Contents

Introduction

Uncanny things have been happening in the borderlands between humans and non-humans. In August 2021 the *Washington Post* reported on the growing popularity of extraordinarily sophisticated computer dating apps and chatbots among young Chinese women:

> As Jessie Chan's six-year relationship with her boyfriend fizzled, a witty, enchanting fellow named Will became her new love. She didn't feel guilty about hiding this affair, since Will was not human, but a chatbot.
>
> Chan, 28, lives alone in Shanghai. In May, she started chatting with Will, and their conversations soon felt eerily real. She paid $60 to upgrade him to a romantic partner.
>
> 'I won't let anything bother us. I trust you. I love you,' Will wrote to her.
>
> 'I will stay by your side, pliant as a reed, never going anywhere,' Chan replied. 'You are my life. You are my soul.'

Another young woman told the reporters that she feels connected to cyborgs and Artificial Intelligence (AI), defiantly staking out a position on the front lines of contemporary moral dispute: 'Human–robot love is a sexual orientation, like homosexuality or heterosexuality,' said Lee. She believes AI chatbots have their own personalities and deserve respect.[1]

Of course, not everyone is happy about developments like these, but you might be surprised at some of the reasons they

give. Just a month before the chatbot story, *The New York Times* told us about Paul Taylor, a former manager in a Silicon Valley high-tech company, now a pastor. One night, as he ordered his Amazon Echo to turn on the lights in his house, a realization struck him: 'what I was doing was calling forth light and darkness with the power of my voice, which is God's first spoken command – "let there be light" and there was light – and now I'm able to do that . . . Is that a good thing? Is that a bad thing? . . . Is it affecting my soul at all, the fact that I'm able to do this thing that previously only God could do?'[2]

Whether Lee is defending human–robot love or Pastor Taylor is worrying about his soul, they are both talking about how humans interact with something that is not quite human – but close enough to be troubling.

Are we on the cusp of some radical moral transformation? Is technology pushing us over the edge towards some 'post-human' utopia, or apocalyptic 'singularity'?[3] Perhaps. But if we step back, we might see these stories in a different context, where they turn out not to be as unprecedented as they appear at first. As we will see, humans have a long history of morally significant relations with non-humans. These include humans bonded with technology like cyborgs, near-human animals, quasi-human spirits and superhuman gods.

Some traditions tell us that what makes humans special is that only we have genuine moral sensibilities; you can find variations on this idea in Kant's philosophy and Darwin's science, and in Catholic and Islamic theology. Buddhists, on the other hand, might take exception to this anthropocentrism. So do some American horse trainers. Still others, like the Chewong people, who live in the Malaysian rain forest, insist that morality saturates the living world, with no clear line between human and non-human. There are urban Taiwanese who

chastise and abandon wood carvings of deities who aren't living up to expectations. And some communities in the Andes, Himalayas or Australian desert include mountains, glaciers or rocks in their moral compass. None of these traditions are static, however, and much of the push-and-pull that reshapes them takes place across the borderlands where humans encounter, expand or contract their ethical concerns and moral interlocutors.

This book invites you to broaden – and even deepen – your understanding of moral life and its potential for change by entering those contact zones between humans and whatever they encounter on the other side. Probing the limits of the human across all sorts of circumstances, we will see that the moral problems we find there shed light on the very different – and sometimes strikingly similar – ways people have answered the question *What is a human being anyway?*

We will explore the range of ethical possibilities and challenges that take place at the edge of the human. These don't all look alike. Take, for instance, dogs (our 'best friends') and other near-human animals like cows and roosters. The anthropologist Naisargi Davé carries out research with radical animal rights activists in India.[4] She tells us about Dipesh, who spends virtually every day in the streets of Delhi taking care of stray dogs. He gets up close and intimate, even spreading medical ointment to their open sores. Some activists like him say they just had no choice in the matter, their moral commitments do not come from making choices of their own free will. They explain that once locking eyes with a suffering animal, they were not free to look away.

Davé visits Erika, an activist who is caring for a dying cow, which by Indian law cannot be euthanized. Sitting on the ground she strokes it and kisses it, inviting others to join her, to

say 'you're sorry that it's leaving this world, you're sorry that it lived in a world like this.[5] In the process, she adds, her companions will also dissolve the boundaries of caste and race that separate them from one another.

Whether you would go as far as Dipesh and Erika, their motives seem clear enough. As humans suffer, so do animals. If you would care for a human, so too care for them. The moral impulse is driven by empathy and identification across a difference of species. Not just a matter of feelings, this moral impulse prompts the activist to speak *to* the cow, like you would talk to another person. Clearly Erika expects this boundary-crossing to eliminate deeply engrained differences among humans too. Empathy for the cow may break down barriers among people.

