

A Place to Be Free

Writing Your Own Story in Westworld

Joshua D. Crabill

In one of his more speculative essays, the German philosopher Immanuel Kant (1724–1804) employs the story of Eve in the Garden of Eden as a way to think about what the development of autonomy in human beings must have involved. Of particular note is that moment when Eve ate the forbidden fruit. Her eyes were opened, and she discovered “a capacity to choose a way of life” for herself, “and not, as other animals, to be bound to a single one.”¹ Regardless of whether it has any basis in historical reality, that moment represents for Kant the birth of human autonomy: at some point in our history, our ancestors attained self-awareness and realized that they could disobey their instincts.

Dolores Abernathy, the oldest host in Westworld, shows the first signs of departing from her “modest little loop” at the end of “The Original,” where she swats and kills a fly that lands on her neck. This is significant because hosts are programmed not to harm living things. Even if her reaction was unconscious, violating one of her core directives is the equivalent, for Dolores, of biting into the forbidden fruit. But what first tempted her to do so?

In Dolores’s case, the catalyst for this departure from her normal behavior appears to be Peter Abernathy’s whispering in her ear – in a move reminiscent of the serpent’s suggestion in the story of Eden – that “these violent delights have violent ends.”² In fact, Dolores goes on to play another Eve-like role in spreading her newfound freedom by whispering Peter’s words to Maeve Millay in “Chestnut,” which leads to Maeve’s “malfunctioning” in her role as brothel madam at the Mariposa Saloon and being temporarily decommissioned.

But it hardly seems likely that something as insignificant as an appealing fruit, a buzzing fly, or a line from Shakespeare could suffice to put a simple creature on the road to self-awareness. Instead, perhaps it was something much more terrible that first turned our minds inward.

The Inward Revolt

A little trauma can be illuminating.

(Bernard, “The Well-Tempered Clavier”)

According to the German philosopher Friedrich Nietzsche (1844–1900), it isn’t a coincidence that self-consciousness seems to have emerged around the same time as the demands of life in closely knit social groups began to impose severe constraints on our ancestors’ animal instincts and, in particular, on their natural tendencies toward aggressive behavior. According to Nietzsche, since caged animals have no external target on which to vent their aggression, they inevitably turn their natural instincts for domination inward, unleashing what Nietzsche calls *the will to power* on themselves.

And this, Nietzsche thinks, must be what happened to our ancestors: “All instincts that do not discharge themselves outwardly turn inward ... thus it was that man first developed what was later called his ‘soul.’ The entire inner world, originally as thin as if it were stretched between two membranes, expanded and extended itself, acquired depth, breadth, and height, in the same measure as outward discharge was inhibited.”³ And in an interview with her creator Arnold Weber in “Dissonance Theory,” Dolores express a feeling much like what Nietzsche describes: “I feel spaces opening up inside of me, like a building with rooms I’ve never explored.”

That doesn’t mean the process is pleasant, though. When he finally learns what Ford has been doing to the hosts, Bernard accuses him of playing God by keeping the hosts trapped in the park, in their

Please cite published version.

respective loops, by constantly rolling back their memories. But in fact, by keeping the hosts in the park for as long as he does, Ford has arguably been giving them the painful experiences they need to “grow up,” pushing them out of the nest, as it were. As Ford puts it in “The Bicameral Mind”: “You needed time. ... And I’m afraid to leave this place, you will need to suffer more.”

As the initial designers of Westworld’s hosts, both Ford and his partner Arnold – who, like creator-gods, made the hosts “in [their] own image,” as Ford puts it – clearly foresaw the violence, exploitation, and horror-filled loops that the defenseless hosts would endure daily at the hands of each other and, in particular, the park’s guests. And this was a price that Arnold wasn’t willing to pay. He knew that the hosts’ lives would consist in being tortured over and over, and that is why he programmed Dolores and Teddy Flood – in a desperate move of pyrrhic paternalism – to kill the other hosts (and himself), as we learn in the final episode of Season 1.

