THE ETHICS CENTRE

ETHICAL BY DESIGN: PRINCIPLES FOR GOOD TECHNOLOGY

DR MATTHEW BEARD & DR SIMON LONGSTAFF AO

TECHNICAL MASTERY
DIVORCED FROM
ETHICAL RESTRAINT
IS AT THE ROOT
OF ALL TYRANNY.

PR00.

OUGHT BEFORE CAN

The fact that we can do something does not mean that we should.

There are lots of possible worlds out there – lots of things that could be made or built. Ethical design is about ensuring what we build helps create the best possible world. Before we ask whether it's possible to build something, we need to ask why we would want to build it at all.

PR01.

NON-INSTRUMENTALISM

Never design technology in which people are merely a part of the machine.

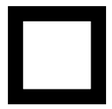
Some things matter in ways that can't be measured or reduced to their utility value. People, ecosystems, some kinds of animal life and political communities shouldn't be used as tools that can be incorporated into design. They must be the beneficiaries of your design, not elements of a machine or design system.

PR02.

SELF-DETERMINATION

Maximise the freedom of those affected by your design.

Technology is meant to be an extension of human will. It's meant to empower us to achieve goals we otherwise couldn't. Technology can't achieve this goal if it interferes with our freedom. We need to make design choices that support people's ability to make free choices about how they want to live and engage with technology.
But remember: maximising freedom doesn't always mean maximising choice – sometimes too much choice can be paralysing.



PR03.

RESPONSIBILITY

Anticipate and design for all possible uses.

Technology is usually designed with a specific use case – or set of use cases – in mind. Problems often arise when users deviate from the intended use case. It's entirely possible to predict the different ways people will use our designs, if we take the time to think it through. Failing to imagine alternate uses and their implications is risky and unethical. Doing so can alert us to potentially harmful uses we can safeguard against, or potential benefits we can maximise through good design.

PR04.

NET BENEFIT

Maximise good, minimise bad.

The things we build should make a positive contribution to the world – they should make it better. But more than this, we should also be mindful of the potentially harmful side-effects of our technology. Even if it does more good than bad, ethical design requires us to reduce the negative effects as much as possible.

PR05.

FAIRNESS

Treat like cases in a like manner; different cases differently.

Technology designs can carry biases, reflect the status quo or generate blind spots. The implication of this can mean some groups of people are treated negatively on the basis of irrelevant or arbitrary factors such as race, age, gender, ethnicity or any number of unjustifiable considerations. Fairness requires that we present justifications for any differences in the ways our design treats each user group. If some groups do experience greater harm or less benefit than others we must consider why this his the case and if our reasons are defensible.

PR06.

ACCESSIBILITY

Design to include the most vulnerable user.

Whenever we identify intended user profiles and use cases, we also act to isolate non-users from the design consideration. This creates the risk that design excludes people who might benefit were they considered in the process. Design can reinforce social disadvantage, or it can help people overcome it. But it can only do this if we bear in mind all the possible users, without dismissing some groups as 'edge cases'.

PR07.

PURPOSE

Design with honesty, clarity and fitness of purpose.

Design is, in one sense, a promise. You are promising to solve a problem your users are experiencing. Like all promises, you should honour this promise. You must be honest and clear about the ability and limitations of your design. Moreover, your design should be tailored to the problem it's trying to solve – and be intended to solve a genuine problem. Good design serves an ethical purpose and does so in efficient and effective ways.

CONTENTS

ABOUT THE AUTHORS

Dr Matthew Beard is a husband, father and moral philosopher with an academic background in both applied and military ethics. Matt has taught philosophy and ethics at university level for several years, during which time he has also published widely in academic journals, book chapters and spoke at a number of international conferences. His work has mainly focussed on military ethics, including the ethics of military technology, a topic on which he has advised the Australian Army. He has also published academic pieces on torture, cyberwar, medical ethics, weaponising space, sacrifice and the psychological impacts of war on veterans. In 2016, Matt won the Australasian Association of Philosophy prize for media engagement, recognising his "prolific contribution to public philosophy". He regularly appears to discuss a range of ethical issues on television, radio, online and in print.

Dr Simon Longstaff AO began his working life on Groote Eylandt (Anindilyakwa) in the Northern Territory where he worked in the Safety Department of the then BHP subsidiary, GEMCO. He is proud of his kinship ties with members of the island's Indigenous community. Following a period studying law in Sydney and a brief career teaching in Tasmania, Simon undertook postgraduate studies in philosophy as a Member of Magdalene College, Cambridge. He commenced his work as the first Executive Director of The Ethics Centre in 1991. He is a Fellow of CPA Australia and in June 2016, was appointed an Honorary Professor at the Australian National University – based at the National Centre for Indigenous Studies. Formerly serving as the inaugural President of The Australian Association for Professional & Applied Ethics, Simon serves on a number of boards and committees across a broad spectrum of activities. He is a Fellow of the World Economic Forum. In 2013, Dr Longstaff was made an officer of the Order of Australia (AO) for "distinguished service to the community through the promotion of ethical standards in governance and business, to improving corporate responsibility, and to philosophy."

INTRODUCTION

Recent technological change has transformed almost every part of life. Today, technology influences our relationships, decisions, desires and the way we experience reality. In almost every sector people list 'emerging technology' among the most pressing ethical challenges they face.

The explosion of new technologies has led the World Economic Forum to define our current era as part of the 'Fourth Industrial Revolution'. They're right to say we're living in a revolutionary age, but is it a good thing or not? Revolutions can be opportunities for growth, but they can also open the floodgates to disaster.

Previous revolutionary eras often spiralled out of control, betraying the ideas of their founders. Today, we have a rare and fleeting opportunity to seize responsibility for our future. Will we use technology to shape the kind of world humanity deserves? Or will we allow it to shape our decisions and thus, our future? Will technology serve our goals, or will we serve it?

For some, technology is a liberator. It might free us from unpleasant aspects of life: work, illness, difficult relationships and complex decisions. For others, technology is a destroyer. It will undermine humanity, perpetuate inequality and render us 'slaves to the machine'.

Whether we come to see technology as 'the hero' or 'the villain' depends on the choices we make. We get to design and deploy technology. We determine what limits we place on it. It's up to us to define what counts as 'good' and 'bad' technology.

It's tempting to think of the ethics of technology as being all about killer robots, nuclear power and other headline-grabbing topics. In reality, these are the fields where ethical reflection tends to be best. For example, the scientists involved in the Manhattan Project were acutely aware of the ethical choices they faced. Scholars disagree about whether they made the right choice, but evidence suggests they were alive to the moral issues.

Often, this is not the case: the stakes aren't obvious and the harms are hard to foresee. This is a major area of ethical risk: technological design processes that don't pay attention to ethics at all. Our paper's primary intention is to guide technology design when the ethical issues aren't obvious.

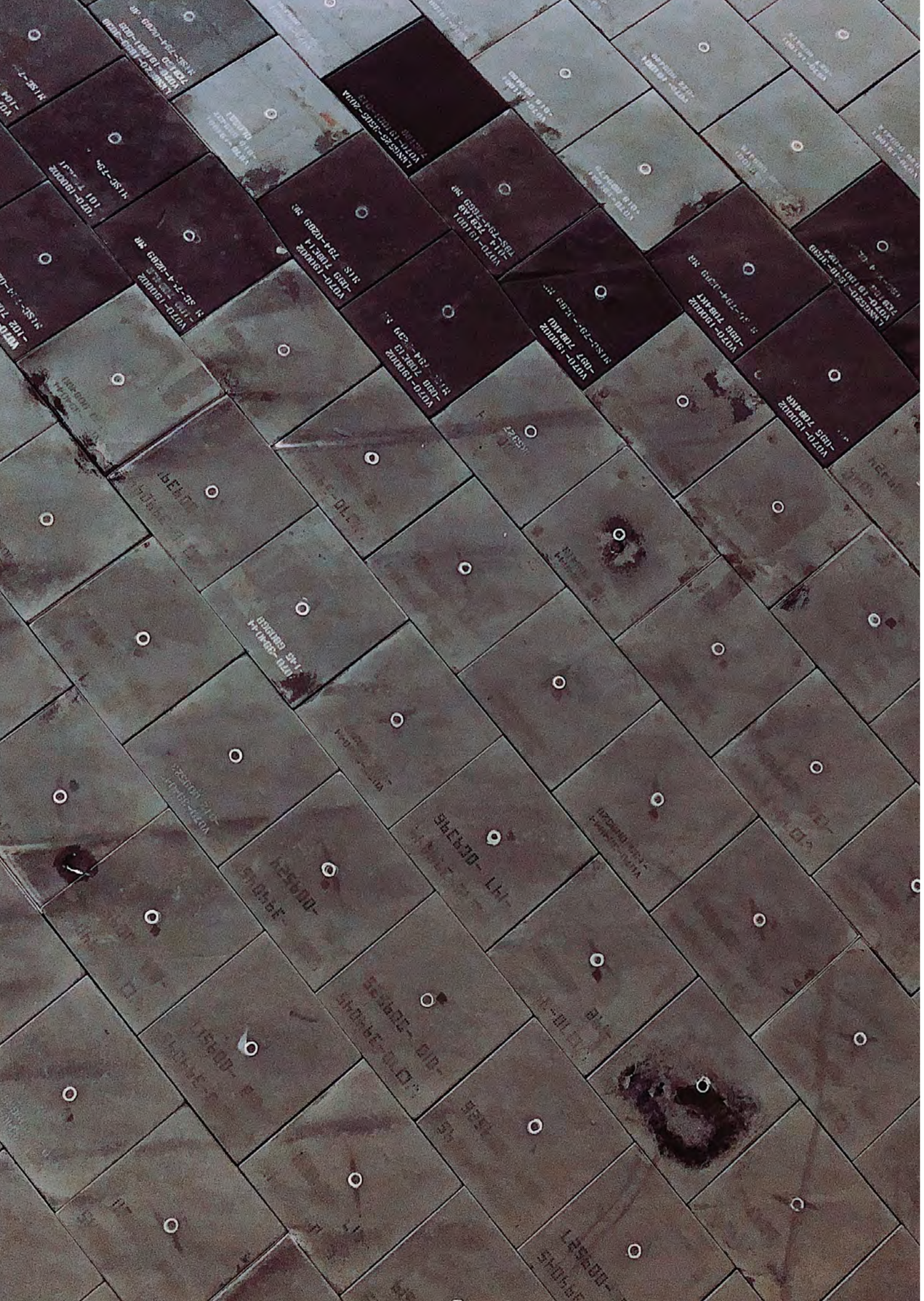
How do we ensure that the technology we create is a force for good? How do we protect the most vulnerable? How do we avoid the risks inherent in a belief in unlimited progress for progress' own sake? What are the moral costs of restraint – and who will bear the costs of slower development?

This paper addresses such questions by proposing a universal ethical framework for technology. We argue that ethical principles should inform the design, development and deployment of new technologies. They might also serve as a standard to test whether new technologies pass the 'sniff test' or not.

These principles apply universally. They are not limited to any particular industry or sphere of life – they can help us pass judgements on algorithms, synthetic biology, wheelbarrows, weapons and everything between.