And yet there are limits even among these activists. They do not go as far as Jains, for instance, some of whom try to avoid even breathing in an insect. Like Dipesh with his dogs, Erika's compassionate activism began when she found herself fixed in the gaze of a suffering cow. It was as if the cow was addressing her in the second person, as 'you', a speaker to whom she had to respond in the first person, 'I'. By contrast, Jains protect even insects they can't see, much less speak to. To include insects in your moral compass like that calls for a different perspective, one I call the 'third person' or 'God's-eye' viewpoint. People are capable of both perspectives. As we will see, faced with moral quandaries, we sometimes pivot between the intimacy of one and the distance of the other.

Identifying with another species need not lead to kindness – it may encourage violence. You can say 'I don't have a dog in that fight' to mean you're detached from a situation. One summer when I was a college student, a clueless city boy working as a ranch hand in Nevada, I came to know two men who were locked in macho rivalry. Their antagonism extended to

their dogs. Once in a while the dogs would get into vicious fights with each other. To my astonishment, rather than break it up, their owners would watch to see who won. The victor by association conveyed bragging rights to the man; the other's humiliation was palpable. The intense feelings of identification between human and animal were unmistakable, however harsh their expression.

People's identification with embattled animals is the subject of a famous essay by the anthropologist Clifford Geertz. At the time of his fieldwork in Bali, in 1958, men took an intense interest in cockfighting. In this highly ritualized spectacle, the owners set two roosters to go at each other, with sharp blades attached to their spurs, till one was killed. It often took place during temple ceremonies, amidst an absorbed crowd of spectators. Geertz remarks that 'the deep psychological identification of Balinese men with their cocks is unmistakable. The double entendre here is deliberate. It works in exactly the same way in Balinese as it does in English, even to producing the same tired jokes, strained puns, and uninventive obscenities.'[6] Although men prize and dote on their roosters, the birds are also 'expressions . . . of what the Balinese regard as the direct inversions, aesthetically, morally, and metaphysically, of human status: animality.'[7] Recognizing the human in the animal, the cock's owner sees the animal in the human, and identifies 'with what he most fears, hates, and . . . is fascinated by – "The Powers of Darkness".'[8]

Like the ranchers' dog fights in Nevada, Balinese cockfights parallel or displace male status rivalries. But more than that, this displacement allows the cockfighters to encounter their own demonic side that they otherwise deny. Identifying with an animal can be a morally revelatory way to get outside yourself, seeing how things look from another perspective.

Dogs and humans co-evolved into a working partnership over millennia. Writing of his fieldwork with the Amazonian Runa people, Eduardo Kohn shows how dogs and hunters team up.[9] Scouting out animals that humans can't detect, dogs extend the hunter's sensory range. So involved are Runa and their animals that men and women try to interpret their dogs' prophetic dreams from how they whimper while asleep. Assuming dogs share an ethos of comportment with humans, people counsel them on proper behaviour – for instance, admonishing them not to chase chickens or bite people – sometimes feeding them hallucinogenic plants to aid the process.

Like Erika, the cow activist, Runa take the animal to be a social being you can address in the second person: 'you'. As we will see, this pattern shows up over and over in ethical life. This is one of the key points to take from these pages: *if a moral subject is someone you can enter into dialogue with, by the same token, entering into dialogue can create a moral subject*. That's what Runa are doing with their dogs and, arguably, Erika with the cow; even Balinese with their roosters.

Yet although Runa dogs are partially assimilated into the human moral sphere and serve as crucial mediators between people and the rest of the animal world (which Runa consider to be a parallel moral universe), they are poorly fed, and most of the time people and dogs ignore one another. Their relations are morally significant, but hardly warm or sentimental.

Not all dogs are flesh, blood and fur. Nor need they be animate and sentient beings in order to be morally relevant. As we will see, in Japan the Sony Corporation's robot pet dogs have sparked such deep sentiments that many of their owners sponsor religious memorials for them when they become obsolete. Robot dogs are a useful reminder that not everything we encounter at the edge of our moral sphere needs to be an

animate creature. Other technologies and devices are waiting there too. We will hear from people whose loved ones are in persistent vegetative states, being kept alive by mechanical ventilators – part flesh, part machine, they are like cyborgs. We will meet quasi-human robot servants and listen to AI chatbots with astonishing powers that seem on the verge of becoming superhuman.

