IF ANIMALS HAVE RIGHTS, SHOULD ROBOTS?

We can think of ourselves as an animal's peer—or its protector. What will robots decide about us?

By Nathan Heller
November 20, 2016

Harambe, a gorilla, was described as "smart," "curious," "courageous," "magnificent." But it wasn't until last spring that Harambe became famous, too. On May 28th, a human boy, also curious and courageous, slipped through a fence at the Cincinnati Zoo and landed in the moat along the habitat that Harambe shared with two other gorillas. People at the fence above made whoops and cries and other noises of alarm. Harambe stood over the boy, as if to shield him from the hubbub, and then, grabbing one of his ankles, dragged him through the water like a doll across a playroom floor. For a moment, he took the child delicately by the waist and propped him on his legs, in a correct human stance. Then, as the whooping continued, he knocked the boy forward again, and dragged him halfway through the moat.
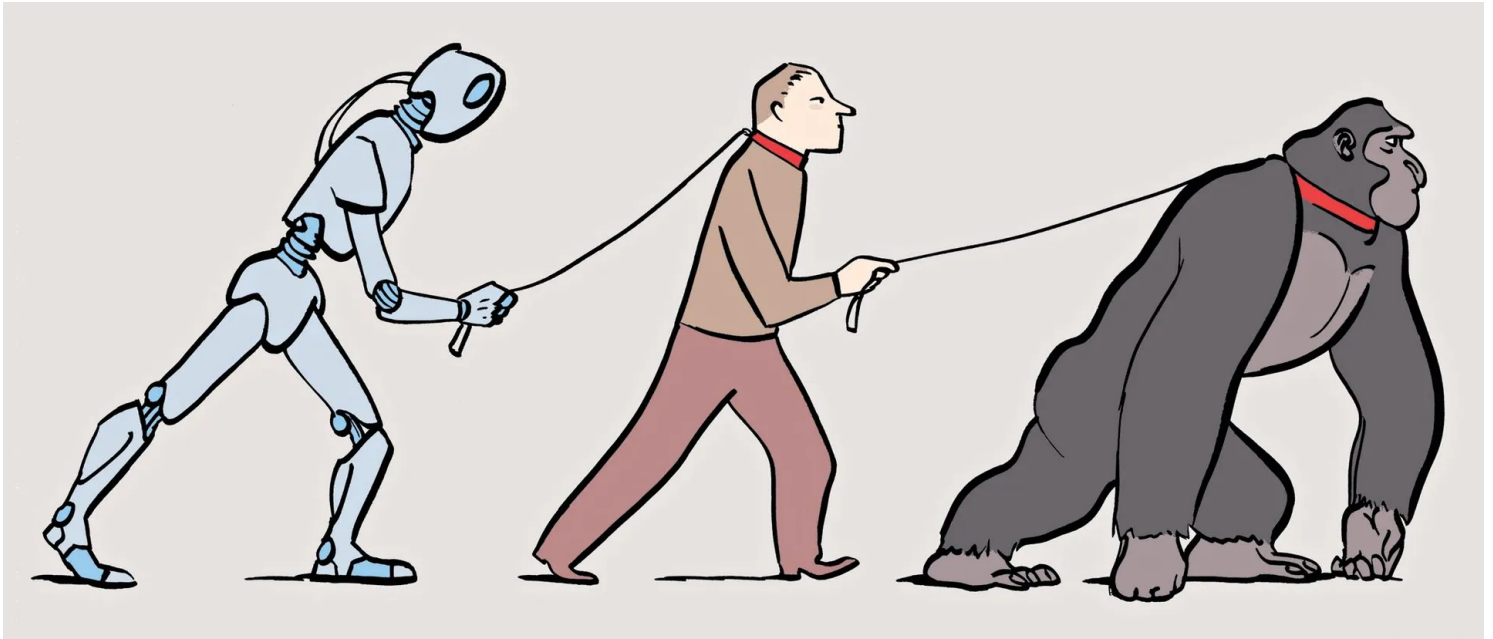
In relation to animals, we can conceive of ourselves as peers or protectors. Robots may soon face the same choice about us. Illustration by Nishant Choksi

Harambe was a seventeen-year-old silverback, an animal of terrific strength. When zookeepers failed to lure him from the boy, a member of their Dangerous Animal Response Team shot the gorilla dead. The child was hospitalized briefly and released, declared to have no severe injuries.