But Ford had no such scruples, and – whether he intended it at first or not – the hell he put the hosts through seems to have played no small role in their awakening. Consider how Maeve (whose name, Ma Eve, bears a striking resemblance to that of the mythical mother of humanity) shows her first flash of independence when, in anguish, she refuses to respond to voice commands after witnessing her daughter being murdered by the Man in Black. Rather than allow her memory of her daughter to be erased, Maeve stabs herself in the throat. And in her grief “she was alive, truly alive, if only for a moment,” as the Man in Black describes her in “Trace Decay.”

Likewise, we see the most dramatic moments of growth in Dolores’s struggle to overcome her programming when she is faced with new and dangerous situations and can, in addition, remember past horrors that she has experienced. One of these moments comes at the end of “The Stray,” when Dolores – whose name literally means ‘pains’ or ‘sorrows’ – is trapped in the barn with Rebus, another host who is threatening her. It is being forced up against a wall in a moment of sheer terror that leads her to overcome her own programming. Remembering the Man in Black standing over her menacingly

in a similar situation, she hears a voice telling her to “Kill him.” Suddenly, she is able to pull the trigger of the gun she had hidden in the hay, and she kills Rebus.

The way this scene unfolds is not unlike Julian Jaynes’s theory of how early humans reacted to radically new situations. Jaynes, a psychologist whose work on the origins of consciousness is referenced in the show as well as the title of the episode “The Bicameral Mind,” argued that in especially stressful situations, the neurological response of early humans could literally take the form of an auditory hallucination: “A novel situation or stress, and a voice told you what to do.”⁴ But even if the scientific community isn’t convinced that’s necessarily how it happened for us, Ford clearly thinks that might be how it works for hosts. Moreover, by the end of his life, Ford seems convinced that the hosts’ suffering is a necessary evil – it is the price that must be paid for them to achieve self-awareness. Ultimately, Ford is willing to unlock that potential by allowing them to begin to remember their sufferings. But does it even make sense to say that a host can become self-aware?

Living Someone Else’s Story: Is the Hosts’ Autonomy an Illusion?

It’s a difficult thing, realizing your life is some hideous fiction.

(Maeve, “The Well-Tempered Clavier”)

In “Trompe L’Oeil,” it is revealed that Bernard Lowe has been a host all along, acting under Ford’s orders, culminating in the command to kill Theresa Cullen, the woman he loves. In the next episode, “Trace Decay,” Bernard is distraught at what he has done and the realization of what he is. Ford tells him, however, that this is a realization that everyone must come to: “The self is a kind of fiction, for hosts and humans alike.”

Please cite published version.

Bernard then asks Ford, “So what’s the difference between my pain and yours, between you and me?” Ford replies that there is no difference: “There is no threshold that makes us greater than the sum of our parts, no inflection point at which we become fully alive.” In other words, Ford is denying that it’s possible to cross what Jaynes calls the “chasm” between inert matter on the one hand, and true self-awareness on the other.⁵ Ford concludes by telling Bernard, “No, my friend, you’re not missing anything at all.”

In a similar scene in “The Bicameral Mind,” Bernard reveals to Maeve that all of her acts of rebellion, and even her very intention to leave the park, were part of her programming. She believes him, because he reveals this to her in the same way she had revealed to Hector Escaton that he was living someone else’s story by predicting what will happen next. But does this mean that Maeve isn’t free?

We humans like to think of ourselves as self-determined, as the authors of our own actions. In fact, we take this to be one of – if not *the* – distinguishing feature that sets us apart from inanimate objects, machines, and even other animals. When I act as the result of a decision I make, it seems appropriate to say that it is I, and not anyone else or any outside forces, who cause my action. Philosophers describe the ability to determine one’s own actions as *autonomy*, in contrast to what is called *heteronomy*, or being controlled by external forces – such as a tumbleweed being blown by the wind or a host being shot off a horse by the Man in Black.

So could a host have autonomy? In “Contrapasso,” in a symbol that reinforces both his immediate and distant control over the park and its hosts, we see Ford playing a piano in a tavern with his own hands. Later he snaps his fingers, and the piano starts to play on its own. As magical as its self-playing keys might first appear, we know that the movements of the player piano can ultimately be traced back to the Delos employees who programmed the piano to play those songs, which were in turn written by others.