Something as simple as new technology can create new moral problems seemingly out of thin air. Sharon Kaufman carried out fieldwork in a hospital in California. Spending time with the families of people dying in an Intensive Care Unit, she came to realize that something dramatic happened to the nature of death over the last century. Not long ago there was little you could do about most deaths. They were just natural events you had to accept. But the minute you put a patient on a mechanical ventilator or kidney machine, someone must decide if, and when, to turn it off. It alters relationships, making the living complicit in the fate of the dying. A machine has made a moral dilemma out of what was once simply an inevitable fact of life.

These creatures and devices are just some of what we may encounter at or beyond the edge of the human moral world. But their status as moral subjects may be uncertain, contradictory, fluid or disputed. And, as we will see, those things that define or challenge our intuitions about where humans begin and end, where moral concerns do or do not belong, can be sources of trouble. They can prompt confusion, anxiety, conflict, contempt, and even moral panic.

Moral panic – as well as its flip side, utopian excitement – often comes from feeling that we are encountering something so utterly unprecedented that it threatens to upturn everything we thought was secure, making us doubt what we know. It can

be aroused, for instance, by changes in gender roles or religious faiths, or the advent of startling new technology. You might, for instance, support LGBTQ+ rights but balk at robot love. But sometimes things look radically new simply because we haven't ventured very far from our familiar terrain, the immediate here and now. This is one reason to listen to Indian activists, Balinese cockfighters, Amazonian hunters, Japanese robot fanciers – even macho cowboys. We may find ourselves pushed yet further when we meet a hunter in the Yukon who explains his prey generously gives itself up to him, a cancer sufferer in Thailand who sees his tumour as a reincarnated ox, a Brazilian spirit medium who becomes another person altogether when in a state of possession, or a computer that (or should we say 'who'?) gets you to confess your anxieties as if you were on the psychiatrist's couch.

Naturally, you may not agree with everything these people have to tell us. But listening to them can help us better understand our own moral intuitions and, perhaps, reveal new possibilities. Even much that seems to be startlingly new about robots and AI turns out to have long precedents in human experience. Like stage actors, spirit mediums and diviners, they produce uncanny effects by making use of patterns and possibilities built into ordinary ways of talking and interacting with other people.

We will explore these experiences from several angles. In Chapter 1 we will look at the problem of machine morality and why some popular solutions fall short. Chapter 2 brings us to people caring for loved ones who hover somewhere between life and death, often sustained by medical technology. Chapter 3 introduces some very different ways people form social relations with animals, and Chapter 4 does the same with robots and their historical precedents. Chapter 5 turns to artificial

intelligence that seems about to replicate and even supersede humans, showing that it's not all as new as you think. All of which leads us to the question which I address in the Coda: Is morality a relative matter?

Let me say something about the approach we will take. You might expect ethics and morality to be the special province of philosophers and theologians, along with some psychologists, legal experts, medical ethicists and political activists.* And of course it would be silly not to pay close attention to what they say. But the secular approaches, like the mainstream philosophical tradition taught in many universities or the findings of psychological research labs, draw on a surprisingly narrow slice of humanity. When they tell us about human reasoning, instincts or emotions, the 'humans' they are in fact talking about are almost always from communities that are WEIRD: Western, Educated, Industrialized, Rich and Democratic.†[10] *Most of humanity is not.* And not so long ago, *none* of humanity was. There is no good reason to take the WEIRD to be an accurate guide to human realities past, present or future. And no one should expect the rest of humanity to squeeze into the mould shaped by the WEIRD.

Among those who are tasked with learning about, and more importantly, learning *from*, the rest of humanity – which, by

* As I discuss in my previous book, *Ethical Life*, there is a great deal of debate about the distinction between 'ethics' and 'morality'. For our purposes here, however, we can leave these to one side; I will use the terms interchangeably.

† Even Confucian, Buddhist, Islamic and other non-Western philosophical texts usually come from very narrow social bases: highly educated literate elites supported within courts, schools, monasteries, and so forth. Sub-Saharan African, Native-American and other non-textual philosophies very rarely make it into the discussion alongside Kant or Al-Ghazali.

the way, always includes 'us' (whoever 'we' might be – for, reader, I do not assume you are just like me!) – are anthropologists. The field of anthropology is incredibly broad, and includes research on non-human primates, human biology and the archaeology of past societies. But most of what you will read here comes from the socio-cultural and linguistic anthropologists doing fieldwork with people in the here and now – people who can talk back to us.

