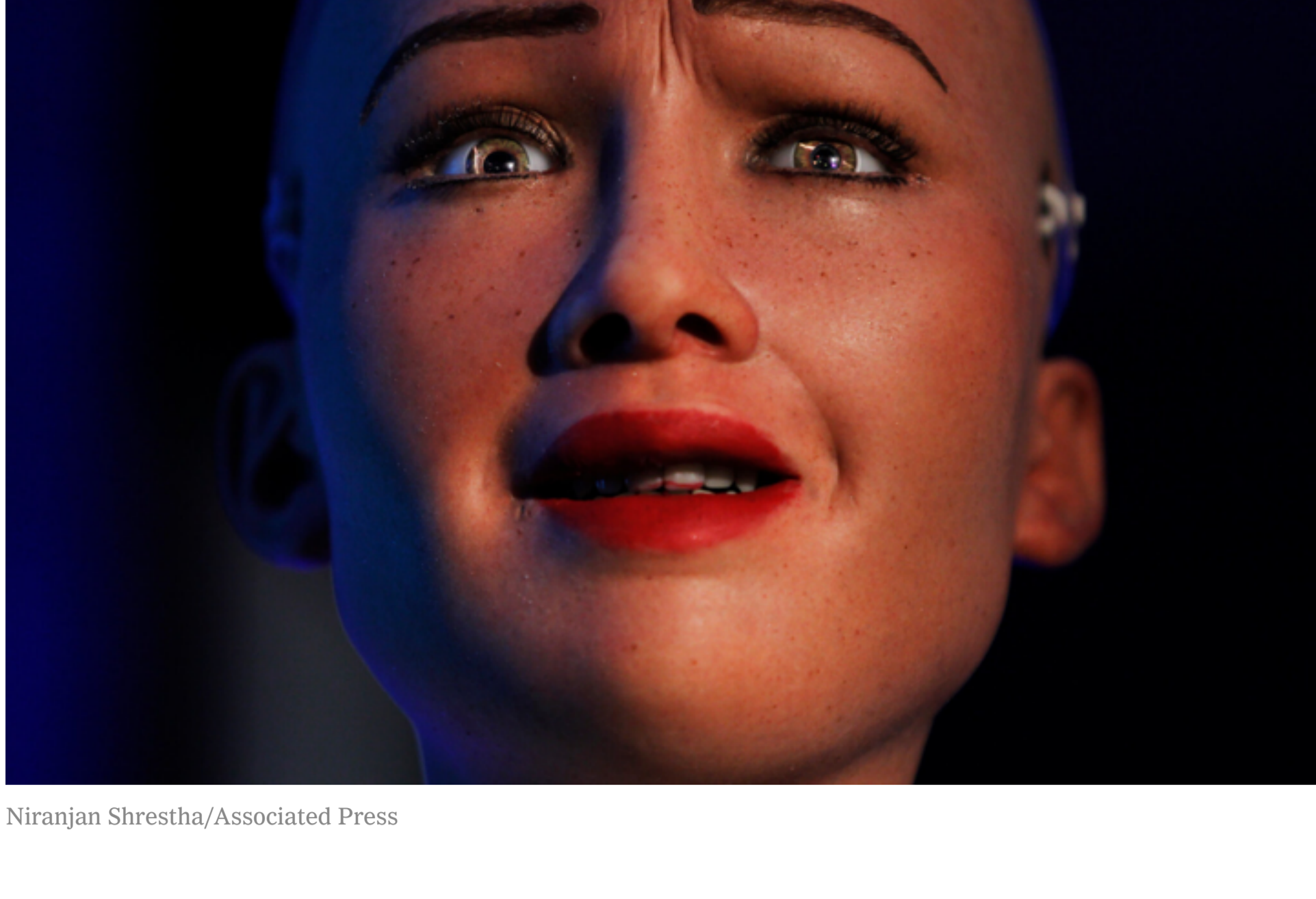


SCIENCE TIMES

Why A.I. Should Be Afraid of Us

Because benevolent bots are suckers. Plus, racism in medical journals, the sperm-count “crisis” and more in the Friday edition of the Science Times newsletter.



Niranjan Shrestha/Associated Press

By Alan Burdick

June 5, 2021

Artificial intelligence is gradually catching up to ours. A.I. algorithms can now consistently beat us at [chess](#), [poker](#) and [multiplayer video games](#), [generate images of human faces indistinguishable from real ones](#), [write news articles](#) (not this one!) and [even love stories](#), and drive cars better than most teenagers do.

