

Chapter 1:

Trouble in paradise

In Chapter 1, Bowles writes, “Over the coming decades, our industry will ask people to trust us with their data, their vehicles, and even their families’ safety. Dystopian science fiction has already taught people to be skeptical of these requests; unless we tackle the ethical issues that are blighting the field, this trust will be hard to earn.”

How do you feel sentiment around data and trust has changed over the past few years? Is the sentiment universal or different based off of geography or other identity markers?

In Chapter 1, Bowles describes two opposing viewpoints on the role of technology; Instrumentalist vs. Technological Determinism. The instrumentalist stance, “argues technology is just a tool, one that people can use for good or misuse for harm. Instrumentalists argue that since bad actors will always twist technology for evil, the only ethical recourse is to educate and plead for proper use.” Technological Determinism, “argues that technology is anything but neutral; instead, it’s so powerful that it moulds society and culture, acting more as our master than our servant.”

What are the strengths and weaknesses of each viewpoint? Is one better than the other? How do these differ from Philosopher Peter-Paul Verbeek’s Mediation Theory which states that, “technology is a medium through which we perceive and manipulate our world.”?

In Chapter 1, Bowles makes the argument that ,”Design is applied ethics.”

Do you agree or disagree with this statement. Why or why not?

Chapter 2: Do no harm?

In Chapter 2, Bowles discusses unintended consequences. He says, “According to the law of unintended consequences, there will always be outcomes we overlook, but unintended does not mean unforeseeable. We can—and must—try to anticipate and mitigate the worst potential consequences.”

In what ways does Artefact’s approach to Responsible Design, anticipate and mitigate potential consequences? How do those align with some of the concepts and examples Bowles uses to discuss this topic (i.e., externalities, algorithmic bias, data shadows, etc.)

In Chapter 2, Bowles talks about “The technocracy trap.” He writes, “Elected officials make laws; citizens slowly develop social conventions. But disruptive technologies often burst onto the scene without warning, before these social or legal norms can emerge.”

In what ways has disruptive changed social norms in ways that elected officials haven’t caught up to? What might be an approach to balance influence between government, society, and technology?

“If you live near a Whole Foods, if no one in your family serves in the military, if you’re paid by the year, not the hour, if most people you know finished college, if no one you know uses meth, if you married once and remain married, if you’re not one of 65 million Americans with a criminal record — if any or all of these things describe you, then accept the possibility that actually, you may not know what’s going on and you may be part of the problem.”

— Anand Giridharadas

In Chapter 2, Bowles talks about futurist, and uses the Futures Cone, a model to discuss probable, plausible, possible, and preferable futures (which cross between probable, plausible, and possible futures). He then mentions some criticism of the model, mentioning that, “Cameron Tonkinwise points out the direction of the beam depends on who’s holding the torch: in an unequal world, everyone starts from a different ‘now’. The idea of a preferable future is also loaded: preferable to whom?”

How do we account for this bias in our own work of creating “preferable futures”?

In Chapter 2, Bowles talks about Core Values. He writes, “Core values – essentially a list of the company’s stances and commitments – are a widespread and important vehicle for ethics, and usually carry senior support. Project teams can also create more localised rules, such as design principles that govern the design decisions within a particular product line or project. Strong core values and design principles taken to heart

are powerful tie-breakers for ethical dilemmas: in the event of moral emergency, consult the agreed tenets for guidance.”

What are Artefact’s core values, and how does this influence how you work and make decision on a daily basis?

Chapter 3: Persuasive mechanisms

In Chapter 3, Bowles discusses Coercion and nudging at the beginning of the chapter. Later on he builds arguments on persuasion and experimentation, and mentions, “If we can exploit someone’s weaknesses on cue, when does a nudge become a shove?”

What do you see as the different between Nudge Theory and Dark Patterns?

In Chapter 3, Bowles write, “Any project that learns from user behaviour is a user research project, yet the industry has tacitly chosen to exempt experimentation from research ethics. Users are given no right to withdraw from studies.”

What safeguards do we put on our own research in lieu of an institutional review board (IRB) that would happen in academia?

In Chapter 3, Bowles writes, “Designers haven’t interrogated the difference between enjoyable and habitual use, and the rhetoric of designing for delight has directly contributed to addiction.”

What methods could we employ to ensure we’re designing for the correct outcomes?

Chapter 4:

The data deluge

In Chapter 4, Bowles writes, “It [data] has become a river, a torrent. Data is the new oil: immense, uncontrollable, and with a nasty habit of leaking.”

Building upon the oil metaphor, how might we predict how our relationship around data might change in the future? What might we consider doing now to shape a preferable future around data?

In Chapter 4, Bowles claims that, “While we can simply tap into existing (albeit diminishing) reservoirs of oil, extracting data is not a passive act. The choice of what data to collect and what to omit, the technologies we use to collect and process it, and the techniques we use to analyse data are themselves laden with implicit assumptions and biases.”