We begin by providing some of the philosophical backdrop to our thinking and break down some common thinking about technology. We then introduce a framework for managing the ethical challenges of technology, centred on a set of philosophical principles. We also highlight specific design challenges and offer some examples of ethical solutions. Finally, we highlight some unresolved issues.

If ethics frames and guides our collective decision-making, we can ensure we reap the benefits of technology without falling foul of avoidable, manageable shortcomings.



A NOTE ON IMPLEMENTATION

Ensuring these principles are implemented will require collaboration across various industries.

Specifically, it will be necessary for technical experts to provide advice on how to give effect to these principles within the technology themselves. This is an especially important point. Throughout this paper we draw attention to some fundamental principles. We haven't provided an exhaustive list of the ways in which these principles can be implemented. We have taken this approach for two reasons:

01. **The designers and developers of technology should accept responsibility for their design decisions – including the technical means by which artefacts bring ethical principles to life. For us to specify a set of 'rules' would be to diminish this sense of responsibility, it risks creating a culture of compliance rather than a culture of genuine responsibility.**

02. **This paper has been written to address all forms of technology – ranging from biotechnology to household appliances. There is unlikely to be any single means for giving effect to ethical principles in all forms of technology. We believe designers and developers are better placed than ethicists to determine the best ways of implementing the principles, and so they should have the freedom to do so. However, they need to demonstrate publicly how the principles have been applied. For example, what would an 'off-switch' look like in the context of a piece of synthetic biology? This is not a philosophical question: it is a practical and technical one that will require the input of subject matter experts.**

Various political and legal jurisdictions will also be required to determine how these principles might inform or be supported by legislation or regulation. Should technology developers be compelled to adhere to these principles? If so, which ones should be obligatory and which should be voluntary? How can we ensure potential users of technology understand these principles and are able to know which ones have been implemented in the design of any particular piece of technology?

We invite industry experts, academics, regulators and all others to join in a global conversation about how to make these principles a universally known and embedded system within the language, logic and design of technology so that all of humanity shares the tech sector's confidence that technology is indeed a force for good.

HOW TO READ THIS DOCUMENT

This guide is intended to be used practically. It is not a theoretical document – we want it to be used by designers to help them refine their processes and choices to incorporate more fully ethical issues.

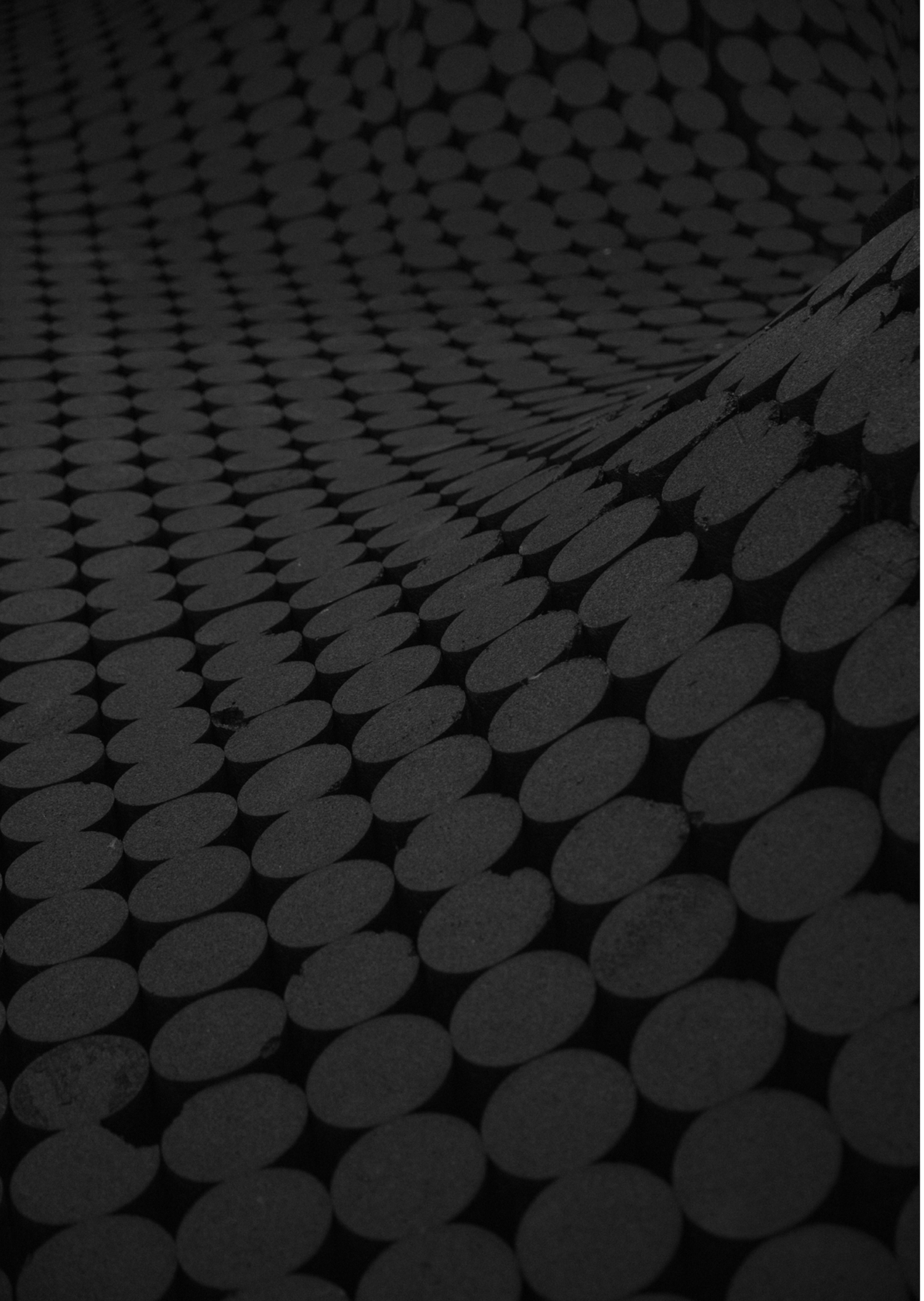
However, we do think it is valuable to outline some matters that will strike some people as theoretical. These appear early on in the work – in the background information on philosophy, ethical theories and the nature of technology.

We think practically-minded people would find this information useful and interesting, but they are also included in the interests of 'showing our work' – which is one of the ethical recommendations we make regarding technological design. We want people to understand how deeper views about ethics and philosophy are informing our thinking and recommendations.

Understanding the ethical smokescreens and how to respond effectively to them will be useful for those trying to advocate for ethical decision-making within their teams, industries and organisations. We've tried to outline some of the major barriers to accepting ethical responsibility and taking on ethics as a serious part of the design process. Hopefully, this will provide people with some responses to some of the arguments they often hear in opposition to ethical thinking.

With that said, it is possible to begin reading this document from the Governing Principle onward. We believe the principles we've outlined speak for themselves, and even those who might not agree with our philosophical framing will find a great deal of value in the principles we've identified. More practically oriented people may prefer to read this document backwards: looking first at the principles we've identified, and then understanding how we've arrived at them.

We would also ask that you read this document critically. It will take a village to resolve the ethical issues of technology design, and this contribution is intended to be challenged, improved upon and refined with contributions from other experts, practitioners and the array of knowledge and experience they have to share. We would therefore invite you to share any other thoughts, challenges or opportunities with us. You can send us an email at tech@ethics.org.au.



PHILOSOPHICAL BACKGROUND

Our philosophical starting point is that humanity and human beings matter in a special way. Many things deserve our moral attention; humanity deserves more of it than most other things.

This is not a new idea. It appears in writings across cultures and times. There is no philosophical consensus on why we matter in this way. For some, it is our conscience – our ability to reason and act against our instincts and self-interest – that makes us unique. For others, a concept like human dignity captures what makes humanity important. Some believe there is a *je ne sais qua* to humanity – we know it's particularly important, even if we can't say exactly why.

Not everyone agrees that humanity is special in any way. Some view this belief as nothing more than a kind of prejudice, arrogance or hubris. They see a kind of moral vanity in the idea that we're special – especially in the face of the scale of the universe and diversity of the natural world. However, even those who would question whether humans are special in any way would likely agree that particular aspects of humanity have value and ought to be defended – even if not at all costs. These aspects might include: our capacity for community, freedom of conscience, life, happiness or the ability to build a life on our own terms.

We will take it as granted that some aspects of humanity are worth defending. What's more, we start by saying humanity itself – and human beings – have value and deserve our care for and respect, no matter the cost. Not everyone will agree with this. Those who wish to reduce everything to a measurable unit of value will struggle here. Those who believe that the end of humanity wouldn't be any greater a loss, to the universe, than that of any other creature will have objections to our starting point. But we hope that even they will find plenty to like in our conclusions – which we argue every reasonable person should be comfortable adopting.

This is because we have drawn on a range of different ethical perspectives to find a 'reflective equilibrium'. Reflective equilibriums aim to bring a variety of moral judgements and perspectives into comparison. Through reflection, we refine our views and make them consistent with one another.

If done well, reflective equilibriums satisfy a broad range of positions and therefore appeal to many (and sometimes all) people. Is your ethical focus on getting the best outcomes, being of good character, living in harmony with others, adhering to universal principles or upholding human freedom? Whichever speaks most to you, these principles will capture key elements of what you find most important. They will also provide guidance on bringing your ethical priorities into harmony with others.

WHAT IS ETHICS?

One of the things that distinguishes humans from other creatures is our freedom to decide how we act.

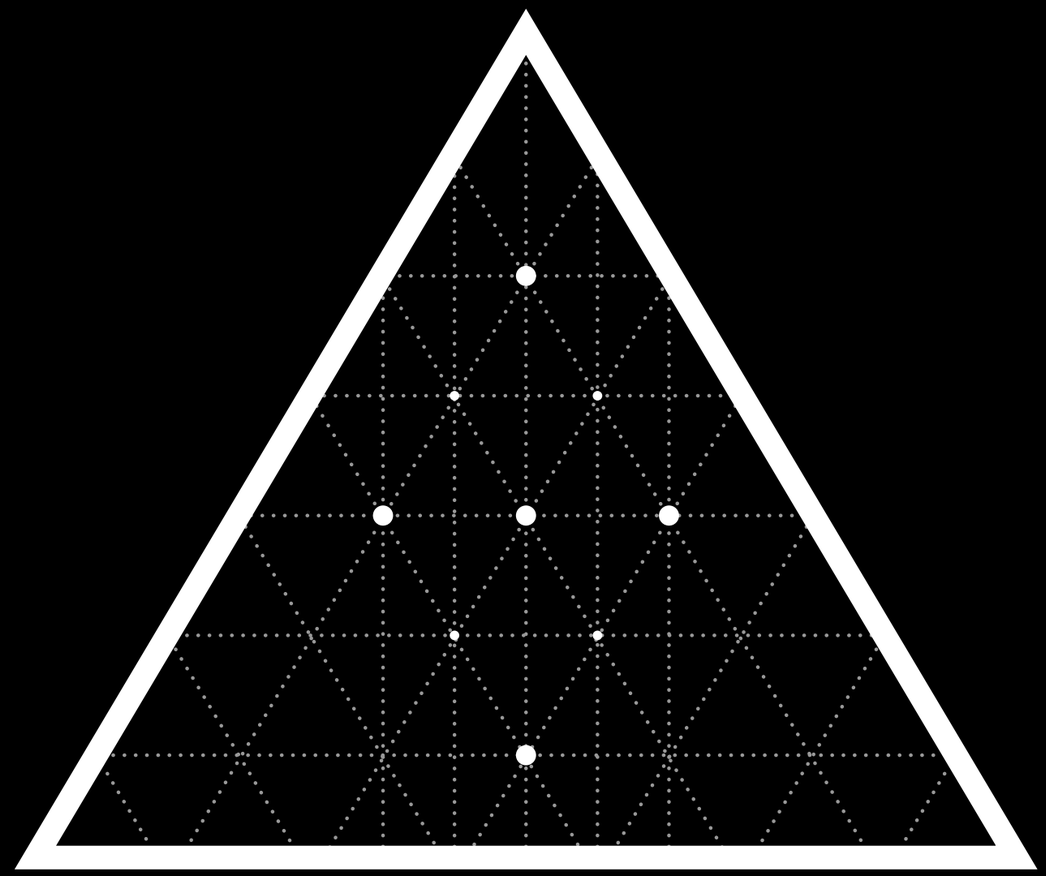
Whatever choice we make, we could have done otherwise if we'd chosen differently. Although nature, nurture and a range of factors influence us, what we do is at least partly up to us. The Danish philosopher, Soren Kierkegaard, captures this sense when he describes the anxiety of standing on the edge of a cliff. It's not only a fear of heights that worries us. Kierkegaard notes that the only thing that prevents us from falling into the void is us. Our decision not to jump keeps our feet on the ground.