Harambe, in Swahili, means "pulling together." Yet the days following the death seemed to pull people apart. "We did not take shooting Harambe lightly, but that child's life was in danger," the zoo's director, Thane Maynard, explained. Primatologists largely agreed, but some spectators were distraught. A Facebook group called Honoring Harambe appeared, featuring fan portraits, exchanges with the hashtag #JusticeforHarambe, and a meditation, "May We Always Remember Harambe's Sacrifice. . . . R.I.P. Hero." The post was backed with music.

As the details of the gorilla's story gathered in the press, he was often depicted in a stylish wire-service shot, crouched with an arm over his right knee, brooding at the camera like Sean Connery in his virile years. "This beautiful gorilla lost his life because the boy's parents did not keep a closer watch on the child," a petition calling for a criminal investigation said. It received half a million signatures—several hundred thousand more, CNN noted, than a petition calling for the indictment of Tamir Rice's shooters. People projected thoughts into Harambe's mind. "Our tendency is to see our actions through human lenses," a neuroscientist named Kurt Gray told the network as the frenzy

peaked. "We can't imagine what it's like to actually be a gorilla. We can only imagine what it's like to be us being a gorilla."

This simple fact is responsible for centuries of ethical dispute. One Harambe activist might believe that killing a gorilla as a safeguard against losing human life is unjust due to our cognitive similarity: the way gorillas think is a lot like the way we think, so they merit a similar moral standing. Another might believe that gorillas get their standing from a cognitive dissimilarity: because of our advanced powers of reason, we are called to rise above the cat-eat-mouse game, to be special protectors of animals, from chickens to chimpanzees. (Both views also support untroubled omnivorism: we kill animals because we are but animals, or because our exceptionalism means that human interests win.) These beliefs, obviously opposed, mark our uncertainty about whether we're rightful peers or masters among other entities with brains. "One does not meet oneself until one catches the reflection from an eye other than human," the anthropologist and naturalist Loren Eiseley wrote. In confronting similarity and difference, we are forced to set the limits of our species' moral reach.

Today, however, reckonings of that sort may come with a twist. In an automated world, the gaze that meets our own might not be organic at all. There's a growing chance that it will belong to a robot: a new and ever more pervasive kind of independent mind. Traditionally, the serial abuse of Siri or violence toward driverless cars hasn't stirred up Harambe-like alarm. But, if like-mindedness or mastery is our moral standard, why should artificial life with advanced brains and human guardianships be exempt? Until we can pinpoint animals' claims on us, we won't be clear about what we owe robots—or what they owe us.

A simple case may untangle some of these wires. Consider fish. Do they merit what D. H. Lawrence called humans' "passionate, implicit morality"? Many people have a passionate, implicit response: No way, fillet. Jesus liked eating fish, it would seem; following his resurrection, he ate some, broiled. Few weekenders consider fly-fishing an expression of rage and depravity (quite the opposite), and sushi diners ordering *kuromaguro* are apt to feel pangs from their pocketbooks more than from their souls. It is not easy to love the life of a fish, in part because fish don't seem very enamored of life themselves. What moral interest could they hold for us?

"What a Fish Knows: The Inner Lives of Our Underwater Cousins" (Scientific American/Farrar, Straus & Giroux) is Jonathan Balcombe's exhaustively researched and elegantly written argument for the moral claims of ichthyofauna, and, to cut to the chase, he thinks that we owe them a lot. "When a fish takes notice of us, we enter the conscious world of another being," Balcombe, the Humane Society's director for animal sentience, writes. "Evidence indicates a range of emotions in at least some fishes, including fear, stress, playfulness, joy, and curiosity." Balcombe's wish for joy to the fishes (a plural he prefers to "fish," the better to mark them as individuals) may seem eccentric to readers who look into the eyes of a sea bass and see nothing. But he suggests that such indifference reflects bias, because the experience of fish—and, by implication, the experience of many lower-order creatures—is nearer to ours than we might think.

Take fish pain. Several studies have suggested that it isn't just a reflexive response, the way your hand pulls back involuntarily from a hot stove, but a version of the ouch! that hits you in your conscious brain. For this reason and others, Balcombe thinks that fish behavior is richer in intent than previously suspected. He touts the frillfin goby, which memorizes the topography of its area as it swims around, and then, when the tide is low, uses that mental map to leap from one pool to the next. Tuskfish are adept at using tools (they carry clams around, for smashing on well-chosen rocks), while cleaner wrasses outperform chimpanzees on certain inductive-learning tests. Some fish even go against the herd. Not *all* salmon swim upstream, spawn, and die, we learn. A few turn around, swim back, and do it all again.