What practices do you have in your design and/or research methods to ensure you’re collecting (or the product is collecting) the appropriate amount of data?

“Some theorists describe privacy as a right to choose seclusion, to decline a role in the public realm. Others define privacy as confidentiality, including legal scholar Richard Posner, who says what people mean by privacy is ‘concealment of information about themselves that others might use to their disadvantage’. These forms of privacy may no longer be sustainable. Someone who wants to fully conceal their personal data must essentially opt out of modern society... Perhaps it’s better to understand modern privacy as control and self-determination. Where possible, people should be able to disclose whatever information they want, to whomever they wish, whenever they want, and be able to reverse those decisions too. Privacy, then, is a hand on the dial, not just a padlock.”

In Chapter 4, Bowles discusses transparency around data. He claims, “In the name of seamlessness, designers have convinced users they have no business looking under the hood.”

How have you experienced this tension between seamlessness and transparency in your work? How did you balance transparency with understanding and value?

In Chapter 4, Bowles introduces another ethics theory: utilitarianism. He writes, “Since utilitarianism asks us to consider the overall sum of happiness, not just our own, utilitarians need a global and inclusive outlook. In the words of philosopher Henry Sidgwick, a utilitarian must consider ‘the point of view of the universe.’”

How does utilitarianism as an ethical theory add to, or contradict, human-centered design as an approach to our work?

Chapter 5: Seeing through new eyes

“The social signal of your principled refusal could in turn send ripples of positive influence through your peer groups.”

In Chapter 5, Bowles talks about speculative design, and how it can “be dangerous to even discuss potentially harmful technology.” He asks if this should be applied to speculative design too. “Isn’t there a risk that we introduce and normalise dangerous ideas to the very people capable of bringing them to life?” He goes on to reference an incident where a design fiction project escaped the walls of Google. Much of our work could be considered speculative, prototypes, or envisioning work.

How do we ensure this work is seen in the proper context? What do we do to ensure our work leads to an ethical future of whatever it might inform?

In Chapter 5, Bowles begins to consider AI and Robots as a new type of stakeholder. He writes, “When the concept of a user breaks down, so does user-centred design. Emerging technologies aren’t just products or tools: they’re often physical entities acting semi-autonomously; in other words, robots. Designers have to plan not only how these robots look and behave, but also their intended position in society.”

How might we consider ‘robots’ as a new stakeholder. What might research including robots, or AI, as a stakeholder look like?

“Technologists have a responsibility to help people form accurate mental models of changing technologies. Many countries require toy guns to be brightly coloured, or at least have an orange tip, so people will know the capabilities of the weapon. There’s a strong ethical case that smart and autonomous objects should be similarly distinguished, through either design or labelling, from their inert counterparts.”

In Chapter 5, Bowles introduces the concept of a counterfactual in relation to machine learning to propose a way to make algorithmic decision making more transparent. “...A counterfactual describes the smallest change that would have achieved a different outcome: ‘You were denied a loan because your annual income was £30,000. If your income had been £45,000, you would have been offered a loan.’ This statement doesn’t mean that everyone who earns £45,000 will get a loan; if it were that simple, you could replace your expensive algorithm with an ‘if... then’ statement. The counterfactual merely shows the nearest threshold at which you would have got a different outcome. This is straightforward

to find: once the system takes a decision, run it again with slightly different values until the outcome changes, then report what tipped the balance.”

Do you believe employing counterfactuals to algorithmic decision making would be an appropriate solution for understanding and transparency of how a particular AI works?

In Chapter 5, Bowles introduces the concept of Virtue Ethics, which is an approach more concerned with overall moral character. Bowles uses a general ethical test of “Would I be happy for my decision to appear on the front page of tomorrow’s news?”

What might be a strength and a weakness to a Virtuist approach to ethics?

Chapter 6:

You have twenty seconds to comply

At the beginning of Chapter 6, Bowles talks about the challenge of moderation at scale. He mentions that most companies, “waited far too long to establish and enforce proper rules of conduct, acting only once abuse became rampant.” These firms have found that moderation is difficult work, and have implemented algorithmic-based moderators alongside thousands of humans. However, Bowles states that, “Ideally, abuse is better prevented than treated; better to stop harm from even occurring than handle it after the event.”

What methods does Bowles mention to help prevent abuse? Are there any others that you might be able to imagine?

Bowles talks a little bit about the collection of data and governments. He mentions that so few tech companies have a plan for which government could be a threat to users. “The data we give to private companies is also, under the right legal and technical conditions, available to the state.”

What role might we have as citizens and designers in considering how data could be abused by governments?

Bowles talks about the ethics of encryption, and how this can cause a conundrum in a democratically elected government, and whether privacy is an unconditional right. “If we see rights as universal applications of moral law, we have to be consistent: if people have the right to privacy, so do terrorists, and restricting encryption is unsustainable.”

What should the expectation of privacy be, and how might we consider that in our own designs?

