The New York Times

THE STONE

It's Westworld. What's Wrong With Cruelty to Robots?

By Paul Bloom and Sam Harris
Paul Bloom is a psychologist. Sam Harris is a neuroscientist.

April 23, 2018

Suppose we had robots perfectly identical to men, women and children and we were permitted by law to interact with them in any way we pleased. How would you treat them?

That is the premise of "Westworld," the popular HBO series that opened its second season Sunday night. And, plot twists of Season 2 aside, it raises a fundamental ethical question we humans in the not-so-distant future are likely to face.

Based on the 1973 film, "Westworld" depicts a futuristic playground modeled after the Wild West, where the characters — bartenders, prostitutes, sheriffs, bandits — are robotic "hosts," programmed to interact as naturally as possible with their human guests. These intelligent machines look and act exactly like people. Indeed, viewers are often confused or misled about who is a host and who is a person.

The guests can behave however they please. Some assume heroic roles, but others choose to act out their darkest impulses, participating in torture, rape and murder — including the murder of robots that are indistinguishable from human children. The hosts have been designed so that they can't harm the guests; so these are acts of pure sadism, without risk of reprisal.

It's hardly a spoiler to say that things go wrong for the humans in "Westworld." But

we are interested here in the show's premise — and what our reaction as viewers to these lifelike robots suggests about human nature and the future of technology.

The biggest concern is that we might one day create conscious machines: sentient beings with beliefs, desires and, most morally pressing, the capacity to suffer. Nothing seems to be stopping us from doing this. Philosophers and scientists remain uncertain about how consciousness emerges from the material world, but few doubt that it does. This suggests that the creation of conscious machines is possible.

Suppose, as many philosophers and scientists believe, that consciousness arises in a sufficiently complex system that processes information. There is no reason to think that such a system need be made of meat. Conscious minds are most likely platform-independent — ultimately the product of the right software. It seems only a matter of time before we either emulate the workings of the human brain in our computers or build conscious minds of another sort.

And even if one believes that only biological systems can be conscious, this development, too, lies within our reach — perhaps through a combination of artificial intelligence and genetic engineering. Indeed, in "Westworld," it's hinted that the hosts are partly biological. We certainly see them bleed.

If we did create conscious beings, conventional morality tells us that it would be wrong to harm them — precisely to the degree that they *are* conscious, and can suffer or be deprived of happiness. Just as it would be wrong to breed animals for the sake of torturing them, or to have children only to enslave them, it would be wrong to mistreat the conscious machines of the future.

But how will we know if our machines become conscious? Descartes argued that one's own consciousness is beyond any possibility of doubt. In the case of others, we're never absolutely sure. Many of us have entertained, if only for a moment, the idea that everyone else might be a zombie: laughing, crying, complaining, rejoicing, but with no one home. Perhaps scientists will eventually discover the signature of consciousness, and then we will be able to test for it in our robots, as well as in animals and one another. But it is certain that we will build machines that *seem*

conscious long before we arrive at that point.

Anything that looks and acts like the hosts on "Westworld" will appear conscious to us, whether or not we understand how consciousness emerges in physical systems. Indeed, experiments with AI and robotics have already shown how quick we are to attribute feelings to machines that look and behave like independent agents. Multiply that temptation to anthropomorphize a thousandfold: Think not of a machine with visible wires, cartoonish eyes and a voice that sounds like Siri but of a beautiful stranger who engages you in intelligent conversation and who may be more aware of your emotions than your spouse or best friends ever were. It would be irresistible to see this creature as a person, whatever your philosophical views and regardless of what its creators told you about how it was built.

This is where actually watching "Westworld" matters. The pleasure of entertainment aside, the makers of the series have produced a powerful work of philosophy. It's one thing to sit in a seminar and argue about what it would mean, morally, if robots were conscious. It's quite another to witness the torments of such creatures, as portrayed by actors such as Evan Rachel Wood and Thandie Newton. You may still raise the question intellectually, but in your heart and your gut, you already know the answer.

Watching the show, you also discover how you feel about the people who rape, torture and kill these robots. We have no idea how many people would actually behave this way in a place like Westworld (the show implies that there's no shortage of such customers), but there is something repugnant about those who do. In this scenario, the robot hosts are the most human and the humans who abuse them are monsters.

Kant had odd views about animals, seeing them as mere things, devoid of moral value, but he insisted on their proper treatment because of the implications for how we treat one another: "For he who is cruel to animals becomes hard also in his dealings with men." We could surely say the same for the treatment of lifelike robots. Even if we could be certain that they weren't conscious and couldn't really suffer, their torture would very likely harm the torturer and, ultimately, the other people in his life.

This may seem like an extreme version of the worry that many have about violent video games. It has long been speculated that enacting violence in a virtual world desensitizes people to violence in the real one. The evidence for such an effect turns out to be weak. In fact, as video games have become increasingly realistic, the rate of violent crime has dropped.

But the prospect of building a place like "Westworld" is much more troubling, because the experience of harming a host isn't merely similar to that of harming a person; it's identical. We have no idea what repeatedly indulging such fantasies would do to us, ethically or psychologically — but there seems little reason to think that it would be good.

The issues here extend beyond sadism. Machines are created to improve the lives of human beings, and one of the attractions of advanced AI is the prospect of robot maids, butlers and chauffeurs (also known as self-driving cars). This is all fine with the sorts of machines we currently have, but as AI improves, we run a moral risk.

After all, if we do manage to create machines as smart as or smarter than we are — and, more important, machines that can feel — it's hardly clear that it would be ethical for us to use them to do our bidding, even if they were programmed to enjoy such drudgery. The notion of genetically engineering a race of willing slaves is a standard trope of science fiction, wherein humankind is revealed to have done something terrible. Why would the production of sentient robot slaves be any different?

For the first time in our history, then, we run the risk of building machines that only monsters would use as they please.

Paul Bloom is a professor of psychology at Yale and the author of "Against Empathy: The Case for Rational Compassion." Sam Harris is a neuroscientist, the author of "Waking Up: A Guide to Spirituality Without Religion," and the host of the "Waking Up Podcast."

Now in print: "Modern Ethics in 77 Arguments," and "The Stone Reader: Modern Philosophy in 133 Arguments," with essays from the series, edited by Peter Catapano and Simon Critchley, published by Liveright Books.

Follow The New York Times Opinion section on Facebook and Twitter, and sign up for the Opinion Today newsletter.