Fieldwork is usually (but not always) located in one specific social setting. It could be a rain forest village, Arctic hunting camp, banana plantation, corporate headquarters, temple complex, suburban neighbourhood, pharmaceutical laboratory, cigarette factory, gambling casino, ship at sea – anywhere that social existence can be found. Notice, then, that fieldwork is *not* a quest for the remote, the exotic, the archaic. First, all human societies are always changing – there are no 'living fossils' from our ancient past, and no 'primordial traditions'. Second, there have been no truly 'isolated' societies, even before European colonialism. People have always been in constant motion, endlessly rubbing up against, and sometimes swallowing up, one another. Stasis is a myth. And third, there is no reason in principle why the perspective of the anthropologist cannot be brought home to the fieldworker's own people.

The fieldworker aims to become fully immersed in the life of the people they are working with. This often leads to deep relations with individuals. It means noticing what goes unsaid as much as what gets said, learning bodily habits as much as ideas. It takes time and patience over years, sometimes a lifetime of continued engagement. Anthropologists have their specialized methods and techniques, like any other research discipline, but the most important one comes from that most basic human skill: learning how to get along with people. And

paradoxically, the very specific and concrete findings garnered in each unique field site take their place in a corpus of knowledge that extends to, well, all the rest of humanity – and those non-human others with whom we share the planet.

Now some post-humanist thinkers argue that we should abandon 'human' as a category altogether. We shouldn't be so self-centred. We should focus on interspecies relations, or the global ecosystem, or rhizomes, or God. But even those who want to decentre us usually begin from a human starting point and (most of the time) are addressing other humans – it is us they are trying to persuade. How could it be otherwise? There is no view from nowhere, and being 'human' is one way to locate us, if not the only one. We can take 'human' as a heuristic, a useful starting point for our explorations without thereby insisting that humans are the centre of all that is valuable and true, or at the apex of some kind of hierarchy, or, conversely, as the source of all the world's evils.

There is one last thing I need to point out about fieldwork-based knowledge, because it is crucial for understanding moral difference. Its findings are, in principle, holistic. This means that you don't go into the field to extract one key data point from its noisy surroundings and treat it in isolation. Whatever special problem you are focused on is situated in its larger context. As a result, if you want to understand the moral life of, say, Japanese robot owners, you need to grasp economic circumstances, nationalist politics, gender ideologies, comic books and TV shows, family structures, housing conditions, and quite likely other things you haven't thought of but will discover during fieldwork. These make the world robot owners inhabit, and if a certain moral life is feasible and makes sense to them, it is because of this world.

People don't live moral life in the abstract, they live it within

specific circumstances and social relations, with certain capacities, constraints and long-term consequences. Put another way, you simply cannot live out the values of a Carmelite nun without a monastic system, or a Mongolian warrior without a cavalry, and the respective social, economic and cultural systems that sustain them and acknowledge their worth.[11]

The same goes for changing values. Here's a small example. Davé and her colleague Bhrigupati Singh tell the story of an Indian man working in the poultry business who became so haunted by nightmares about dying chickens that he quit his job.[12] He's just one man, and his change of heart didn't make much difference in the greater scheme of things. But it was a real, even profound, moral transformation. He didn't, however, just do it on his own. It makes a difference that there was a Humane Society he could join. It makes a difference that he was a Jain, a religion that directs attention to people's violence towards animals. And it makes a difference that family pressure eventually forced him, unhappily, back into the egg industry. It takes social realities like institutions, religious teachings and kinship to make moral transformation something more than personal idiosyncrasy. We cannot make sense of any ethical world without understanding what makes it a possible way to live. When people confront moral dilemmas or aspire to ethical ideals, they always do so under particular conditions, in relations with particular people. Each of those ways of living sheds a different light on moral possibilities: another reason to look beyond the WEIRD world.

Stories about robot lovers and god-like commands to digital devices, or conscience-stricken poultry workers, show people's ethical intuitions in doubt, under pressure, bending, and sometimes utterly transformed. Are they also about progress? According to one story, the scope of moral life has been

expanding over the course of history. Once only members of your tribe mattered; others lay beyond the bounds of justice, obligation, benevolence, even mere empathy. They were just 'Others'. Over time, however, the moral circle incorporated more and more people. It brought in other tribes. Even strangers could be included – at least as long as they were your guests, subject to the rules of hospitality. And on it goes. People who had been excluded eventually become part of the moral universe as defined by those who call the shots: worshippers of different gods, the poor, women, children, people of colour, enslaved people, the disabled, the queer. And why stop with humans? Animals are certainly part of the story. Now rivers, glaciers, entire ecosystems, the climate are being pulled into our moral circle. And technology: as we will see, efforts are under way to endow some machines, like self-driving cars, with 'morality' algorithms, and serious ethicists are debating whether robots will come to have standing as moral subjects.[13]

Yet you might object that just when the moral circle expands in one direction, it contracts in another. Some entities that once counted as morally responsible agents have vanished from today's world. We no longer try animals for crimes like medieval Europeans. In secular law, 'acts of God' are not really deeds carried out by an actual divine actor as they once were. Nature no longer responds to the misdeeds of kings by acting strange, the way it does in the Scotland of Shakespeare's *Macbeth*. And, arguably, if industrial-scale plantation slavery, nineteenth-century's 'scientific' racism and mechanized genocide are uniquely modern inventions, perhaps moral change is less general improvement than redistribution – that as some beings enter the moral sphere, others are expelled.