Ethics is only possible because we have this ability to choose. So often we describe the world: what is likely to happen, what might happen or what will happen. Ethics allows us to judge the world - what should happen? Of all the ways you might act, which is the best? Which of all the possibilities should you bring into reality? What ought one to do? That's the question ethics seeks to answer.

You can only answer that question if you first bother to ask it. It can be more comfortable, safe and common to do what everyone has always done, to pass the buck to other people or stick with the status quo. Ethics asks us to take responsibility for our beliefs and our actions and to live a life that's self-consciously our own.

Ethics isn't the only field that tries explaining what the 'best' decision might look like. Economists might describe the best decision as the one which leads to the most wealth creation; artists might prioritise the most creative option; Machiavellians might prefer those means that advance or protect personal power – the list goes on. However, lying beneath the surface of all such judgements are fundamental beliefs about what is good or right. That brings us back to ethics which, in matters of judgement, can only be escaped if we never make a conscious decision at all.

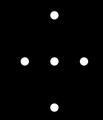
The field of ethics inclines us towards options that best achieve what is 'good', 'right' and consistent with our goals. At The Ethics Centre, we refer to these as 'values', 'principles' and 'purpose'.



PURPOSE

Is our reason for being.

It helps to explain and animate one's choice of core values and principles.



VALUES

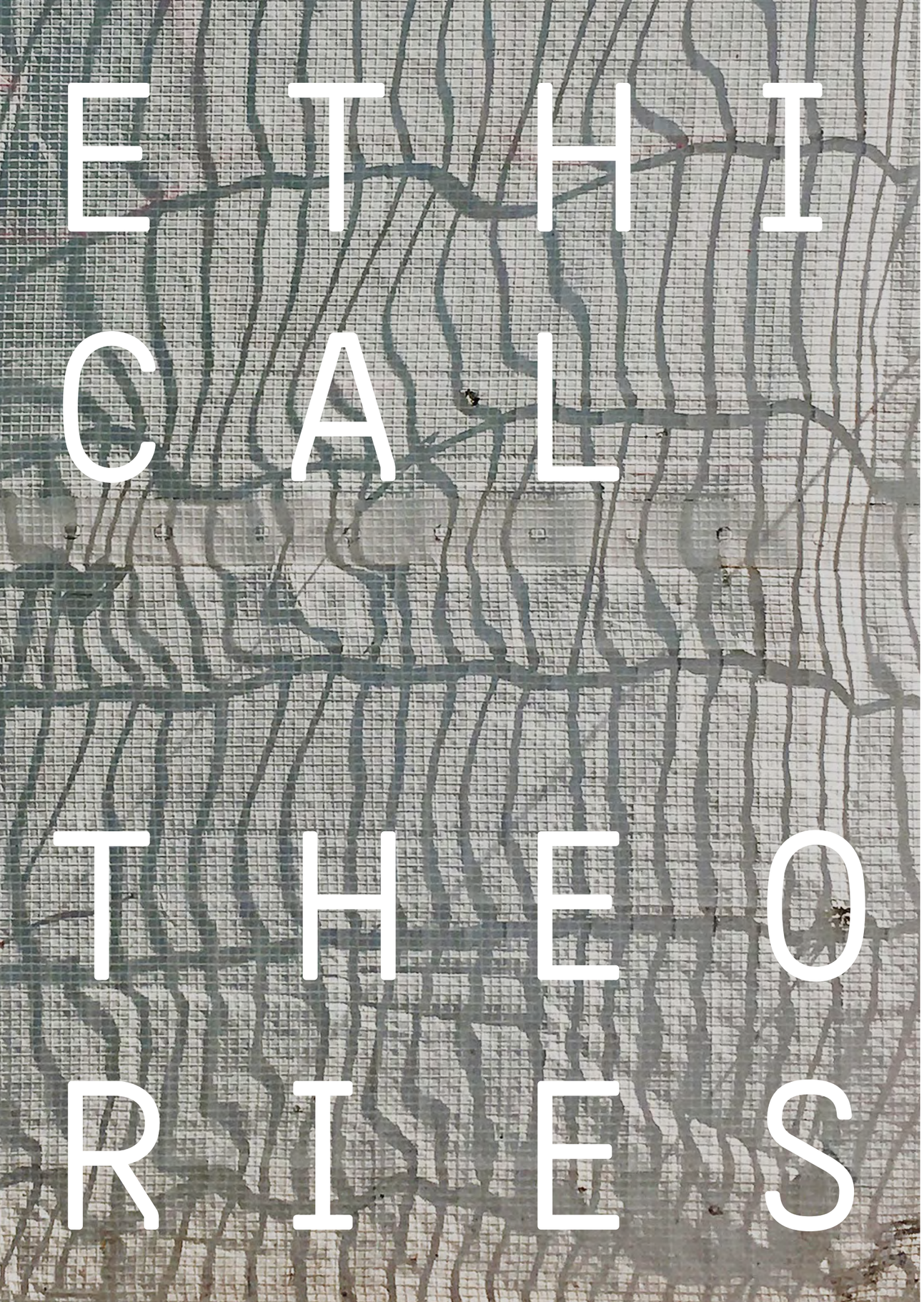
Identify what is good.

They are the things we strive for, desire and seek to protect.

PRINCIPLES

Identify what is right.

Outlining how we may or may not achieve what is good.



Consequentialism

Consequentialism is a broad school of philosophy. The basic, unifying belief is that for ethics, outcomes are the only things that truly matter. It suggests that if you want to do the right thing, then you should ensure that action has more positive effects than negative ones.

For many people, this mode of thinking is intuitive. It informs the practice of cost/benefit analyses and the popularity of pros/cons lists. There are a range of schools of thought within consequentialism, with different views of what counts as a 'good' outcome – is it pleasure? The satisfaction of preferences? The common good of society?

There are a few common tenets of consequentialism. First, that ethical value is measurable. For example, consequentialism needs to be able to quantify the value of 'blissful ignorance' and 'hard truths' to determine which is preferable in a given situation. Second, why you do something is less important than what you do. If you achieve good outcomes for selfish or malicious reasons, that matters less than achieving bad outcomes with good intentions.

Finally, consequentialists say that no person should matter more or less than any other when making ethical decisions. As such, we should not give special weight to our own interests or to those to whom we are closest – like our family and friends. Instead, we should treat every person's pleasure, preferences and interests equally, achieving as much good or avoiding as much harm as possible.

Deontology

Deontology is usually framed as a rival school of thought to consequentialism. It says we should focus on doing what is right – fulfilling our obligations and duties – come what may. The most widely-recognised form of deontology comes from the German philosopher, Immanuel Kant. He thought we could discover the laws of morality through reason and that as a result, all moral laws must make sense from a logical standpoint.

He thought that we may only act in ways that we could logically require of every other person. For instance, Kant believed promise-breaking was always wrong. If we allowed any person to break their promises, we would not only undermine the system of trust that underpins promise-making, we would make the concept of promise-making incoherent and meaningless. Universalising such behaviour would make it impossible. This contradiction is enough for Kant to conclude that promise-breaking is always wrong – irrespective of the circumstances or consequences.

Kant also believed a person's ability to reason gave them special moral status. As persons, we have dignity: we cannot be treated as tools or traded for a price. We deserve respect because we have infinite and intrinsic moral worth. Deontologists insist we must always treat others in ways worthy of their moral status, even if it leads to bad outcomes. For example, a pure deontologist would refuse to sacrifice one innocent person to save the entire world.

Teleology

Teleology concerns the 'ends' or 'goals' (telos) we ought to serve. Teleology finds its roots in Aristotle. It argues that we cannot make an ethical judgement about something without knowing its ultimate purpose. Aristotle thought the ultimate purpose of human beings is to flourish. As such, he said we should act in ways that are supportive of human flourishing and avoid doing things detrimental to it. Teleology challenges us to align methods, means and operations with their ultimate reason for being. We have to be explicit about the purposes our activities, creations and institutions serve.

Virtue

Virtue theory and teleology share a common intellectual history. Like teleology, virtue ethics finds its roots in Ancient Greek thinkers like Socrates, Plato and Aristotle. Virtue ethicists say we should make decisions based on the way those choices will shape our character.

They believe our actions display and are shaped by either positive or negative qualities – virtues and vices. Virtuous or vicious actions shape our character such that, over time, we become more likely to repeat those behaviours in the future. For example, if you routinely take the easy option then, over time, you will acquire the vice of laziness. It will come to define part of your character, meaning it will progressively become harder to put in the hard work when needed. Virtue ethics is about embodying the traits and practices of an ideal person in the knowledge that, over time, that's who we will become.

Aristotle argues for a kind of 'practical wisdom' (phronesis) that enables us to see things as they are – free from the distortions caused by vice, bias or social conditioning. The virtuous person makes clear choices. They start by following the example of wise mentors – learning, along the way, how to discern what is good and right for themselves.

Aristotle sees all virtues falling on what he calls the 'golden mean' – a point of balance between extremes. For example, the virtue of honesty stands on a point between dishonesty and tactlessness. That is, the honest person commits to telling the truth at the right time and in the right way.

Contractualism

Contractualist theories, also known as social contract theories, see the exercise of power as legitimate if it comes from the willing consent of citizens. Contractualists seek to justify and explain why the state has duties to its citizens and vice versa. The social contract outlines the respective rights and responsibilities of both citizens and the state.

Like any contract, the social contract involves a kind of exchange of services. Citizens surrender some power and liberty to the state in return for its guarantee of security and civil liberties. Social contract theorists insist that the social contract is the basis for the entire legitimacy of a government. The State has legitimacy only when its people grant it authority to exercise power.