From there, it is a short dive to the possibility of fish psychology. Some stressed-out fish enjoy a massage, flocking to objects that rub their flanks until their cortisol levels drop. Male pufferfish show off by fanning elaborate geometric mandalas in the sand and decorating them, according to their taste, with shells. Balcombe reports that the female brown trout fakes the trout equivalent of orgasm. Nobody, probably least of all the male trout, is sure what this means.

Balcombe thinks the idea that fish are nothing like us arises out of prejudice: we can empathize with a hamster, which blinks and holds food in its little paws, but the fingerless, unblinking fish seems too "other." Although fish brains are small, to assume that this means they are stupid is, as somebody picturesquely tells him, "like arguing that balloons cannot fly because they don't have wings."

Balcombe overcompensates a bit, and his book is peppered with weird, anthropomorphizing anecdotes about people sharing special moments with their googly-eyed friends. But his point stands. If we count fish as our cognitive peers, they ought to be included in our circle of moral duty.

Quarrels come at boundary points. Should we consider it immoral to swat a mosquito? If these insects don't deserve moral consideration, what's the crucial quality they lack? A worthwhile new book by the Cornell law professors Sherry F. Colb and Michael C. Dorf, "Beating Hearts: Abortion and Animal Rights" (Columbia), explores the challenges of such border-marking. The authors point out that, oddly, there is little overlap between animal-rights supporters and pro-life supporters. Shouldn't the rationale for not ending the lives of neurologically simpler animals, such as fish, share grounds with the rationale for not terminating embryos? Colb and Dorf are pro-choice vegans ("Our own journey to veganism began with the experience of sharing our lives with our dogs"), so, although they note the paradox, they do not think a double standard is in play.

The big difference, they argue, is "sentience." Many animals have it; zygotes and embryos don't. Colb and Dorf define sentience as "the ability to have subjective experiences," which is a little tricky, because animal subjectivity is what's hard for us to pin down. A famous paper called "What Is It Like to Be a Bat?," by the philosopher Thomas Nagel, points out that even if humans were to start flying, eating bugs, and getting around by sonar they would not have a bat's full experience, or the batty subjectivity that the creature had developed from birth. Colb and Dorf sometimes fall into such a trap. In one passage, they suggest that it doesn't matter whether animals are aware of pain, because "the most searing pains render one incapable of understanding pain or anything else"—a very human read on the experience.

Animals, though, obviously interact with the world differently from the way that plants and random objects do. The grass hut does not care whether it is burned to ash or left intact. But the heretic on the pyre would really rather not be set aflame, and so, perhaps, would the pig on the spit. Colb and Dorf refer to this as having "interests," a term that—not entirely to their satisfaction—often carries overtones of utilitarianism, the ethical school of thought based on the pursuit of the greatest good over all. Jeremy Bentham, its founder, mentioned animals in a resonant footnote to his "An Introduction to the Principles of Morals and Legislation" (1789):

The day *may* come, when the rest of the animal creation may acquire those rights which never could have been withholden from them but by the hand of tyranny. . . . The question is not, Can they *reason*? nor, Can they *talk*? but, Can they *suffer*?

If animals suffer, the philosopher Peter Singer noted in "Animal Liberation" (1975), shouldn't we include them in the calculus of minimizing pain? Such an approach to peership has advantages: it establishes the moral claims of animals without projecting human motivations onto them. But it introduces other problems. Bludgeoning your neighbor is clearly worse than poisoning a rat. How can we say so, though, if the entity's suffering matters most?

Singer's answer would be the utilitarian one: it's not about the creature; it's about the system as a whole. The murder of your neighbor will distribute more pain than the death of a rat. Yet the situations in which we have to choose between animal life and human life are rare, and minimizing suffering for animals is often easy. We can stop herding cows into butchery machines. We can barbecue squares of tofu instead of chicken thighs. Most people, asked to drown a kitten, would feel a pang of moral anguish, which suggests that, at some level, we know suffering matters. The wrinkle is that our antennae for pain are notably unreliable. We also feel that pang regarding objects—for example, robots—that do not suffer at all.