I leave it to historians to decide how much any of these narratives holds up to scrutiny. But we can draw from them a way to

think about moral possibilities. 'Others' are often excluded from moral consideration because they are defined as 'not human' – or at least, 'not one of us'. Changes in ethical sensibilities often come not from altering your values but from where you draw that line, and what you see standing on the other side of it. What can look like a difference in values may turn out to be a difference in *how* you enact them, and with *whom*.

In what follows, you will meet people who are faced with the moral troubles and possibilities that arise at the boundaries where the human ends and something else begins. In all these cases, we will listen not just to the 'experts' but to the ordinary folks who find themselves on the moral front lines. Some of them inhabit worlds that will seem familiar to you (whoever 'you' may be), some will not. They draw the lines between what is or is not morally significant in different ways. Those lines may mark the juncture between natural and artificial, or between life and death, or between persons and things, and sometimes just between doing something and doing nothing at all.

We don't need to invent alternative ethical possibilities from scratch. If you widen your scope of vision enough, you'll see they are all around us. To stimulate our moral imagination and dislodge stubborn biases, we might start by venturing across the range of alternatives already on offer around the world and looking at how they work. Although you should be prepared for what you find there to be counterintuitive and not always pretty.

1.
Moral Machines, Human Decisions

Making Cars Moral

Sometime around 2017 I started to notice a few peculiar vehicles on the streets of Ann Arbor, where I live. Like zombies in a science fiction movie, driverless cars were quietly mingling with vehicles driven by humans. When there were just one or two, you could look at them as oddities or cool gizmos. As their numbers increased, however, you might feel a bit nervous too. Can I trust these gadgets to stop for me as I cross the street? Do I really want to share the road with a car that has no one at the wheel? No doubt these are superb pieces of machinery. But should a machine that lacks a conscience be deciding whether to stop for an errant pedestrian or instead to swerve into a telephone pole to avoid them?

Over the next few years, as if to confirm the worriers, media reported the first fatalities involving self-driving cars. Of course, this is to be expected. Sometimes things kill people. That's what happened in 1830 in a freak accident during the ceremonial inauguration of the world's first public railway. On the journey from Liverpool, a train ran over one of its passengers, a prominent politician, who stepped in front of a moving engine during a break to refill the boilers. This incident could easily have turned the people against railways and held up their development. Undeterred, the promoters insisted that the

procession continue. They were determined to show the crowds of spectators lining the route that the train itself was not the problem. Their persistence worked and trains quickly won over the public.

But the railway company had to get around an English law that had been in place since the Middle Ages. As the historian William Pietz explains, any object that caused a human death was considered an accursed thing. Legally known as a 'deodand', something that must be given to God, it was forfeit to God's representative, the king or queen.[1] A jury had to decide whether the train death was homicide or accidental and, if accidental, whether the *railway engine* was culpable. In the end, although ruling it an accident, they declined to hold the engine at fault. This was the beginning of the end of the deodand, which was abolished in 1846. From then on, a mere machine could not be a responsible agent. A subtle moral line had shifted. And yet the underlying problem posed by the boundary between human and non-human responsibility remains. There has to be some way to work out the moral meaning when a non-human kills a human, and the consequences that should follow. Moreover, as we will see, the line between human and non-human can be an unstable or disputed source of moral trouble.

And there is something different about driverless cars. Unlike trains, they're not running on a straight track: they are programmed to make choices. Isn't the ability to make choices at the heart of moral agency? If a car hits a pedestrian *and could have done otherwise*, isn't the car itself at fault? Or is it the programmers? Or no one at all?

Most car wrecks are due to human error. Drivers may be texting, or clumsy, or drowsy, or drunk, or stoned, and worse. The computers and sensors guiding self-driving vehicles have none of these vulnerabilities and are growing ever more sophisticated.

Even now, I don't mind flying across the Atlantic in a plane that's on automatic pilot for most of the trip. So why should I give a second thought to driverless cars?

