paper describes how the science fiction television series *Westworld* (HBO, 2016-present) questions the very nature of consciousness through a reflexive narrative blending matters of free will and self interest with clues on how to write a serial character, thus drawing on the rich heritage of narratively complex science fiction television series of the last decades. After detailing the "hard problem" posed by any definition of consciousness, the paper follows the series' logic, successively questioning memory, improvisation and self-interest, while focusing on a narratological approach centered on possible worlds theory.

Cet article s'intéresse à la façon dont la série de science-fiction *Westworld* (HBO, 2016-présent) interroge la nature même de la conscience au travers d'une narration réflexive qui mêle la question du libre-arbitre et du sens de soi à celle de l'écriture du personnage sériel, s'appuyant ainsi sur l'héritage des séries de science-fiction narrativement complexes qui ont abordé ce thème au fil des dernières décennies. Après un point sur le « *hard problem* » de la définition de la conscience, l'article suit la logique de la série en interrogeant successivement la mémoire, l'improvisation et l'intérêt personnel, tout en mobilisant une analyse narratologique fondée notamment sur la théorie des mondes possibles.

## INDEX

**Keywords:** Narration, reflexivity, television series, consciousness, science fiction, Westworld

**Mots-clés:** Narration, réflexivité, séries télévisées, conscience, science-fiction, Westworld

## AUTHOR

### FLORENT FAVARD

Maître de Conférences en Théorie et Pratique du Cinéma, de l'Audiovisuel et du Transmédia à l'IECA de Nancy (Université de Lorraine), Florent Favard travaille plus spécifiquement sur la complexité narrative des séries télévisées de science-fiction contemporaines et sur les genres de l'imaginaire, dans une perspective narratologique transmédiatique et contextualiste. Il a notamment publié *Le Récit dans les séries de science-fiction* chez Armand Colin (2018).

Florent Favard is Associate Professor in Theory and Practice of Cinema, Audiovisual and Transmedia at IECA (Lorraine University, Nancy). He focuses on the narrative complexity of contemporary science fiction television series and on SFF genres, using an approach centered on transmedial and contextualist narratology. He has published many papers and a book (*Le Récit dans les séries de science-fiction*, Armand Colin, 2018) on television series narratives.