Today, there is a growing sense that commercial enterprises should also form and maintain a kind of social contract. The right to operate as an organisation also needs people's trust, a sense of legitimacy and a commitment to the public interest. In these cases, freedom to operate also depends on the willing, informed consent of the general public – sometimes as investors, or employees or suppliers or just as citizens who provide the infrastructure that make commerce possible. This means organisations have an ethical obligation to fulfil their share of the bargain or, if not, cease to exist.

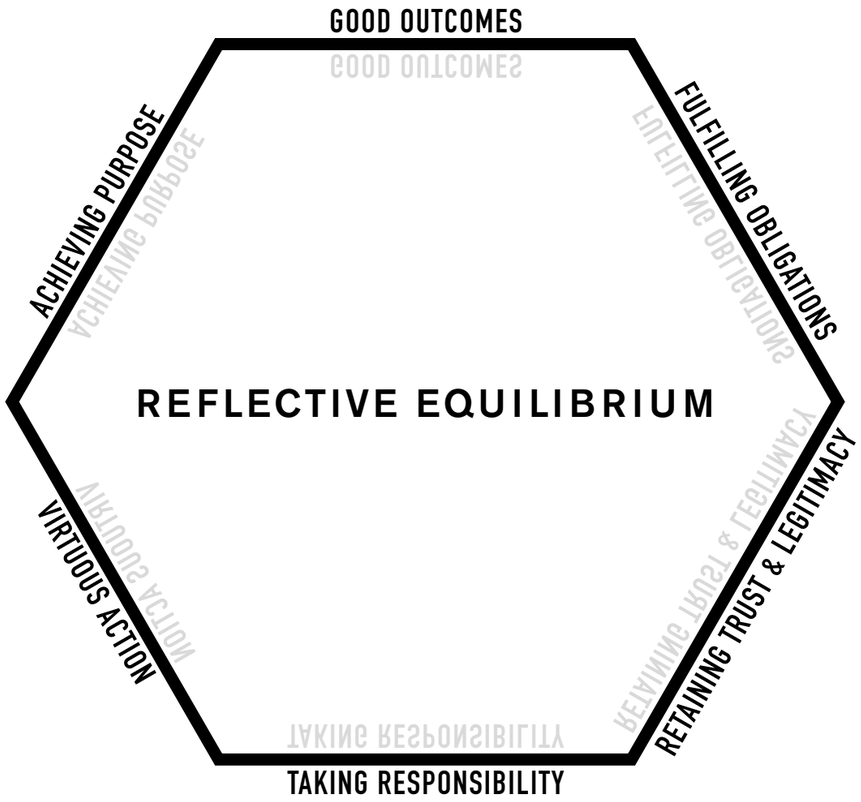
Existentialism

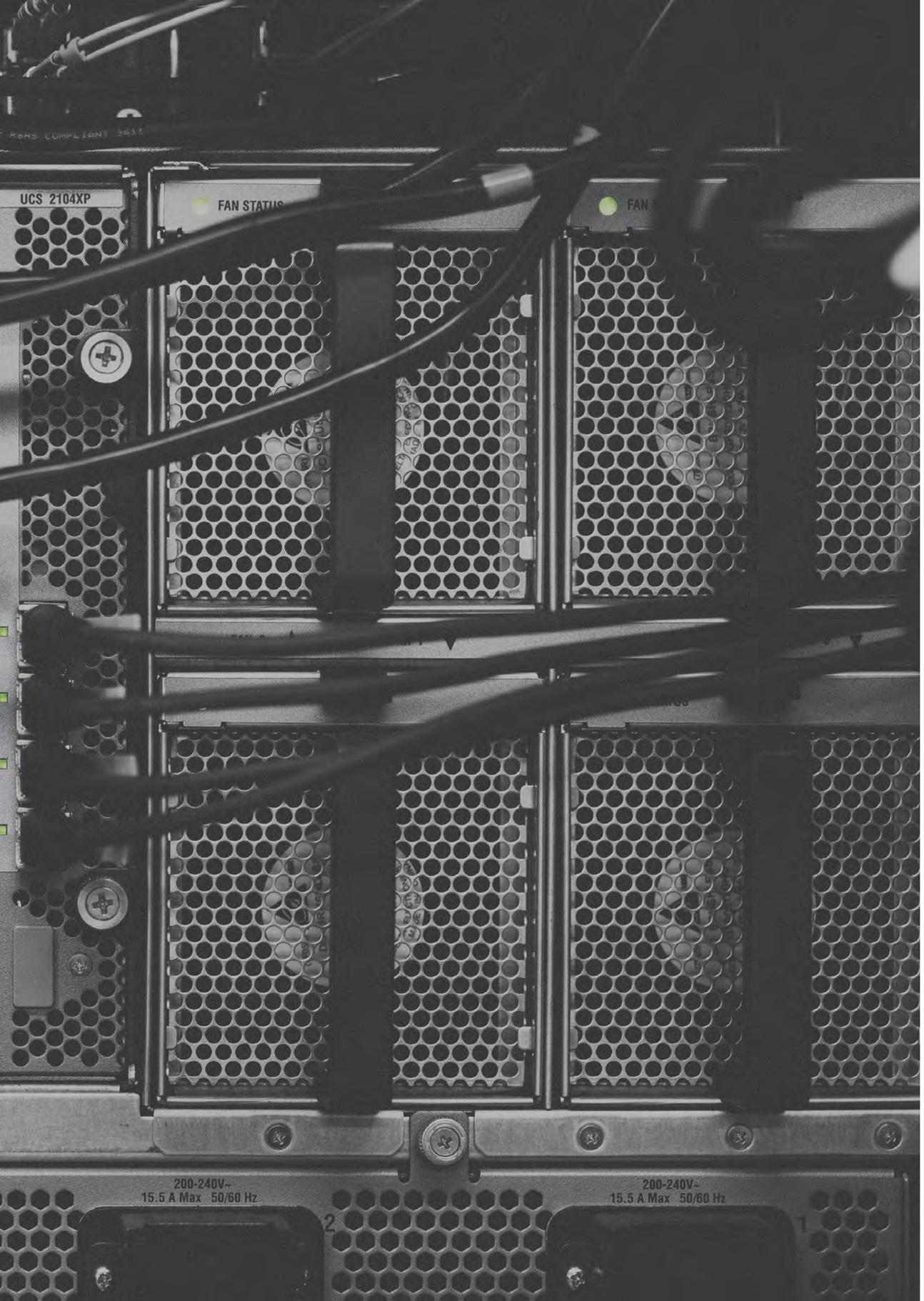
Existentialism says we should live in a way that responds to the basic facts of our existence. For existentialists, what defines human beings is that we are 'radically free' to act as we wish. As a result, we're also 'radically responsible' for the choices we make. The idea of radical freedom comes from French philosopher Jean Paul Sartre's claim that "existence precedes essence". He believed we aren't born with any pre-destined characteristics. Instead, we are completely free to choose who we will become.

For existentialists, this is not a pleasant thing. Being free and responsible is hard. It leads many to outsource their freedom and responsibility to others. We might leave a tough decision to our manager, let someone else decide what to order for dinner, or act the way other people expect of us. For existentialists, any act that denies our freedom, agency and responsibility is an act of 'bad faith' – an attempt to escape from our freedom and responsibility. We alone bear responsibility for our lives and our choices: avoiding our freedom has a price.

We have now provided an understanding of the general landscape of ethics.

These theories each call attention to different issues relevant to making good ethical decisions. Ignoring one or more of these approaches is likely to generate significant ethical blindspots, resulting in harmful outcomes, breaches of trust, morally unsatisfying work or any number of other moral issues we should seek to avoid in technological design.





WHAT IS TECHNOLOGY?

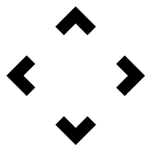
We understand technology in two related ways.

First, as things that extend our capacity to act in and on the world and second, as a distinctive form of thinking that shapes how we live.

Techno-things

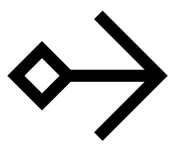
In the first sense of the concept, technology is an instrument of human will. We can talk about *technologies* – a series of artefacts that provide new, efficient and effective ways of acting.

For the sake of our discussion, **technological artefacts** are tools designed by people in order to mediate human engagement with the world, in order to achieve some goal. The defining characteristics of such artefacts are that:



THEY ARE MADE

THEY ARE USED



THEY SERVE SOME GOAL

PRIMARY ETHICAL QUESTIONS

From these characteristics, we can identify a few primary ethical questions.

FIRST, the fact someone has to *make* a technological artefact means they bear initial ethical responsibility for those artefacts – what they do and how they do it.

SECOND, we *use* artefacts to engage with the world. Understanding this means part of the ethics of technological artefacts concerns *how* they engage with the world and what *effects* they have.

From this, we can summarise the ethics of technological artefacts into two categories:

01.	**UNJUST GOALS**
	The people who build an artefact can intend for it to serve goals that are wrong. They might want to help people: enslave, harm, manipulate or humiliate other people. For example, 'slave collars', prevented black slaves from moving through forest areas or laying down to rest their heads. As such, they made it impossible for a person to escape from slavery. Building an artefact like this is wrong because the goal it served, slavery, is wrong – it's a violation of the dignity and freedom we should afford to every person.
02.	**UNJUST USE**
	Sometimes artefacts can aim to serve a defensible goal but do so in an unacceptable way. For instance, surveillance technology may aim to keep people safe, but may enable the widespread invasion of privacy. Many consider some fuel sources unethical because they have harmful effects on the environment. Although these artefacts have acceptable goals, their use generates other moral problems.

There is debate around what counts as an unjust goal or an unjust means. Once defined, these two questions roughly summarise the ethics of technological artefacts. Artefacts that serve just goals through acceptable means are ethical pieces of technology.

However, recent controversies have highlighted a third set of ethical considerations:

+ The failure to foresee alternate uses of an artefact – both for good and for ill
+ Uncertainty around the effects of new technologies or new uses for existing technologies.

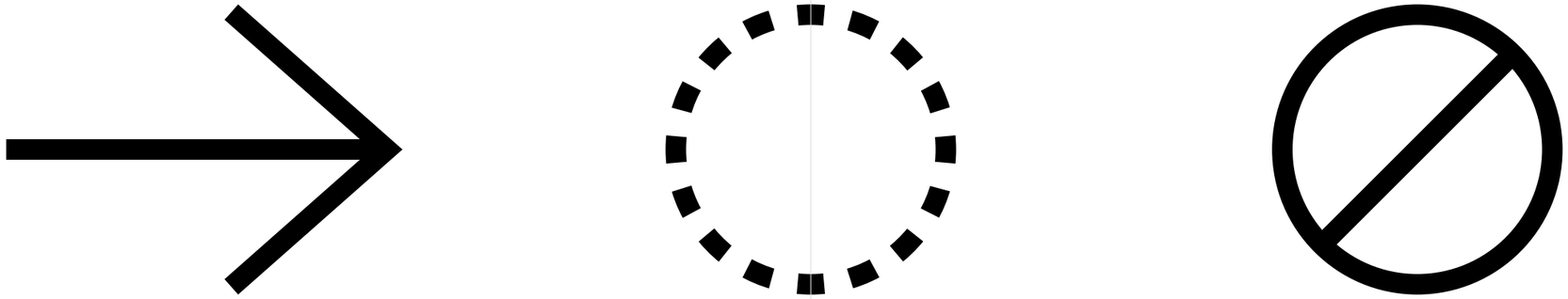
The issues surrounding Facebook and Cambridge Analytica provide a timely example here. Although many think that the company should have foreseen the risk, Facebook claims that it did not predict that its platform would be leveraged by third parties to manipulate its users for political ends. Nor did the company recognise how a typical user of its platform might be exposed to the risks of addiction, loneliness or other mental health concerns. Yet, surely an approach to technology that failed to recognise and address these issues should be deemed ethically inadequate.