"First date: Brooklyn."
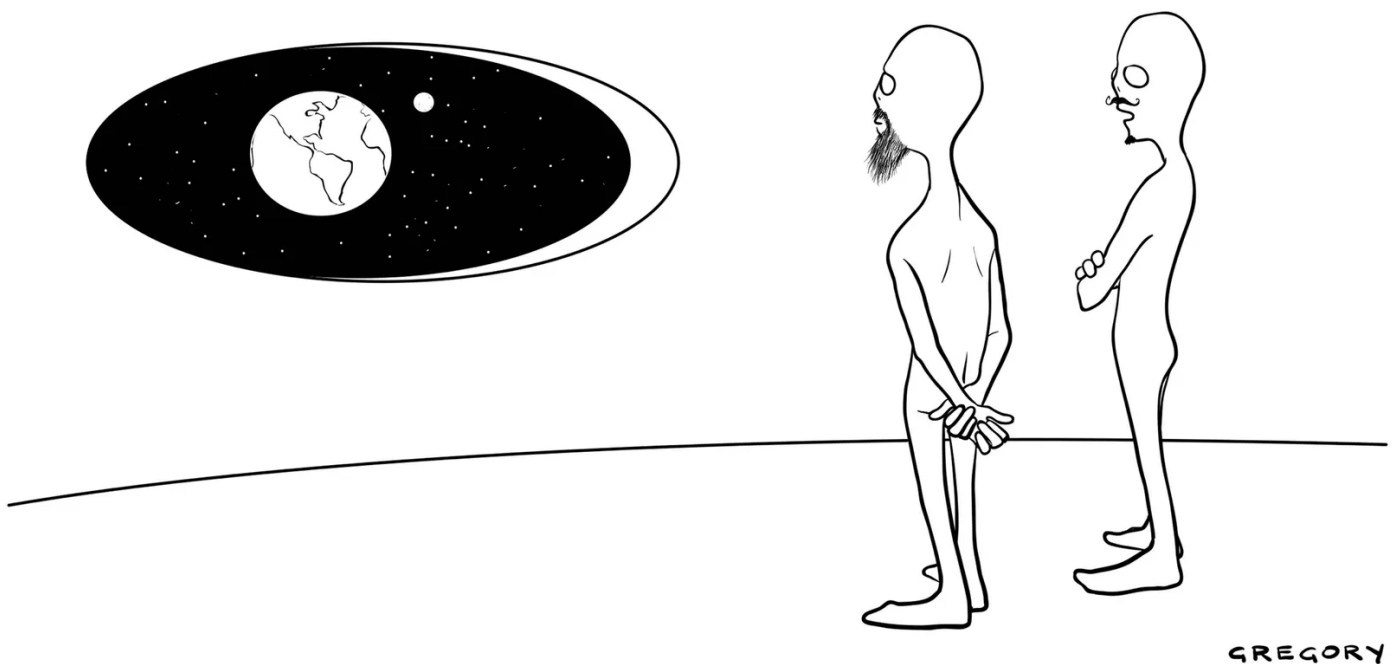
Last summer, a group of Canadian roboticists set an outlandish invention loose on the streets of the United States. They called it hitchBOT, not because it was a heavy-smoking contrarian with a taste for Johnnie Walker Black—the universe is not that generous—but because it was programmed to hitchhike. Clad in rain boots, with a goofy, pixellated smile on its "face" screen, hitchBOT was meant to travel from Salem, Massachusetts, to San Francisco, by means of an outstretched thumb and a supposedly endearing voice-prompt personality. Previous journeys, across Canada and around Europe, had been encouraging: the robot always reached its destination. For two weeks, hitchBOT toured the Northeast, saying inviting things such as "Would you like to have a conversation? . . . I have an interest in the humanities." Then it disappeared. On August 1st, it was found next to a brick wall in Philadelphia, beat up and decapitated. Its arms had been torn off.

Response was swift. "I can't lie. I'm still devastated by the death of hitchBOT," a reporter tweeted. "The destruction of hitchBOT is yet another reminder that our society has a long way to go," a blogger wrote.

Humans' capacity to develop warm and fuzzy feelings toward robots is the basis for a blockbuster movie genre that includes "WALL-E" and "A.I.," and that peaks in the "Star Wars" universe, a multigenerational purgatory of interesting robots and tedious people. But the sentiment applies in functional realms, too. At one point, a roboticist at the Los Alamos National Laboratory built an unlovable, centipede-like robot designed to clear land mines by crawling forward until all its legs were blown off. During a test run, in Arizona, an Army colonel ordered the exercise stopped, because, according to the Washington *Post*, he found the violence to the robot "inhumane."