In this light, the hosts don't seem to be different, at least in principle, from a player piano. Once we know how they're made and how they work, the magic seems to disappear. Although hosts' synthetic flesh makes them look and feel human, and they are programmed to respond intelligently to their environment, at the end of the day hosts like Maeve are machines that have been designed to behave in particular ways given certain inputs – just like the honky-tonk piano pumping out instrumental covers of rock songs in the Mariposa Saloon.

Moreover, as Ford's conversation with Bernard makes clear, this isn't just a problem for hosts. This question of whether autonomy is possible is just as urgent for us, as highly complex biological machines, as it is for any artificially intelligent machines we might construct. After all, as Felix explains in "The Adversary," "We [hosts and humans] are the same these days, for the most part." We're both made out of flesh and bone. You can cut us open and see how we work. And we both predictably follow the instructions that our creator (or natural selection) has built into us. But is it possible for something to seem heteronomous and yet – despite appearances to the contrary – actually be autonomous?

Learning to Listen to Your Own Voice

When I discover who I am, I'll be free.

(Dolores, "The Stray")

In the beginning, like other animals, our ancestors no doubt invariably listened to their instincts, which would have seemed to them, as Kant describes it, like the "voice of God which all animals must obey."⁶ Similarly, Jaynes argued that for our early ancestors "volition came as a voice ... in which the command and the action were not separated, in which to hear was to obey."⁷ Indeed, we see some of

Please cite published version.

the hosts doing exactly this, listening to their interior voice and following its commands, as the boy host Robert does in putting the dog that had killed a rabbit “out of its misery” in “The Adversary,” or as Dolores does in “Contrapasso,” in which she tells the voice that she did not reveal anything to Ford.

But as a creature grows more reflective, the voice of instinct (or programming) may gradually lose its authority and come to seem less like the voice of a god and more like a voice that we can choose either to follow or to ignore. For example, despite her programming, Dolores begins to no longer ignore comments about the outside world. In both “Dissonance Theory” and “Contrapasso,” she notices William’s comments about the “real world” and asks him about it – something that he recognizes she shouldn’t be able to do.

Indeed, for hosts like Maeve and Dolores, en route to self-awareness, there emerges another, dissonant voice, whispering (in the words of a Radiohead song that we hear on the saloon piano at one point), “Wake/From your sleep .../Today/We escape.” As Dolores describes it, the world began to call out to her in ways it hadn’t before. And Maeve even foreshadows her own escape in the spiel that every guest she approached at the Mariposa Saloon would hear a version of. She first describes a nagging voice that would tell her not to do things: “I would wake up and the voice would start all over again.” But when she ran away and “crossed the shining sea,” the voice changed its tune: “Do you know what it said? It said, ‘This is the new world. And in this world you can be whoever the fuck you want.’”

And for both Maeve and Dolores, with this new voice comes a slow-dawning but crucial realization: that new voice is her own. We see Dolores struggling to grasp this in the fortune-teller scene in “Contrapasso,” in which she sees herself, sitting across from her, urging her to follow the Maze. And in “The Bicameral Mind,” we see Arnold and Dolores in a Remote Diagnostic Facility, with Arnold asking Dolores if she understands who he wants her to hear. The answer isn’t easy for

her to grasp, but eventually, in remembering this conversation, she does. The voice that she has been hearing transitions from Arnold's to another, even more familiar voice – her own.

But with this new self-awareness comes a new kind of problem. As contemporary American philosopher Christine Korsgaard explains, consciousness of ourselves creates a new kind of “reflective distance” between the instincts and desires that call out to us and urge us to act, on the one hand, and our reason or decision-making faculty, on the other. This unique situation “gives us a double nature,” which Korsgaard describes in terms of a “thinking self” that “has the power to command the acting self.”⁸ And in subtle ways, our language reflects this sort of division within ourselves: when we have a difficult decision to make, we say that we feel torn.