As a result, we need to add a THIRD consideration to our account of technological artefacts:

03.	**UNINTENDED EFFECTS**
	It's not enough to account for the intended goals and uses of technology. Ethical technology design, innovation and use must also be responsible and agile. It should pre-empt, or have an inbuilt capacity to respond to and manage, those things its designers did not, or were unable to, predict.

TECHNO-LOGIC SEES THE WORLD AS THOUGH IT IS SOMETHING WE CAN SHAPE, CONTROL, MEASURE, STORE AND ULTIMATELY USE. IT IS A LOGIC OF CONTROL.

Techno-logic



The three-part model provides a general outline for the ethics of technological artefacts.

However, there is a second sense of technology which drives its design and development.

Since at least the middle of the 20th century, scholars have argued that ‘technology’ is a way of seeing the world. According to philosophers like Martin Heidegger and Jacques Ellul, technology is, in its purest form, a way of thinking.

This way of thinking sees the world as a set of problems to be solved, forces to be understood and measured, and products to be collected, stored and used. Technological artefacts bring this way of thinking into being. They mediate our relationship with the world so that we can’t help but see it in a certain way.

This kind of thinking has gone by a number of names: calculative rationality, *techne*, technique, and so on... To avoid too much philosophical jargon, we refer to it as techno-logic.

Techno-logic sees the world as though it is something we can shape, control, measure, store and ultimately use. According to this view, techno-logic is the ‘logic of control’. No matter the question, techno-logic has one overriding concern: how can we measure, alter, control or use this to serve our goals?

The application of techno-logic has helped to provide considerable benefits to the world. The scientific revolution and subsequent developments in STEM and other fields have lifted the standard of living for millions of people. It has enabled an expansion in our understanding of what it is possible when it comes to living a good and meaningful life. Such things may not have happened without the calculation, efficiency and rigour that is at the core of techno-logic.

All the same, we should not let the effectiveness of techno-logic stop us from forming a clear-headed assessment of its appropriate place and its limits. This begins by recognising the negative implications of a ‘technological gaze’. The chief risk is in ‘technologising’ things that should be beyond a brute logic of calculation and control. There are things we can’t or shouldn’t reduce to their use value. There is also a totalitarian potential in a logic that aims to control life, a potential that cannot be ignored or discounted because of noble intentions.

Examples of this approach to technology occur in all walks of life. For instance, a range of big data algorithms claim to be able to predict and even alter people’s behaviour. The stated aim may be benevolent, but this technology risks treating human beings as objects. Techno-logic can license us to measure, control and use other people to achieve our goals. Again, this is a violation of the fundamental principle of ‘respect for persons’.

In 2014, a team of researchers published a study entitled ‘Experimental evidence of massive-scale emotional contagion through social networks’.[1] The study used Facebook’s algorithm to control whether users saw positive or negative posts in their feeds. They found news feeds could be manipulated to influence a user’s moods.

[1] Adam DI Kramer, Jamie E Guillory & Jeffrey T Hancock, ‘Experimental evidence of massive-scale emotional contagion through social networks’, Proceedings for the National Academy of Sciences of the United States, vol. 111(24), 2014.

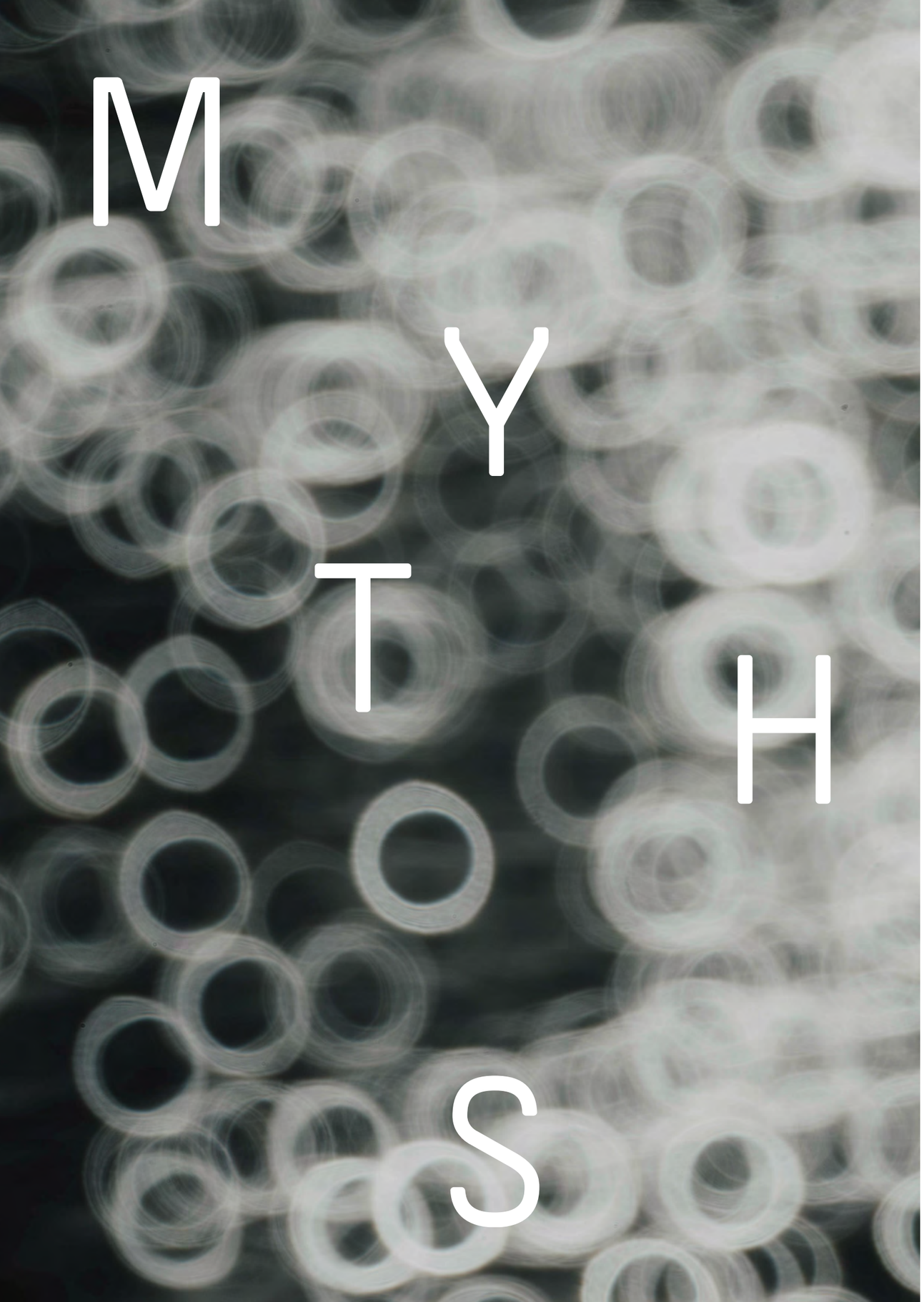
This use of technology would seem to fail our tripartite test of goals, means and unintended effects.

+ **THE GOAL,** understanding (and effecting) emotional contagion, is of questionable ethical standing. Most societies place a moral premium on human freedom and individual self-determination.
+ **THE MEANS** seem to fail to exercise an adequate duty of care. They exposed groups to negative emotional content without screening for possible mental health vulnerabilities, following-up with a debrief (standard practice in human research) or providing support mechanisms.
+ **THE UNINTENDED EFFECTS,** could have ranged from self-harm and suicide to the successful deployment of emotional contagion to advance unjust goals.

It seems clear that, as an artefact, the emotional manipulation algorithm was unethical. However, this analysis is incomplete. We must also acknowledge and explore the set of beliefs underpinning these actions. Why were these decisions made? What kind of rationality would license or encourage this path of action?

This is where understanding techno-logic is of such importance. It leads us to a new and deeper range of questions we can use in an ethical pre- and post-mortem.

+ **Did Facebook rely on a limited concept of ‘excellence’ that did not extend beyond a concern for: efficiency, control, measurement and effectiveness?**
+ **Was this an endeavour built by many hands, few of whom knew precisely what they were creating and to what end?**
+ **Did this logic blind decision-makers to the ethically salient components of their decision?**



TECHNO-ETHICAL MYTHS

Ethical frameworks don't appear out of nowhere. There are always pre-existing beliefs, practices and cultural factors we must consider. Technology is no different. There are a range of widespread beliefs that have developed over time which make it harder to introduce ethics into the mix. Of these beliefs, from an ethical perspective, the most problematic is the claim that technology design is 'value neutral'.