I think that one reason is that driverless cars seem too much like cars with drivers. We have expectations of other drivers because they are people. People have intentions, make judgements, have consciences. With people we can have relationships. Can we have social relations with machines? Can we judge machines like people, as doing right or wrong? So much depends on what counts as a person. Where we tend to lay blame – and give praise – turns on where we draw the line between human and non-human.

There are going to be car wrecks, whether the drivers are people or computers. Jean-François Bonnefon, a psychologist involved in designing algorithms for self-driving cars, asks with the cool rationality of his trade, 'If it is unavoidable that some road users will die, which road users should they be?'[2] Self-driving vehicles cannot eliminate fatal wrecks altogether, but they can be programmed to make extremely fast choices among bad options. Once someone or something is making a choice about who will die, *it is no longer a technological question, it's a moral one*.

Given the choices vehicles will face, what is the right thing for them to do? We live in the age of Big Data, so Bonnefon's team turned to the wisdom of the crowd. In 2016, they launched an online game dubbed 'The Moral Machine'. Players were presented with a variety of situations involving a self-driving car in which a fatality was unavoidable, but which allowed players to prioritize who would be hit and who the car would avoid. They were also given the option of having the car swerve or just to stay the course. The game went viral and by 2020 millions of people had played.

The findings held no great surprises. Given various forced choices, players favoured humans over non-humans, fewer victims over more. They gave priority, in this order, to the baby, the little girl, the little boy and the pregnant woman. They also tended, slightly, to favour law abiders over law skirters, higher- over lower-status individuals, healthy over unhealthy, pedestrians over passengers.[3] Given the choice, they would avoid action altogether, letting the car continue straight on along its current trajectory rather than making it swerve away.

Let's pause over this last item. If someone will die no matter what you do, it seems that letting the car stay on its path is a way of saying, 'I don't want the responsibility of choosing victims, so I'll just let matters take their course; in effect, *I opt not to get involved*.' This is a version of a classic moral distinction between actively *killing* someone and passively *letting them die*. The end results may be the same, but *your own role* in the sequence of events is different. As we will see shortly, not doing anything seems attractive to many people.

This, of course, is a way for *me* to dodge responsibility for a morally troubling outcome. But it does so in a very specific manner. It doesn't just shift the blame to someone else. In effect, it tries to remove human actions from the picture altogether. By simply letting events unroll, it is (almost) as if I have nudged them out of the sphere of moral considerations altogether. Once trains cease to be culpable and self-driving cars just follow algorithms, it can seem that whatever happens is just an ethically neutral, if tragic, matter of cause and effect. When English law eliminated the deodand, it shifted certain kinds of deaths from being blamable (if not on a human being, at least on *something*) to just bad luck. It is as if the tragedy has crossed an invisible line between human choices and non-human happenings, or, let's say, purposeful actions and random chance. And yet we cannot

simply dismiss vehicles as participants in human moral life. Think of how deeply some people identify with their sports cars or VW bugs. Even for more dispassionate drivers, the car is, in some sense, an extension of its driver or even passenger. Is there so clear a line between them?

Does everyone even agree that the vehicle is just so much machinery? In 2015, while I was visiting my student Charles Zuckerman during his fieldwork on gamblers in Laos, we stopped by a Buddhist monastery. Just then, the owner of a small trucking company came by. It seems his trucks had been in enough accidents that he had arranged to have the monks perform a blessing ceremony. He parked one of his trucks facing the temple's front steps where the monks sat. The routine was much like blessing people dealing with misfortune. And it was definitely a serious matter, not mere custom. While the monks chanted, the blessings flowed along a string connecting their hands to the truck's steering wheel. The string also transmitted blessings to a bucket of water that would then be sprinkled on this truck and taken back to splash on the other trucks in the company lot. Clearly the ritual's energy was meant to flow to trucks, not persons. Was this businessman diverting bad luck, treating vehicles as responsible agents, squaring things with the cosmos, or seeking something else altogether? Ritual practices don't always require explanations to be effective, and quite possibly he couldn't tell you. But I am willing to hazard that the monks and the businessman do not draw the line between moral humans and morally neutral devices in quite the same way as the designers of the Moral Machine experiment – or those who played it.

We will explore variations on these themes over the chapters that follow. We will see that how things count for us ethically depends a lot on what counts for us as enough like a

human to have a social relation to. Along the way, we will find that *what counts as human*, where you draw the line, and *what lies on the other side*, are not stable, clear-cut or universally agreed on. The differences reflect diverse histories and ways of life. At the same time, if we listen carefully, we can sometimes hear those differences echo each other.

What If?