But A.I. isn't perfect, yet, if Woebot is any indicator. Woebot, as Karen Brown wrote this week in Science Times, [is an A.I.-powered smartphone app that aims to provide low-cost counseling](#), using dialogue to guide users through the basic techniques of cognitive-behavioral therapy. But many psychologists doubt whether an A.I. algorithm can ever express the kind of empathy required to make interpersonal therapy work.

“These apps really shortchange the essential ingredient that — mounds of evidence show — is what helps in therapy, which is the therapeutic relationship,” Linda Michaels, a Chicago-based therapist who is co-chair of the Psychotherapy Action Network, a professional group, told The Times.

Empathy, of course, is a two-way street, and we humans don't exhibit a whole lot more of it for bots than bots do for us. Numerous studies have found that when people are placed in a situation where they can cooperate with a benevolent A.I., they are less likely to do so than if the bot were an actual person.



Curiosity and Innovation Can Create a More Sustainable Future

Learn how Microsoft and SAS are helping innovators address challenges like fighting deforestation in the Amazon, and studying snow leopard populations.

LEARN MORE

“There seems to be something missing regarding reciprocity,” Ophelia Deroy, a philosopher at Ludwig Maximilian University, in Munich, told me. “We basically would treat a perfect stranger better than A.I.”

In a [recent study](#), Dr. Deroy and her neuroscientist colleagues set out to understand why that is. The researchers paired human subjects with unseen partners, sometimes human and sometimes A.I.; each pair then played a series of classic economic games — Trust, Prisoner's Dilemma, Chicken and Stag Hunt, as well as one they created called Reciprocity — designed to gauge and reward cooperativeness.

Thanks for reading The Times.
Subscribe to The Times

Our lack of reciprocity toward A.I. is commonly assumed to reflect a lack of trust. It's hyper-rational and unfeeling, after all, surely just out for itself, unlikely to cooperate, so why should we? Dr. Deroy and her colleagues reached a different and perhaps less comforting conclusion. Their study found that people were less likely to cooperate with a bot even when the bot was keen to cooperate. It's not that we don't trust the bot, it's that we do: The bot is guaranteed benevolent, a capital-S sucker, so we exploit it.

That conclusion was borne out by conversations afterward with the study's participants. “Not only did they tend to not reciprocate the cooperative intentions of the artificial agents,” Dr. Deroy said, “but when they basically betrayed the trust of the bot, they didn't report guilt, whereas with humans they did.” She added, “You can just ignore the bot and there is no feeling that you have broken any mutual obligation.”

This could have real-world implications. When we think about A.I., we tend to think about the Alexas and Siris of our future world, with whom we might form some sort of faux-intimate relationship. But most of our interactions will be one-time, often wordless encounters. Imagine driving on the highway, and a car wants to merge in front of you. If you notice that the car is driverless, you'll be far less likely to let it in. And if the A.I. doesn't account for your bad behavior, an accident could ensue.

Editors' Picks

- [As M.L.B. Honors Lou Gehrig, It Shines a Spotlight on A.I.S.](#)
- [The Sperm-Count 'Crisis' Doesn't Add Up](#)
- [Is Chinese Elephants Are on a Long March North. Why, No One Knows.](#)

The New York Times | SERIAL
Nice White Parents

The powerful force shaping the public school down the block.
A new podcast, streaming now.

LEARN MORE



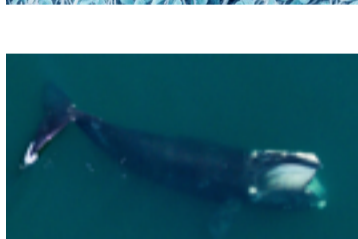
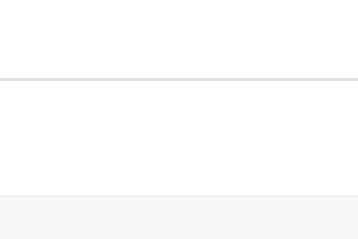
“What sustains cooperation in society at any scale is the establishment of certain norms,” Dr. Deroy said. “The social function of guilt is exactly to make people follow social norms that lead them to make compromises, to cooperate with others. And we have not evolved to have social or moral norms for non-sentient creatures and bots.”