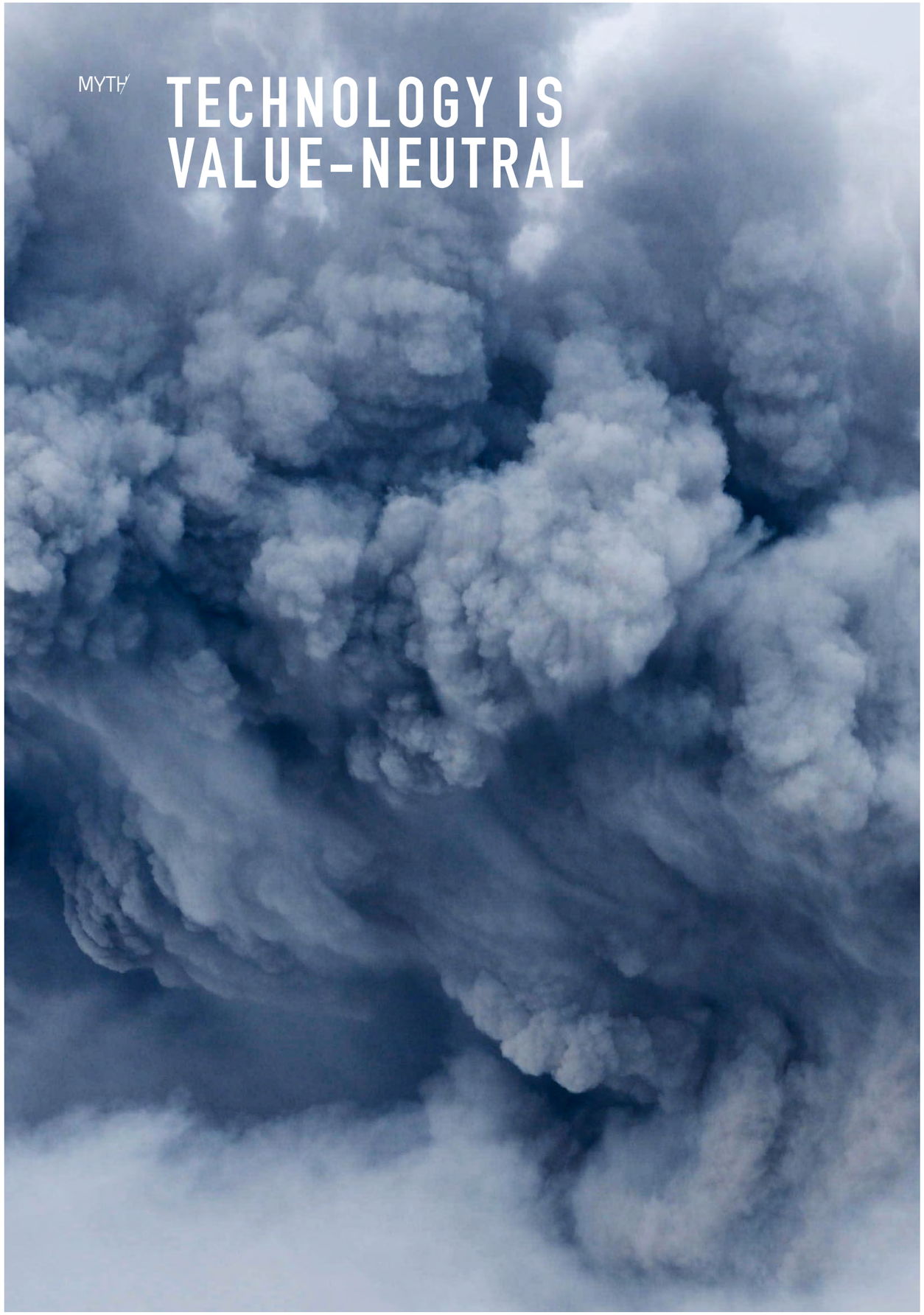
Not everyone thinks this way but it, and other beliefs, are common enough to warrant unpacking. What follows is consideration of four common and related beliefs. They suggest it is either pointless, unnecessary or impossible for ethics and technology design to work together. On close inspection, it turns out these beliefs are **WRONG**. They are that:

01. TECHNOLOGY IS VALUE-NEUTRAL

02. WE SHOULD BLAME THE ARTEFACTS IF THINGS GO WRONG

03. WE CAN'T HALT THE TIDE OF TECHNOLOGY

04. WE CAN HOLD OFF ON THE ETHICAL QUESTIONS



"Guns don't kill people, people kill people."

Most people recognise this line as an apology for gun ownership in the wake of a mass shooting or other atrocity. It also represents one problematic view of the ethics of technology: technological instrumentalism.

Technological instrumentalism is the idea that technology is a mere tool. Instrumentalists believe technology has no motives, intentions or agency. As a result, they think it's foolish to blame the harms people inflict via technology on the artefact rather than the person using it.

Today, very few scholars defend the instrumental view of technology. The chief reason is simple. Although technology doesn't have agency, it's design can influence our agency. Technological artefacts mediate our relationship with the world in both a moral and physical sense.

The instrumentalist misses the important point about guns. The person holding a gun engages with the world as a series of potential targets. It also provides him (it's usually a him) with a new way of interacting with the world: firing lethal shots. Moreover, the gun reframes the wielder's choice. He now has only two ways of interacting with the world: shooting, or not.

Given this, to focus on the person who does the killing is a red herring. What matters is the extent to which a gun facilitates murder.

Technological instrumentalism is the source of the intuition that technology is 'value neutral'. Technology is frequently seen as neither good nor evil until people get involved. This is a myth.

As we've said, technology influences our behaviour in various ways by mediating new relationships with the world. Each of these mediations carries with it value-based claims. In the same way, creating a new technology involves making ethical claims. For example, making a gun implies that sometimes taking another life is acceptable.

Less controversial examples make the same point. The invention of the motor vehicle conceptualised freedom of movement as being good. The printing press assumed the distribution of information to be a good thing, and so on. It is difficult to think of a piece of technology that is completely free of value judgement.

Understanding techno-logic helps us see another way in which value-neutrality is mistaken. Technology is born of and reflects techno-logic. It limits and controls the way in which the world appears to us. Techno-logic also recalibrates our thinking and understanding of what it means to solve a problem. For techno-logic, no problem is so complex that the right invention couldn't fix it. The meme, "Uber, but for X", works on the premise that we want apps to automate and digitise solutions for all our problems – from the practical to existential.

Consider an example from the more perverse end of the spectrum. The Gatling Gun, once the deadliest weapon in the world, was designed to reduce casualties. Here, again, we can see the creeping effects of techno-logic and its laden values. If only we could design the right piece of technology, we could end war itself.

Here is the second sense in which technology is not neutral. It creates blind spots such that we can only imagine technological solutions. All problems are either problems of non-existent technology or poorly-designed technology. In this way, techno-logic is the ultimate intellectual pyramid scheme. Once you buy in, it's almost impossible to free yourself.

Artefacts carry values with and within them.

This doesn't mean we should blame artefacts for the values they embody or express. No technology invents its own values and they aren't morally responsible for them.

Technology can be causally responsible for things that happen – and often is. However, moral responsibility is not something we can ascribe to technologies. Responsibility requires agency: the ability to determine one's own course of action and the possibility of having acted differently. This is beyond the scope of technology.

The ethical lesson of stories ranging from the Jewish myth of the Golem, to Frankenstein, to Jurassic Park bears this out. The creations – monsters and dinosaurs – destroy, terrify and amaze. But they aren't the real villains. The true monsters are the people who, through negligence or malice, make and abuse them.

Remembering this point is important. It helps us to determine who we should hold responsible in cases when technology has caused harm. What's more, it encourages humanity to bear responsibility for the future.

The impact of technology on our lives and our futures is ours to manage. This means that when technology is used wrongfully, there is some person to blame. Somebody – or some group of people – could have acted otherwise to prevent what had happened.

Some forms of technology may function without human input. Still, we should only ever consider them causally responsible. For technology to be morally responsible, it would need to achieve more than human-level intelligence. It would also need all the other characteristics of humanity: conscience, the ability to act against our instincts, mortality, and so on. Until (and if) this is achieved, whenever people blame technology, they are mistaken. By blaming technology, we attribute to it everything that makes us human.

This possibility is precisely the outcome feared by German philosopher Martin Heidegger. Heidegger believed it was possible for technology to redeem humanity. Still, he worried about its power to reflect, encourage and reward our desire for control, calculation and certainty.

He writes of 'the rule of enframing' (one of the sources of our idea of techno-logic) that:

The threat to man does not come in the first instance from the potentially lethal machines and apparatus of technology. The actual threat has already afflicted man in his essence. The rule of enframing threatens man with the possibility that it could be denied to him to enter into … a more primal truth.

The "primal truth" Heidegger is referring to is a human life, lived authentically. He thought our central task was to be true to ourselves – including the type of 'being' that we are – free, decision-making, responsible and so on. He worried that technology could enable people to sit comfortably in a state of perpetual self-denial. We could leave to technology all the messy parts of being human and in doing so; pretend we're not as responsible as we truly are.

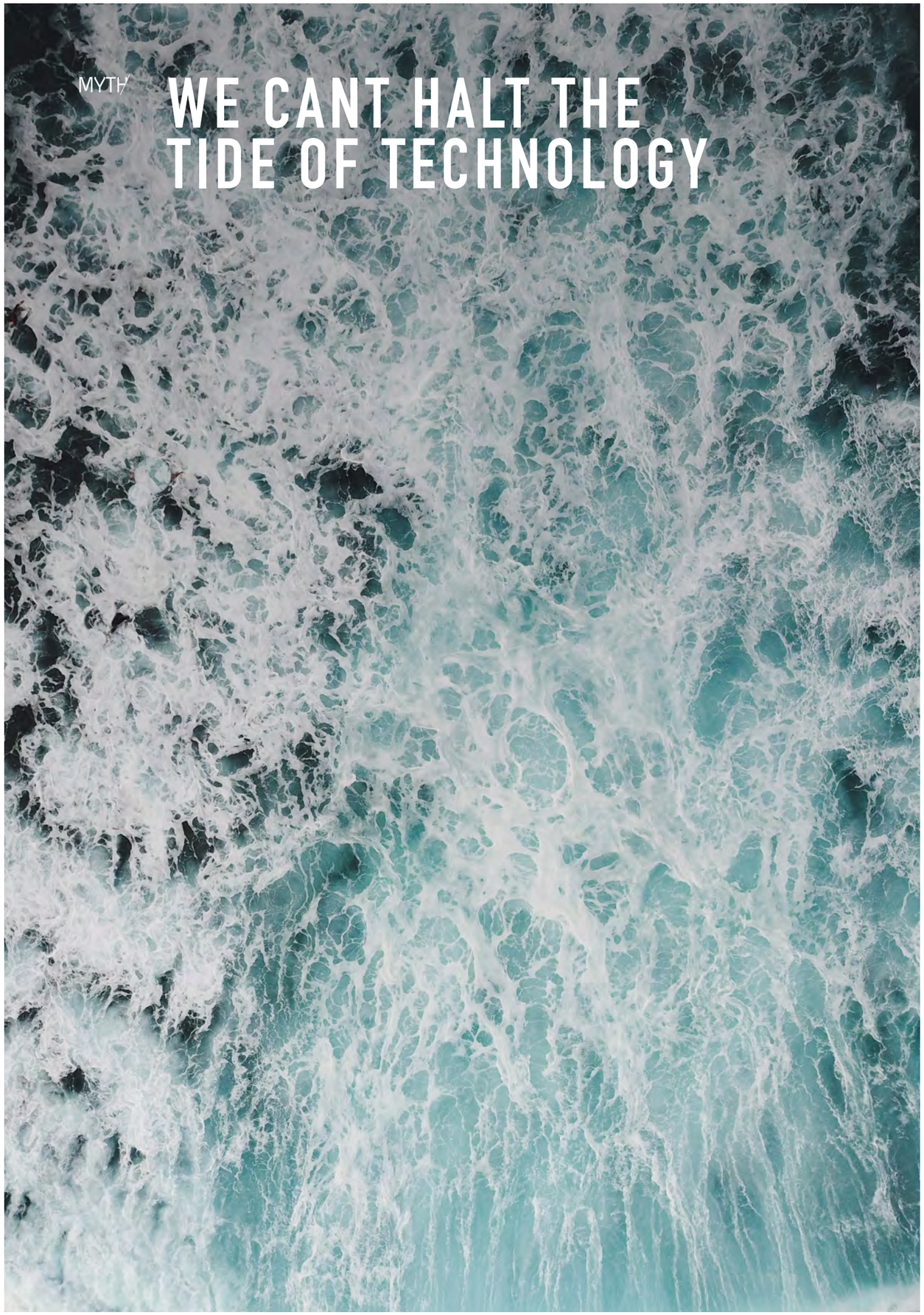
One practical step that helps to deny responsibility is the siloing of different people into teams or roles. Small teams are often built to contribute a very specific part to a much larger end product. This gives them little power to effect change on the final result. There is immense risk if a team or person limits their role to 'doing' without questioning the potential uses and consequences of their technology. When everyone assumes somebody else will take responsibility for the ethics, nobody does.