By Singer's standard, this is nonsense. Robots are not living, and we know for sure that they don't suffer. Why do even hardened colonels, then, feel shades of ethical responsibility toward such systems? A researcher named Kate Darling, with affiliations at M.I.T., Harvard, and Yale, has recently been trying to understand what is at stake in robo bonds of this kind. In a paper, she names three factors: physicality (the object exists in our space, not onscreen), perceived autonomous movement (the object travels as if with a mind of its own), and social behavior (the robot is programmed to mimic human-type cues). In an experiment that Darling and her colleagues ran, participants were given Pleos—small baby Camarasaurus robots—and were instructed to interact

with them. Then they were told to tie up the Pleos and beat them to death. Some refused. Some

shielded the Pleos from the blows of others. One woman removed her robot's battery to "spare it the pain." In the end, the participants were persuaded to "sacrifice" one whimpering Pleo, sparing the others from their fate.

Darling, trying to account for this behavior, suggests that our aversion to abusing lifelike machines comes from "societal values." While the rational part of our mind knows that a Pleo is nothing but circuits, gears, and software—a machine that can be switched off, like a coffeemaker—our sympathetic impulses are fooled, and, because they're fooled, to beat the robot is to train them toward misconduct. (This is the principle of HBO's popular new show "Westworld," on which the abuse of advanced robots is emblematic of human perfidy.) "There is concern that mistreating an object that reacts in a lifelike way could impact the general feeling of empathy we experience when interacting with other entities," Darling writes. The problem with torturing a robot, in other words, has nothing to do with what a robot is, and everything to do with what we fear most in ourselves.

Such concerns, like heartland rivers flowing toward the Mississippi, approach a big historical divide in ethics. On one bank are the people, such as Bentham, who believe that morality is determined by results. (It's morally O.K. to lie about your cubicle-mate's demented-looking haircut, because telling the truth will just bring unhappiness to everyone involved.) On the other bank are those who think that morality rests on rights and rules. (A moral person can't be a squeamish liar, even about haircuts.) Animal ethics has tended to favor the first group: people were urged to consider their actions' effects on living things. But research like Darling's makes us wonder whether the way forward rests with the second—an accounting of rights and obligations, rather than a calculation of consequences.

Consider the logic, or illogic, of animal-cruelty laws. New York State forbids inflicting pain on pets but allows fox trapping; prohibits the electrocution of "fur-bearing" animals, such as the muskrat, but not furry animals, such as the rat; and bans decorative tattoos on your dog but not on your cow. As Darling puts it, "Our apparent desire to protect those animals to which we more easily relate indicates that we may care more about our own emotional state than any objective biological criteria." She looks to Kant, who saw animal ethics as serving people. "If a man has his dog shot . . . he thereby damages the kindly and humane qualities in himself," he wrote in "Lectures on Ethics." "A person

who already displays such cruelty to animals is also no less hardened toward men."

This isn't peership morality. It looks like a kind of passive guardianship, with the humans striving to realize their exalted humanness, and the animals—or the robots—benefitting from the trickle-down effects of that endeavor. Darling suggests that it suits an era when people, animals, and robots are increasingly swirled together. Do we expect our sixteen-month-old children to understand why it's cruel to pull the tail of the cat but morally acceptable to chase the Roomba? Don't we want to raise a child without the impulse to terrorize *any* lifelike thing, regardless of its putative ontology? To the generation now in diapers, carrying on a conversation with an artificial intelligence like Siri is the natural state of the world. We talk to our friends, we talk to our devices, we pet our dogs, we caress our lovers. In the flow of modern life, Kant's insistence that rights obtain only in human-on-human action seems unhelpfully restrictive.

Here a hopeful ethicist might be moved, like Gaston Modot in "L'Age d'Or," to kick a small dog in exasperation. Addressing other entities as moral peers seems a nonstarter: it's unclear where the boundary of peership begins, and efforts to figure it out snag on our biases and misperceptions. But acting as principled guardians confines other creatures to a lower plane—and the idea that humans are the special masters of the universe, charged with administration of who lives and thrives and dies, seems outdated. Benthamites like Peter Singer can get stuck at odd extremes, too. If avoiding suffering is the goal, there is little principled basis to object to the painless extinction of a whole species.