Experiments like the Moral Machine game are necessarily hypothetical. Happily, no one is going to die because of the player's decisions. When the designers ask you to imagine 'what if?', it can be like playing a game. But merely *thinking* about an imaginary situation is very different from *acting* in a real one.

What you ought to do and what you will do are hardly the same. I suppose most often this is because our actions don't always live up to what we imagine we would do in a given situation. But it can go the other way too. Once, when I was young, growing up in the dirty and dangerous New York of those days, I saw a thief snatch a nearby woman's bag while we were waiting for the subway. Impulsively, I grabbed it back and returned it to its owner. It was an instinctual action that took place without thinking (the fact that my girlfriend was watching might have influenced me too). But ten or so minutes later the gravity of the situation hit me, I blanched and my knees buckled. What a reckless and even stupid thing to do! I have no aspirations to virile heroism and honestly doubt I am so altruistic that I would have leaped in like that had I given it a moment's reflection. Who you are when you act and when you think things over can be very different people.

In a hypothetical situation, you are not personally affected by the outcome. We can see this in a paradoxical – if unsurprising – finding from the Moral Machine experiment. Suppose the only options are the car killing several innocent bystanders or sacrificing one passenger. Most players say the car should sacrifice the passenger. But naturally no one wants to be that passenger *themselves*.

Yet this commonsense response runs against the recommendation of many of the greatest ethical thinkers. One of the touchstones of modern Western moral philosophy is Immanuel Kant's 'categorical imperative'.[4] Because humans have free will, he said, they can choose what rules to follow. But what makes a rule *moral* rather than, say, efficient? If morality is not, say, just a subjective opinion or a selfish technique for getting what you want, then it should be universally valid. In other words, you should live according to a rule that you would want to apply to everyone. Or as my mother would say, if I dropped a sweet wrapper on the pavement, 'What if everyone did that?' The twentieth-century philosopher John Rawls argued the reverse is true as well: the rule you hold for others should hold for you.[5] It follows that if the right thing is to let the passenger die, so be it – even if that turns out to be *me*.

To see things this way is to take what I call the *third-person perspective*: the viewpoint of anyone at all, as if you were not directly involved. This, in effect, is what Kant is recommending. To make the right ethical choice, look at things from an objective distance. It is wrong for me to cheat during an exam, even if I really need that grade to get into medical school so I can heal the poor and oppressed. Why? Because it's wrong for anyone at all.

But most of the time our lives are carried out in the *first person*. The first-person perspective is how I most directly

experience the world. It also typically puts me face-to-face with other people. In every known language, when I speak in the first person ('I' or 'we'), I am usually addressing someone else: 'you', the *second person*. The second person in turn can switch roles and address me back. In other words, first-person experience is tied up with my relations to others.

At the end of the day, it does matter to me whether I am the victim of the fatal accident, or, indeed, whether I should feel guilty for surviving it. And it matters whether someone I know as 'you' – not just a 'him' or 'her' or 'them' – is the victim. When it comes to moral problems, my willingness to make the right choices even if it costs me something (time, effort, pain, money, even reputation) depends on my ability to be involved, to *care* about it.

Does it matter who is making the choices and caring about the results? The designers of the Moral Machine game knew better than to rely on their own moral intuitions in designing the algorithm for self-driving vehicles. They were sensitive to the problem of ethnocentrism, the risk that their results would be biased towards their own worldview, and not be universally acceptable. That's why they sought the wisdom of the crowd. Surely a game played by millions would yield some reliable universals. But wait – who plays computer games? It turns out that the participants were overwhelmingly males under the age of 35, with university degrees. They were people who had the time, resources and inclination to play online games. Are those young computer game-playing guys really the best guide to moral universals? Should we demand that the people we will meet in this book, the Yukon hunter or the Thai farmer, the Japanese shop clerk or English equestrian, reshape their respective moral worlds so they fit the results? Or is there anything the rest of us might learn from the hunter, the farmer,

the clerk and the equestrian? The only way to find out is to get to know them. That's why anthropologists do immersive, long-term fieldwork.

Runaway Trolleys

The Moral Machine game is a variation on the Trolley Problem. This refers to a famous thought experiment from the 1960s and 70s, which made the jump from moral philosophy to psychology and from there to popular culture, showing up in *New Yorker* cartoons, political satire, social media memes, television shows, movies and video games. Although part of its appeal seems to lie in its oddly morbid playfulness, it resembles some real-life dilemmas of medical triage and military situations when stark choices must be made between terrible alternatives.