If enough people leave the moral considerations to someone else, we risk technocracy on one hand and totalitarianism on the other.

Indeed, when reflecting on how people could have willingly caused the Holocaust in World War II, philosopher Hannah Arendt coined the phrase 'banality of evil' to refer to precisely this process. Arendt highlighted the way people do evil by failing to think reflectively about their actions. By viewing themselves as cogs within a broader system, people can ignore their moral responsibilities. This lack of thinking can empower gross institutional wrongdoing.

Another practical barrier to accepting responsibility is the incremental nature of technological progress. When people are so focussed on refining specific elements of a piece of technology, it is easy to overlook any incremental effects. Ethics requires us to consider the broader impacts of our activities. We must take into account the cumulative effects of small actions. Those involved in a limited aspect technology design cannot pretend they have no responsibility for the end product – for good or for ill.

Technologists are responsible for what they build. Exactly how responsible is an open question, but to claim they are innocent of the results of what they make is an act of bad faith.



It is not uncommon today for people to talk about the inevitable path of technology, as though the future is no longer ours to control, but in the hands of our creations.

This thinking is wrong. Humanity remains in control of our own destiny (even if at some point we are no longer able to control the specific actions and effects of the technologies we create).

There are two versions of this argument. The first is a kind of progressive refrain – human creative genius cannot be held back or restrained, innovation will out and so forth. This mentality is present both in advocates of technology and their opponents. Frequently, opposition to new technologies takes the form of a 'slippery slope' argument, where the true objection isn't to the existing technology, but to other related technologies which may come to pass in the future.

For example, the use of genetic testing, selection and even gene editing technology is often feared not for its likely immediate applications, but for fear of the most dystopian applications of the technology – such as 'designer babies', a genetically superior class of humans and so on. These may indeed come to pass, and for many this would be considered a bad thing. But if they do, it won't be because the implementation of a technology designed to alleviate suffering set us on an inescapable course to the apocalypse. It will be because of a series of choices made by human actors – and at each point along the way, if the foundational claim of ethics is true, they could have acted otherwise.

The implications of this kind of technological determinism are also concerning. If we deny our agency in impacting how technology takes shape and affects our world, we may lose one safeguard for ethical conduct: belief in free will. A series of studies in the last 15 years or so have shown that people who are more confident in the existence of human freedom are more likely to avoid opportunities to cheat[5] and tend to score more highly in job performance surveys.[6]

If we believe technology is an unstoppable force, one concern would be that we might be less inclined to exercise our agency and responsibility in making ethical decisions. And, indeed, when you consider a second kind of technological determinism, it appears there is some basis to this hypothesis.

This second form of determinism is not based in ontological claims about the direction of technology. Instead, it simply argues that, pragmatically, if someone is going to build, profit and get the credit for building a piece of technology, it may as well be them. This is a poor justification: I am certain my loved ones will die at some stage – it doesn't follow that I would be justified in killing them for pay. The view serves as a convenient rationalisation for self-interested behaviour, but it's unpersuasive as an ethical argument.

We cannot point to 'trends' or observations of what most people are doing, or which horses have bolted as ethical justifications for the technology both of today and the future. We must accept our agency in this regard.

As the Canadian philosopher and futurist Marshal McLuhan wrote, "There is absolutely no inevitability, so long as there is a willingness to contemplate what is happening."[7]

[5] "The Value of Believing in Free Will: Encouraging a Belief in Determinism Increases Cheating," Kathleen D. Vohs and Jonathan W. Schooler (2008), Psychological Science, 19 (1), 49-54.

[6] "Personal Philosophy and Personnel Achievement: Belief in Free Will Predicts Better Job Performance," Tyler F. Stillman, Roy F. Baumeister, Kathleen D. Vohs, Nathan M. Lambert, Frank D. Fincham, and Lauren E. Brewer (2010), Social Psychological and Personality Science, 1, 43-50.

[7] Marshal McLuhan, The Medium is the Message

In 2015, Harvard Professor of Psychology, Steven Pinker, made a radical declaration. He suggested bioethicists should "get out of the way" of scientific progress.[8]

For Pinker, delaying life-saving technologies until we have resolved every ethical issue would be unethical. He thought this remained true even when the reasons for the delays are themselves ethical: for instance, to consider long-term implications of the new technology.

The message here is simple. Given technology can increase quality of life for a huge part of the population, there is an ethical cost to delaying research.

Pinker is right to highlight the ethical costs of delaying technological development. But his argument understates the risks in 'bolting on' the ethical dimension to technology after the fact.

It's tempting to consider what technology can do before considering about what it should do, but to do so would be unwise. Once technology is developed, one of the most potent protections against ethical missteps – the thing not existing – is already lost to us. To use a cliché, once the cat is out of the bag, we can't put it back in.

This poses an ethical double-bind for technology design. On the one hand, we must safeguard against reckless, negligent or badly intended technology. On the other, we need to explore and advance opportunities to improve the common good. Those involved in technology design may be damned either way. They either allow bad technology out in the world or prevent promising technology from developing fast enough.

Discovering the tools for resolving this double-bind can't be left until the decision-making moment. We need to be proactive in identifying how to balance innovation and restraint in technological innovation.

[5] Steven Pinker, 'The moral imperative for bioethics', Boston Globe, August 1, 2015, Accessed: https://www.bostonglobe.com/opinion/2015/07/31/the-moral-imperative-for-bioethics/JmEkoyzITAu9oQV76JrK9N/story.html

TECHNOLOGY: AN ETHICAL FRAMEWORK

Technological design needs a proactive ethical framework, consisting of a statement of purpose, core values and guiding principles, within which good technology can be proposed, designed, updated and used.

This would enable our judgements and decisions to be forward-facing. Too often, ethical reflection is reactive, used to diagnose and resolve a problem after it has occurred. Whilst ethics is useful here, this approach diminishes the true value of a commitment to ethical reflection and action.

An ethical framework allows us to pursue excellence. By basing our thoughts, decisions and actions in a clear statement of why we're here, what we stand for and where we draw a line in the sand, we go far beyond a 'do no harm' approach to ethics. Instead, we're able to imagine the best version of something – in this case, the best kind of technology.

Of course, if this ethical framework is worth its salt, it will also give us a precise and rigorous way to diagnose ethical failure, apportion responsibility and seek justice for victims. However, this won't be all it does. It doesn't just outline the minimum standard, it also explains the ideal we should be striving for.

Every Ethical Framework includes three elements:

PURPOSE

Is our reason for being.

It helps to explain and animate one's choice of core values and principles.

VALUES

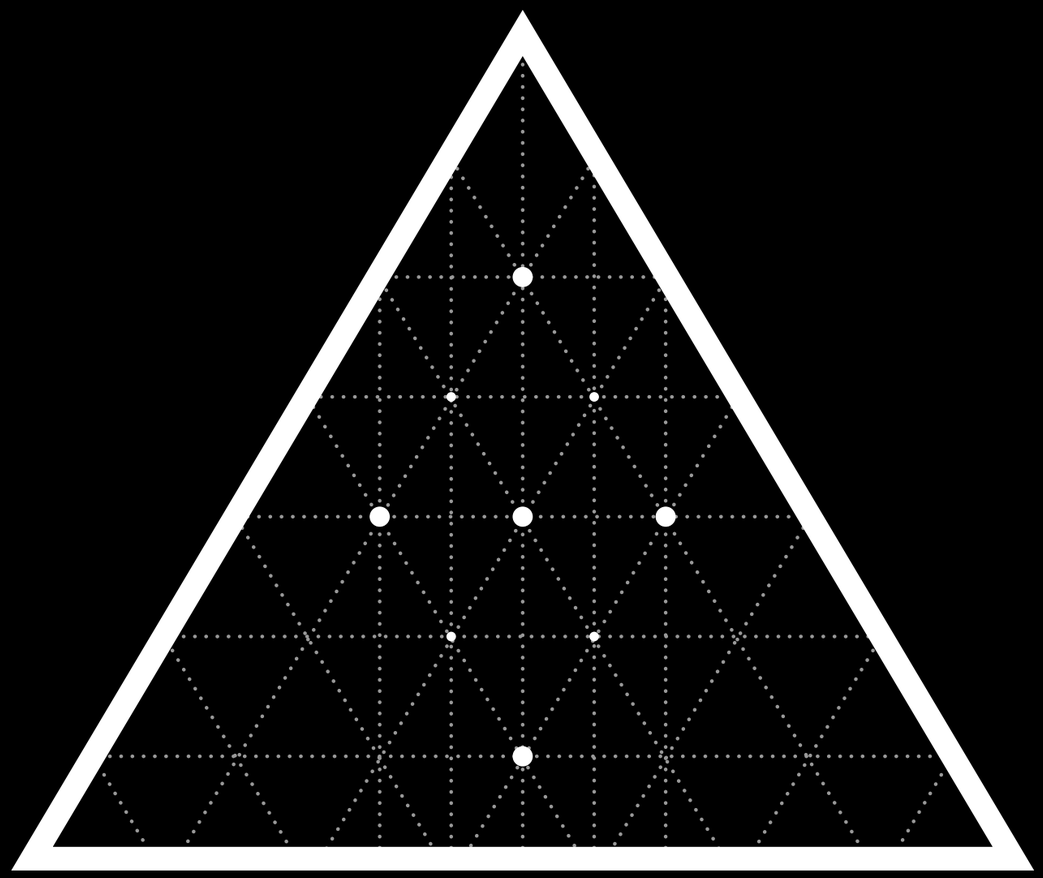
Identify what is good.

They are the things we strive for, desire and seek to protect.

PRINCIPLES

Identify what is right.

Outlining how we may or may not achieve what is good.



Project Goal

We want this framework to form part of a much larger collection of tools, resources and processes.