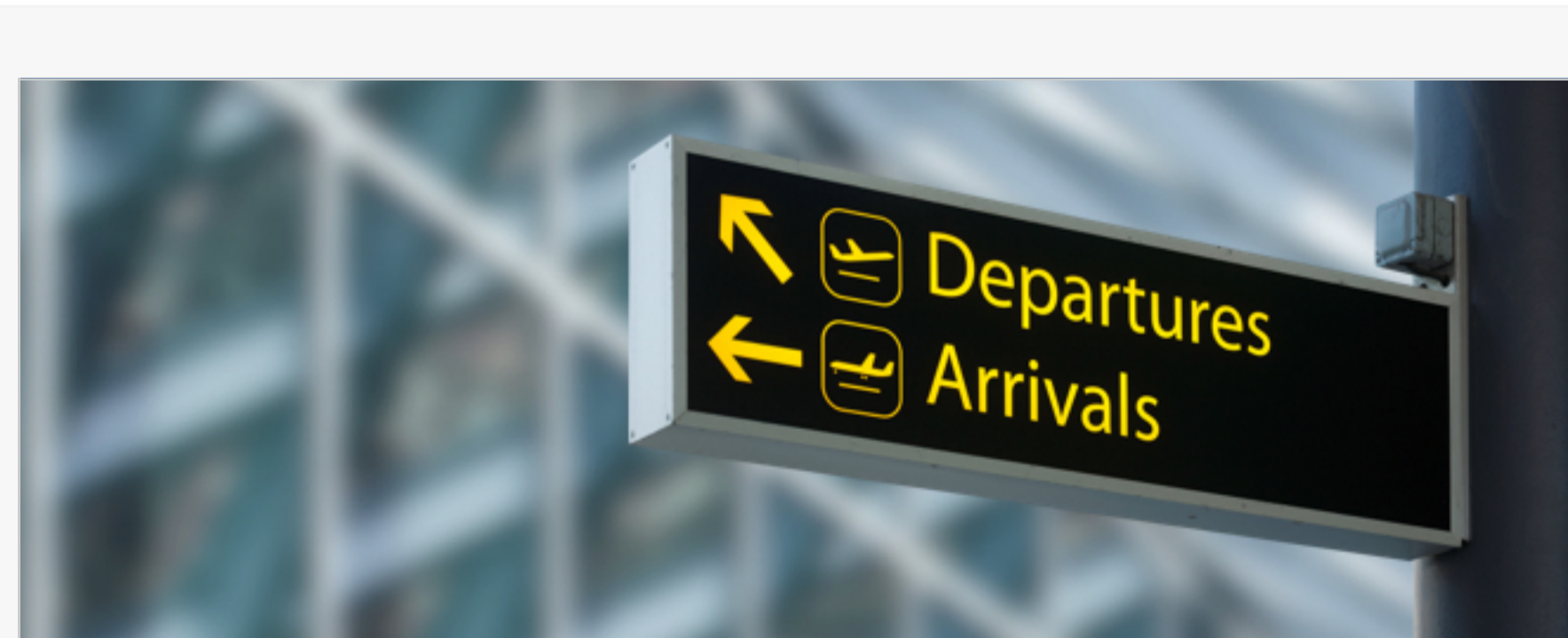
That, of course, is half the premise of “Westworld.” (To my surprise Dr. Deroy had not heard of the HBO series.) But a landscape free of guilt could have consequences, she noted: “We are creatures of habit. So what guarantees that the behavior that gets repeated, and where you show less politeness, less moral obligation, less cooperativeness, will not color and contaminate the rest of your behavior when you interact with another human?”

There are similar consequences for A.I., too. “If people treat them badly, they're programed to learn from what they experience,” she said. “An A.I. that was put on the road and programmed to be benevolent should start to be not that kind to humans, because otherwise it will be stuck in traffic forever.” (That's the other half of the premise of “Westworld,” basically.)

There we have it: The true Turing test is road rage. When a self-driving car starts honking wildly from behind because you cut it off, you'll know that humanity has reached the pinnacle of achievement. By then, hopefully, A.I. therapy will be sophisticated enough to help driverless cars solve their anger-management issues.

Health and Science Highlights

-  **“I had never faced the reality of death”**
June 3, 2021
-  **The influential JAMA journals promise changes in staff diversity and inclusivity**
June 2, 2021
-  **Why the male-fertility “crisis” doesn't add up**
June 4, 2021
-  **On the verge of extinction, these whales are also shrinking**
June 3, 2021



BARCLAYS

What parts of the old normal will be part of the new normal?