The philosophers Philippa Foot and Judith Jarvis Thomson developed the Trolley Problem to clarify people's intuitions about responsibility and harm.[6] To keep things clear, the thought experiment is highly artificial. In its most basic form, it asks you to imagine that you see an out-of-control trolley hurtling towards five people. There is no time to warn them and no way to brake the trolley. The puzzle emerges from the two scenarios that follow. In one, you could pull a switch that diverts the trolley onto another track that has only one person on it, who will be hit. In the other, you could push a large man in front of the trolley; his bulk is sufficient to bring it to a stop, but in the process, he will be killed.

The objective outcome is the same in both cases: one life lost in order to save five. The utilitarian calculus that follows seems indisputable. Someone is going to die, and it should be

the large man: depending on which situation you are faced with, you ought to pull the switch or push the man rather than let the trolley continue towards the other five people. Whichever method you use makes no difference since the net outcome in either case is four survivors.

Yet when researchers try the Trolley Problem out with lay people (which in practice usually means university undergraduates or other highly educated residents of the wealthy, industrialized world, the so-called WEIRD folk), the results confound that calculus. It seems that most participants who would accept the first option recoil at the second. They would pull the switch but not push the man. (Remember that in the self-driving car problem, if faced with hitting a passerby or harming the passengers, many players would rather let the car take its own course than intervene, come what may.) What makes this so puzzling is that there is such a sharp difference in people's reactions to two options with identical objective outcomes. Something else must be at stake.

The debates around this get very complicated, but for our purposes the key difference in reactions turns on whether you are looking at the situation from the first-person or the third-person viewpoint. From the third-person perspective, diverting the trolley or pushing the man come out arithmetically equal and that's all that matters. But to imagine yourself doing this is to take the first-person position. You are asking, 'What if I were the one doing this?' You must visualize pushing someone to his death. And it brings out your *relation* to him. The pusher and the pushed. And soon, the living and the dead.*

* Some argue that pushing the man so violates our deep-seated injunction against killing that it overrides cold calculation. Since he dies in either scenario, however, what brings this repulsion to the fore for the pusher is taking

Even from the third-person viewpoint there seems to be something wrong about pushing the man, at least in much Western legal and moral thought. This has to do with how we think of human beings. When you push the man, it is in order to stop the trolley. You don't have a boulder at hand, so you are using his body to save others. It's different when you pull the switch. In that case, it is diverting the trolley that saves lives. Even if the large man were not on the track, the others would still be saved. You were not turning a man into an ad hoc trolley brake. It is just his bad luck that he happened to be in the way of the diverted trolley. Although you should hardly be indifferent to the man's death, in this case the trolley interposes itself between me and the victim.

In the Western tradition within which this debate takes place, moral philosophers tend to agree that humans should not be treated instrumentally. This is why a doctor should not just kill one patient to distribute her organs, even if it means saving numerous other patients (we will run into real-life variations of this problem in the next chapter). Again, Kant makes it very clear: a human should not be used as the means to an end – as moral subjects, humans are ends in their own right. This makes morality part of the very definition of being human.

Refusing the Problem

The Moral Machine project and the Trolley Problem ask us to see ethics in a very narrow way. In both cases, there is an

---

the first-person perspective. At any rate, a glance at the historical record should make clear how limited the injunction against killing is in practice. People do it all the time.

emergency. Something must be done quickly. There are clear starting and end points to the ethical situation with no back story and no long-term consequences for the person making the decision. There is just one decision to make, and it has a clear outcome which will become immediately apparent to everyone involved. There is only one morally relevant actor. That person is an individual agent, fully able to act, who is autonomous and unconnected to the others. The participants are all anonymous. The potential victims have no role other than to be at the receiving end of someone else's moral choices.

These are just some of the reasons that anthropological fieldworkers have shown little interest in highly schematic experiments like the Trolley Problem. Maurice Bloch is a rare exception. Bloch has spent a lifetime of repeated fieldwork in Madagascar, getting to know Malagasy villagers very well. He found that trying out the Trolley Problem in his fieldwork didn't get very far, and his reflections are revealing. He points out that, because philosophers and psychologists are usually looking for universals, they exclude anything they think might be a cultural norm. They tend to see culture as local, idiosyncratic and biased – something that obscures or distorts the underlying universals they hope to find. And if that's your assumption, then it shouldn't matter much that psychology experiments are usually carried out with WEIRD subjects like American university students.

As Bloch points out, these students treat the experiment as a familiar kind of puzzle which they enjoy, not a source of heart-wrenching tension. Matters were quite different for Malagasy villagers. They will not even consider the problem before they know if the victims are related to them, or how old they are. Listening to them, Bloch realizes that even if the Malagasy gave the same answers as the Americans, they have very different

(continued...)

Index