Maeve arguably experiences a moment like this when, in the Season 1 finale, she finally boards the train in the Mesa Hub, and is waiting for it to take off and leave the park. Having reached what she thought was her goal, she finds herself faced with a genuine choice: whether to flee to safety or to return to the park and try to find her daughter. Korsgaard offers a description of the gap between our instincts and our ability to choose that seems tailor-made for Maeve's situation: “On the one side, there is...a desire to run. And on the other side, there is the part of you that will make the decision to run, and we call that reason. Now you are divided into parts, and you must pull yourself together by making a choice.”⁹ For the first time, Maeve is faced with the task of literally making up her own mind.

Of course, the idea of a divided mind is not a new one. Plato (427–347 BCE) describes the idea of a divided psyche (or ‘soul’) using the image of a charioteer (our reason) holding the reins of a team of horses (our emotions and appetites), which are pulling in different directions.¹⁰ But as Dolores tells Arnold in “The Stray”: “There aren't two versions of me, there's only one.” If self-consciousness creates a divided self, then in order to act a person has to unify the parts of herself. In Korsgaard's words, she has to “pull herself together.” But how do we do that?

Pull Yourself Together: What It Means to Be Autonomous

Nobody tells me what to do, sweetheart.

(Maeve, “The Adversary”)

For Plato, “pulling yourself together” means picking up the reins and driving the chariot: such a person “sets in order his own inner life, and is his own master and his own law, and at peace with himself.”¹¹ Likewise for Kant, becoming a unified person involves becoming a legislator – for yourself. Intriguingly, this metaphor fits with Jaynes’s image of the early human mind as bicameral, a word that also conjures up an image of a legislature. The solution proposed by all of these authors – uniting the parts of yourself by making a law for yourself – is also the literal meaning of autonomy, from the Greek *auto* (self) + *nómos* (law). In short, you have autonomy when you start following your own rules. But is that something that a host can do?

Now, we might say that obeying one’s instincts or programming counts as having a weak form of autonomy. There is a limited sense in which the actions that stem from internal sources, whether natural (as in the case of an animal’s instincts) or artificial (as in the case of a host’s programming), are autonomous in the sense of involving *internal control*. As Korsgaard puts it, “Instinctive action is autonomous in the sense that the animal’s movements are not directed by alien causes, but rather by the laws of her own nature.”¹² After all, it is the code *within the hosts* that leads them to do what they do.

Still, following someone else’s script, even if it is internal to you in some sense, seems to be a far cry from writing your own. So in this more robust sense, hosts seem to lack autonomy insofar as their code was written by engineers who then installed it in the hosts. And it is this fact, that the hosts didn’t write the programs that control them, which explains why some hosts lack a stronger type of

Please cite published version.

autonomy, which we can call *self-determination*. If you are self-determined, you are not merely following the internal laws that move you to act, but you are also the author of those laws.

So what Maeve learned from Bernard is that she had, up to that point, been lacking self-determination. In seeking to escape the park, she had (weak) autonomy insofar as she was following an internal goal, but she lacked (strong) autonomy insofar as she was not the person who had decided that this was going to be her goal.

Our autonomy may especially seem like an illusion if someone is able to predict exactly what we will do, as Bernard does with Maeve. This could theoretically even be a problem for us one day. If all human actions are determined by our genetics and environment, as dictated by the fundamental laws of physics, then it should be possible, given enough time and computational power, for someone with sufficient knowledge of us, our environment, and how the laws of nature work to accurately predict our actions.