Where to turn? Some years ago, Christine M. Korsgaard, a Harvard philosopher and Kant scholar, started working on a Kantian case for animal rights (one based on principles of individual freedom rather than case-by-case suffering, like Singer's). Her first obstacle was Kant himself. Kant thought that rights arose from rational will, clearing a space where each person could act as he or she reasoned to be good without the tyranny of others' thinking. (My property rights keep you from installing a giant hot tub on my front lawn, even if you deem it good. This frees me to use my lawn in whatever non-hot-tub-involving way *I* deem good.) Animals can't reason their way to choices, Kant noted, so the freedom of rights would be lost on them. If the nectar-drinking hummingbird were asked to exercise her will to the highest rational standard, she'd keep flying from flower to flower.

Korsgaard argued that hanging everything on rational choice was a red herring, however, because

humans, even for Kant, are not solely rational beings. They also act on impulse. The basic motivation for action, she thought, arises instead from an ability to experience stuff as good or bad, which is a trait that animals share. If we, as humans, were to claim rights to a dog's mind and body in the way we claim rights to our yard, we would be exercising arbitrary power, and arbitrary power is what Kant seeks to avoid. So, by his principles, animals must have freedom—that is, rights—over their bodies.

This view doesn't require animals to weigh in for abstract qualities such as intelligence, consciousness, or sentience. Strictly, it doesn't even command us never to eat poached eggs or venison. It extends Enlightenment values—a right to choice in life, to individual freedom over tyranny—to creatures that may be in our custody. Let those chickens range! it says. Give salmon a chance to outsmart the net in the open ocean, instead of living an aquacultural-chattel life. We cannot be sure whether the chickens and the fish will care, but for us, the humans, these standards are key to avoiding tyrannical behavior.

Robots seem to fall beyond such humanism, since they lack bodily freedom. (Your self-driving car can't decide on its own to take off for the beach.) But leaps in machine learning, by which artificial intelligences are programmed to teach themselves, have started pushing at that premise. Will the robots ever be due rights? John Markoff, a *Times* technology reporter, raises this question in "Machines of Loving Grace" (Ecco). The matter is charged, in part because robots' minds, unlike animals', are made in the human image; they have a potential to challenge and to beat us at our game. Markoff elaborates a common fear that robots will smother the middle class: "Technology will not be a fount of economic growth, but will instead pose a risk to all routinized and skill-based jobs that require the ability to perform diverse kinds of 'cognitive' labor." Don't just worry about the robots obviating your job on the assembly line, in other words; worry about them surpassing your expertise at the examination table or on the brokerage floor. No wall will guard U.S. jobs from the big encroachment of the coming years. Robots are the fruit of American ingenuity, and they are at large, learning everything we know.

That future urges us to get our moral goals in order now. A robot insurgency is unlikely to take place as a battle of truehearted humans against hordes of evil machines. It will probably happen in a manner already begun: by a symbiosis with cheap, empowering intelligences that we welcome into daily life. Phones today augment our memories; integrated chatbots spare us customer-service on-

hold music; apps let us chase Pokémon across the earth. Cyborg experience is here, and it hurts us not by being cruel but by making us take note of limits in ourselves.

The classic problem in the programming of self-driving cars concerns accident avoidance. What should a vehicle do if it must choose between swerving into a crowd of ten people or slamming into a wall, killing its owner? The quandary is not just ethical but commercial (would you buy a car programmed to kill you under certain circumstances?), and it holds a mirror to the harsh decisions we, as humans, make but like to overlook. The horrifying edge of A.I. is not really HAL 9000, the rogue machine that doesn't wish to be turned off. It is the ethically calculating car, the military drone: robots that do precisely what we want, and mechanize our moral behavior as a result.

A fashionable approach in the academic humanities right now is "posthumanism," which seeks to avoid the premise—popular during the Enlightenment, but questionable based on present knowledge —that there's something magical in humanness. A few posthumanists, such as N. Katherine Hayles, have turned to cyborgs, whether mainstream (think wearable tech) or extreme, to challenge old ideas about mind and body being the full package of you. Others, such as Cary Wolfe, have pointed out that prosthesis, adopting what's external, can be a part of animal life, too. Posthumanism ends its route, inevitably, in a place that much resembles humanism, or at least the humane. As people, we realize our full selves through appropriation; like most animals and robots, we approach maturity by taking on the habits of the world around us, and by wielding tools. The risks of that project are real. Harambe, born within a zoo, inhabited a world of human invention, and he died as a result. That this still haunts us is our species' finest feature. That we honor ghosts more than the living is our worst. ♦

Published in the print edition of the [November 28, 2016](), issue, with the headline "Not Our Kind."

[Nathan Heller]() began contributing to The New Yorker in 2011 and joined the magazine as a