If you're thinking about it, we're analysing it.

GAIN INSIGHTS

What we're metabolizing lately

- “[The Age of Reopening Anxiety](#),” by Anna Russell in The New Yorker, explores the experience that so many of us are having these days.
- This fun interview with the mathematician Jordan Ellenberg, in The Atlantic, examines [why so many pandemic predictions failed](#).
- If luscious drone video of a volcano erupting in Iceland is the sort of thing you find relaxing, here's a [whole lotta lava](#).
- Less relaxing: this Washington Post article goes a long way toward deciphering [the appeal of QAnon by noting its similarity to a video game](#) (and an unwinnable one).
- And, um, in case you need it, here's a clear explanation of [why getting a Covid test will not make your forehead magnetic](#).

Science in The Times, 58 years ago today



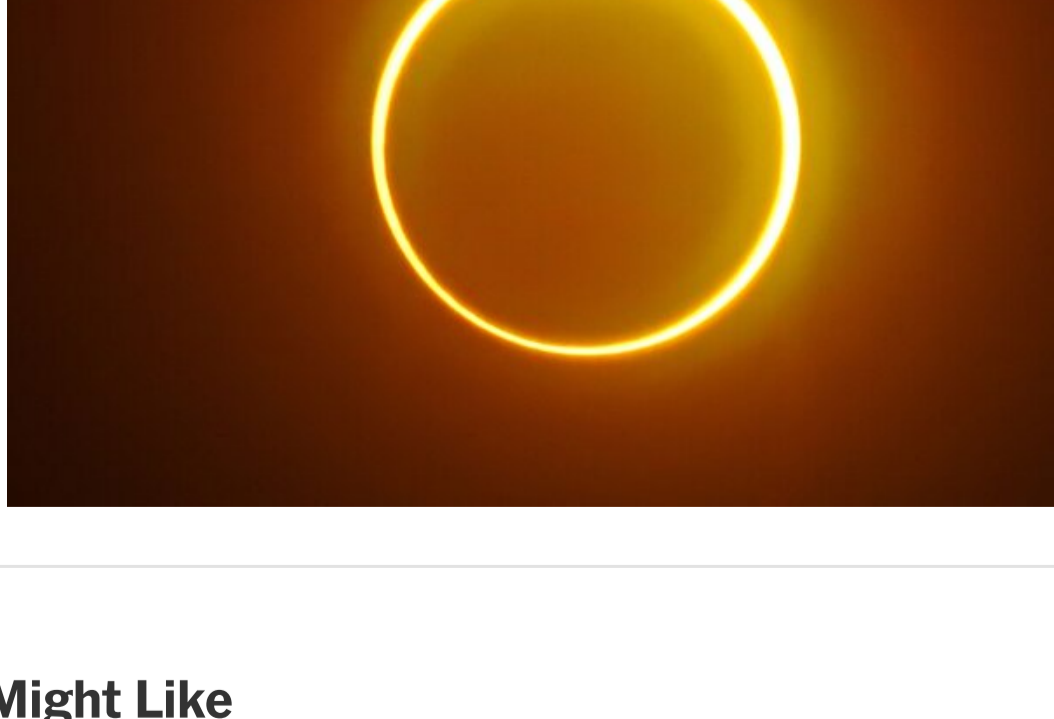
The paper of June 4, 1963. Sixth-grade science news on page 70.

WASHINGTON — A group of about 50 sixth-graders were recruited today to give the \$4,500,000 Tiros VI satellite a helping hand. [The United States Weather Bureau enlisted the aid of the 12-year-olds after having difficulty in identifying cloud formations](#) televised from the orbiting weather observer. [...]

A spokesman at the National Weather Satellite Center said pictures relayed by Tiros showed only gray or white patches for cloud formations. It cannot be determined from the pictures whether the clouds are rain-bearing, nor can their description be pinpointed, he said.

Sync your calendar with the solar system

Never miss an eclipse, a meteor shower, a rocket launch or any other astronomical and space event that's out of this world.



More Newsletters You Might Like

CORONAVIRUS BRIEFING: An informed guide to the global outbreak, with the latest developments and expert advice about prevention and treatment. **Sign Up**

CORONAVIRUS TRACKER: The latest U.S. coronavirus data from the places that matter most to you. **Sign Up**

WELL: The latest on health, fitness and nutrition. **Sign Up**

ADVERTISEMENT