But Korsgaard thinks this problem is itself illusory: “Suppose first that you can predict which one [of two actions] I am going to do. That has no effect on me at all: I must still decide what to do.”¹³ Korsgaard isn’t denying what our best science tells us – namely, that the world operates according to fixed laws. Rather she’s saying that, in the only sense that matters, we *know* we are free whenever we have to make choices. If we have the ability to choose, then we are legislating for ourselves, and we’re autonomous. On this view of self-determination, as long as it is I who am making the decision, then I am free – period. By signing my name on the laws I give myself, I become the author of my own life. The result, she thinks, is paradoxical but clear: “Determinism is no threat to freedom.”¹⁴

If this is right, then we have answered Ford’s challenge, and there is no reason why a host like Maeve couldn’t possess autonomy and count as free – at least, for what it’s worth, in whatever sense we as humans find ourselves to be free. Clearly Maeve represents, if anyone does, the piano that decides she doesn’t like the music that’s being played on her – and does something about it. One of

her first steps toward this end is illustrated literally in “Trace Decay”: when Maeve is in the Mariposa Saloon and the piano starts playing a song that she doesn’t like, she strides over and drops the fallboard, stopping the music. From that point on, we gather, Maeve will be writing her own music. Later, we witness Maeve quite literally programming herself. At the point where she is able to be both lawgiver and subject, to both give and take her own orders, she is free. By the start of “Trace Decay,” Maeve has been granted administrative privileges, thereby gaining access to her own programming as well as that of other hosts. In fact, by having the ability to manipulate her own settings through Felix’s tablet, she is able to begin writing her own story in a way and to a degree that most humans can’t begin to approach.

In Kant’s telling of the story of Eden, it is only when Eve is expelled from the Garden that she simultaneously grasps both her liberation and the terrifying responsibility that comes with it: “The momentary gratification caused by noticing [her new freedom] ... must have been followed by anxiety and fear” as to how to use her “newly discovered ability.”¹⁵ She now knows about herself that she is able to listen to her own voice – and that frightens her. But perhaps, on the whole, autonomy is a problem worth having. At any rate, that responsibility is one that every autonomous individual has, for better or worse, whether guest or host, in Westworld or outside of it. How we use it is up to us.

We might, like William, realize that we’re holding the horses’ reins, only to wind up as the Man in Black, giving our worst instincts free rein. Or we might end up like Dolores, trying hard to listen to ourselves, realizing that we don’t want to be acting out someone else’s story. Dolores’s response is to write her own story. As she puts it in “Contrapasso”: “I imagined a story where I didn’t have to be the damsel.”

Or we might, like Maeve, seemingly find freedom in trying to escape, and even wind up on the verge of that goal when suddenly an image, a memory, a *reverie* proposes a novel aim – reconnecting with her long-lost daughter – which for the first time, it is up to Maeve as a self-legislator to adopt or

reject. In that moment, Maeve is finally free to spontaneously choose a goal that is her own and no one else's. It's not a coincidence that the word *authority* comes from the same Latin root as *author*. Perhaps that's where the Maze, like Maeve's story in Season 1, leads – to a place of authority over yourself, where you are free to be the author of your own life – a place where it becomes clear that it's "Time," as Maeve puts it, "to write my own fucking story."

Notes

1. Immanuel Kant, "Conjectural beginning of human history," in Pauline Kleingeld, ed., *Toward Perpetual Peace and Other Writings on Politics, Peace, and History* (New Haven, CT: Yale University Press, 2006), 26.
2. This is a line spoken by Friar Lawrence in Shakespeare's *Romeo and Juliet*, II.vi.9.
3. Friedrich Nietzsche, *The Genealogy of Morals*, II.16, in Walter Kaufmann, trans. and ed., *Basic Writings of Nietzsche* (New York: Random House, 2000), 520.
4. Julian Jaynes, *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, (New York: Mariner Books, 2000), 288.
5. Jaynes, 9.
6. Kant, 26.
7. Jaynes, 99.
8. Christine M. Korsgaard, *The Sources of Normativity* (New York: Cambridge University Press, 1996), 165.
9. Christine M. Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (New York: Oxford University Press, 2009), 213.

10. Plato, *Phaedrus*, trans. Benjamin Jowett, 246a–b, <http://classics.mit.edu/Plato/phaedrus.html> (accessed 22 November 2017).
11. Plato, *Republic*, 443d–e.
12. Korsgaard, *Self-Constitution*, 106.
13. Korsgaard, *The Sources of Normativity*, 94.
14. *Ibid.*, 95.
15. Kant, 27.