“The West has similar repositories of reputation and power, except we prefer them to be under corporate rather than state control. Credit agencies have logged our financial trustworthiness for decades, and in-car devices are reporting back to insurers on how we drive. Facebook holds a patent for a credit score that is affected by your friends’ scores, although the firm has as yet shown no intention to use it.”

Bowles mentions, “If a technologist thinks their company is about to seriously damage the public good, how should they object? The simplest response is refusal. As we know, ‘If I don’t, someone else will’ is no defence, meaning the projects we take on reveal the values we choose to live by.”

Are there any project you have worked on in the past that challenged your values? In hindsight, what would you have done differently? How might you recognize and make a different decision in the future?

“To live a moral life we must be willing to put ourselves at greater risk, if it’s safe to do so.”

Chapter 7: Software is heating the world

In Chapter 7, Bowles discusses a bit about the environmental impact of work, mentioning that no matter the ethical lens we use, “It’s clear we must protect a future from which we’ve borrowed too much for too long.” For today’s technology he mentions, “It’s helpful to look at the tech industry’s environmental impact across three categories: data centres, networks, and devices themselves.”

When doing your own work, how might these areas be a part of the process for consideration?

Bowles writes, “Audrey Shenandoah, clan mother of the Onondaga Nation, says, ‘There is no word for “nature” in my language. “Nature” in English seems to refer to that which is separate from human beings. It is a distinction we don’t recognize.’”

How does our language and culture affect our frame of the world, and thus what design interventions we might consider?

Bowles talks about the potential for scarcity in the future, and the potentiality that, “The tech industry may have to set its sights lower on the hierarchy of needs than it does today, working on the challenges of safe housing, clean water, and fair sharing of scarce resources.” He also proposes *jugaad* as a guiding design principle, which describes a, “do-it-yourself innovation based on what you need and what you have right now.”

How is this at conflict with the current models, approaches, and values we often design in and for?

“The ethical life is not necessarily a more enjoyable life. Study ethics and you’ll soon notice the contradictions in your decisions. You’ll become your own biggest moral critic, although that’s not to say others won’t criticise too; taking an ethical stand can also make your own imperfections more visible to others. I too have my ethical flaws. I still eat meat; I still rack up frequent-flyer miles. I can hope my other choices offset the environmental and ethical damage these acts do, but I can’t justify them.”

Chapter 8:

No cash, no jobs, no hope

Bowles writes about the world being on the brink of a new industrial revolution, one that is centered around many of the new emerging technologies today. He writes, “The advent of cognitive automation means that almost every job is potentially under threat, causing widespread, permanent disemployment by machines.” He cites AI and machine-learning experts estimating that high-level machine intelligence is about 45 years out.

If we think of environmental considerations to be some of the biggest faults of previous industrial revolutions. How might we try to consider some of the affects of widespread, high-level machine intelligence on future generations?

Bowles writes about the tension between consumption, capitalism, and ethics. He writes, “Businesses that employ a triple bottom line, placing equal priority on social, environmental, and economic returns, tend to be mindful of externalities and unintended consequences, and more sympathetic to ethical appeals. Fully private companies can sometimes resist short-term profit demands more easily than publicly listed corporations, while some companies make deeper commitments, such as B Corps, who publicly pledge to put social and environmental good first, finding an important balance between profit and purpose.”

How does our company measure it’s bottom line? Is it reflective of our values?

“Stephen Hawking conceded “the rise of powerful AI will be either the best, or the worst thing, ever to happen to humanity. We do not yet know which.”

Bowles writes about meaning, consciousness, personhood, and anthropomorphism related to machine and human relations. He writes, “If a machine ever approached consciousness we’d have to completely rethink our interactions with it.” He also writes, “If we find ourselves in a disemployed, jobless future, how will people find meaning? People’s intrinsic and social value is often linked to their work.”

How might you imagine the advancements of machines to change the worlds’ current paradigms? What might we do in advance to prepare?

Chapter 9: A new tech philosophy

“...be sceptical of empathy. Banal overuse has devalued the concept of empathy in design; I’ve intentionally avoided the term in this book. Of course we should try to understand other people’s life experiences, but empathy alone is no shortcut to ethical nirvana. Psychologist Paul Bloom argues that empathy really only extends to in-groups: we empathise with people who we imagine to be like us, or who we find attractive, but not those we consider to be unlike us. Empathy can also mislead us into prematurely thinking we’ve fulfilled our ethical duties.”

In Chapter 9, Bowles makes recommendations of how we might move forward. He writes, “An individual can spark an ethical movement, but it takes a group to sustain it. Aim to be an ethical facilitator rather than an oracle; not the casting vote but a catalyst, giving people new ways to think about and discuss ethics, and engineering the time and space for these conversations to happen.”

How might we serve as ethical facilitators in our work. Can you think of a specific example using a project you’ve worked on recently?

“Your ethical decisions are yours alone. That should daunt and excite you equally. With rare exceptions, such as companies that genuinely have no interests other than profit, I think any individual can make an ethical difference in any organisation.”