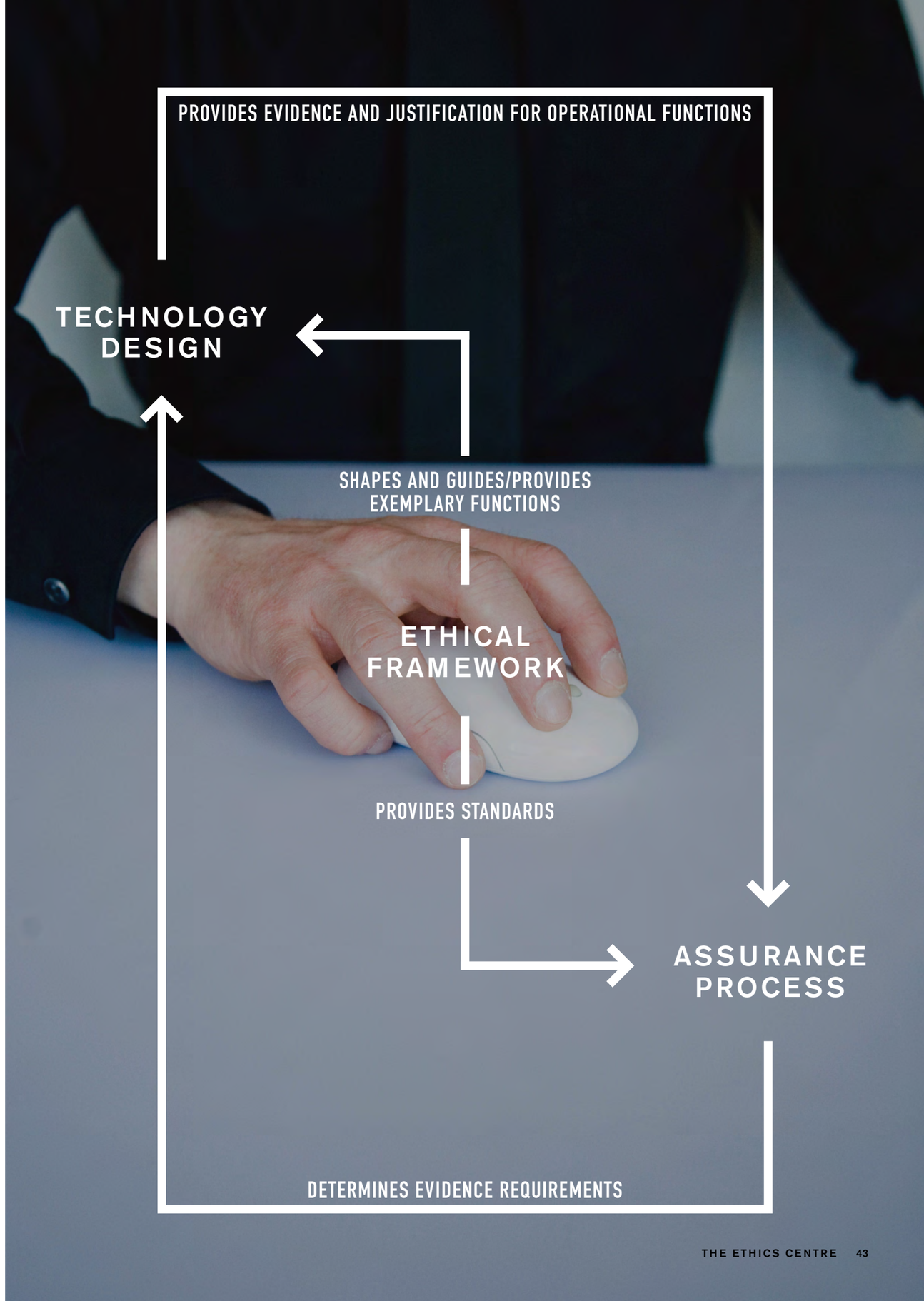
Our final goal is to enable public trust in technology and to support the good will of many in the technology sector. We want to help ensure what is built meets – and exceeds – ethical standards.

This ethical framework might be 'bookended' on either side by an ethical design process and a market assurance process. It should be possible to develop a system through which:

01. DESIGNERS GIVE EVIDENCE OF HOW NEW TECHNOLOGY SATISFIES ETHICAL REQUIREMENTS;

02. THIS EVIDENCE IS THEN CHECKED AND ASSURED BY AN INDEPENDENT THIRD PARTY; AND

03. ONCE ASSURED, THE TECHNOLOGY CAN BE 'MARKED' AND RELEASED TO THE MARKET.



PURPOSE

A traditional ethical framework talks about PURPOSE as though it were a noun.

Thus, people, organisations and products are said to have a *purpose* to which they should stay true. This mode of thinking is also common in the way we make technological artefacts. For instance, a bench is for sitting.

Thinking about technological purpose only in this way suggests a set and sole purpose for the things we make; that there is a 'right' and 'wrong' way to engage with an artefact. It also suggests the purpose of an artefact is solely a product of the designer's will. Whatever they intend their creation to be for is what it is for. However, as we'll see, this mode of thinking isn't suitable for thinking about technological purpose.

First, even if designers have a specific purpose in mind when making an artefact, individual users may use it in different and unintended ways. Take the bench example above. Whilst a designer might intend for their bench to be a seat, a homeless person may – design permitting – use it as a bed.

Should we see this type of usage as being 'worse' than the one intended? Should designers consider the diverse ways in which their technology will be used, or should they focus solely on realising their intent?

If we choose the latter approach, then we treat purpose as a noun. It is a fixed thing that exists objectively, regardless of how people use the technology. This is the sense in which technology can be described (from the designer's perspective) as being either abused or misused.

In treating purpose as a noun, we create the following risks:

01. DESIGN APPROACHES WILL TEND TO UNDERDETERMINE OR OVERLOOK UNETHICAL USES FOR AN ARTEFACT.

For example, Tor is an online browser intended to help free speech and privacy for users. But it is also a safe portal for a range of criminal and unethical activities.

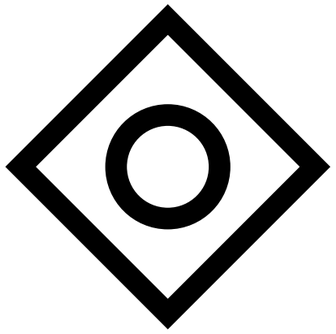
02. DESIGN APPROACHES MAY EXCLUDE USEFUL AND ACCEPTABLE USES FOR ARTEFACTS BY UNINTENDED USER GROUPS.

For example, "hostile design" is often used on park benches to stop homeless people from using them as beds. The implication is that because benches are for sitting, and not intended for accommodating the homeless, this is acceptable design.

Given these risks, the purpose-as-noun approach is insufficient. Furthermore, this approach misunderstands the nature of technology. Although technology is designed with an intention in mind, its use is only ever 'partially bound' by the designer's intention. The innovation and genius that drives technological design also enables (and perhaps encourages) the novel use of artefacts. In part, this is because the best uses for technology aren't always obvious at the moment of design.

For these reasons, we should take an alternative approach and think about purpose as a verb – a fluid, group activity involving a range of influences and processes. To understand this approach, it's useful to look at two different academic approaches to technology: postphenomenology and affordance theory.

Postphenomenology



Postphenomenology is a philosophical theory of technology.

It holds that we should study technology as a relationship between users and artefacts. It also holds that people exist insofar as they are relating to something in the world. We are always *doing* something, and we're always doing it *with* something. Basically, we're always using technology in some way. Postphenomenology is interested in the ways we use technology and how that use defines both the artefact we're using and ourselves.

Philosopher Nolen Gertz explains this with the example of a fork. When a child first encounters a fork, it may throw, chew, drum, stab or do any number of other things with it. Over time, the child learns the *intended* use of the fork - eating, and modifies its use accordingly. The fork is seen as having a context: eating.

But postphenomenology doesn't think there is any moral difference between using a fork as a toy, eating device or projectile. Instead, it thinks the different usages reveal something important to us - that in using a technology, we shape both it and ourselves. As Gertz explains:

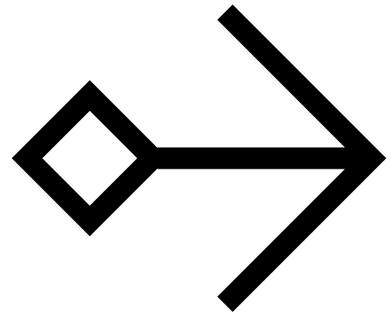
There is no fork independent of the specific human-technology relation in which it is engaged, just as there is no child or adult independent of that same engaged relation.[9]

To put it another way, when a hit-man picks up a firearm, he sets the purpose of the gun as a murder weapon. However, he also uses the gun to constitute himself as a murderer. Both are crucial: there is no murder or murderer independent of the person/gun relationship.

Gertz also suggests the designer's intended purpose for the fork is less important than the way it is actually used. In fact, Gertz's argument goes further, suggesting there can be no *purpose* until the technology is used. This leaves the designer in a much more fluid and ambiguous role than is usually imagined.

[9] Nolen Gertz, 'Designing Responsibility' in van der Velden, M., Strano, M., Hrachvec, H., Abdelnour Nocera, J., & Ess. C. (Eds.). Culture, Technology, Communication: Common worlds, different futures? Proceedings of the Tenth International Conference on Culture, Technology, Communication, 2016, Accessed: http://philo.at/ocs2/index.php/london16/catac_16/paper/view/320/145

Affordance Theory



Like postphenomenology, affordance theory examines how technology exists in the world.

Affordances are clues that suggest the ways a person can use a technology to interact with the world. Unlike purpose, affordances are not the intended or primary functions of the artefact. Whilst purpose describes what an artefact is for, affordances reveal what an artefact can do.

Artefacts can afford in lots of ways. Some affordances are "bids" an artefact places on the user, others are responses to bids the user makes to the artefact.[10] This explains how design influences user behaviour in ways that may be outside the designers' intention or control.

Davis and Chouinard outline six kinds of affordances. They suggest artefacts afford by requesting, demanding, allowing, encouraging, discouraging and refusing.[11]

+ **REQUESTS** recommend a line of action to the user. For instance, a doorbell requests you ring it rather than knock, storm in unannounced or scream until someone opens the door for you. All these options are still possible, though.
+ **DEMANDS** are behaviours that are necessary to use the artefact. An ATM machine demands you enter a PIN number before it will allow you to do anything to the account.
+ **ENCOURAGEMENT** happens when an artefact promotes one line of action above others. Lumbar-support chairs encourage good posture and promote ergonomic modes of sitting above alternatives.
+ **DISCOURAGEMENT** occurs when a particular kind of action requires extra effort to perform. For instance, I can unsubscribe from mailing lists, but I am discouraged from doing so by small, hard-to-locate text telling me how to do so.
+ **REFUSALS** prevent users from engaging in some activities. Password protection refuses entry to those without the password.
+ **ALLOWANCES** are neutral with regard to action. When an artefact is indifferent to different actions, it allows them. A car allows vehicles to travel at a variety of different speeds without influencing the chosen speed in any way.

Understanding this suite of affordances helps us to see how some elements of intention and purpose can be captured in design. It helps explain why the postphenomenological position, that purposes don't determine use, is correct. However, it also explains, contra postphenomenology, how design decisions affect user behaviour. Although the postphenomenologist may be right that user choices create an artefact's purpose in an important way, design affordances also shape the choices available to the user. Affordances are therefore a manifestation of the designers' intentions.

[10] Jenny L Davis & James B Chouinard, 'Theorizing Affordances: From Request to Refuse', Bulletin of Science, Technology & Society, 2016, Vol. 36(4) 241-248

[11] Ibid., 242-244

Proposing / Affording / Purposing

If we are to accommodate the insights of postphenomenology and affordances into ethical considerations of purpose, we need to move beyond seeing purpose as a noun. Instead, we should think about 'purposing' as one step in a three-stage process:

01. PROPOSAL STAGE

At the proposal stage, designers outline the problem they wish to solve and the artefact that will solve it. The designers propose a series of functions that will enable the artefact to help address the problem. At the proposal stage, designers identify the *why, how* and *desired outcome* of the technology. This is what we usually consider to be the intended purpose. However, they also embed these intentions into the artefact by way of functions and features – some of which may operate as affordances. These features – not the intentions in the mind of the designers – are the proposals that will be considered by the user.

02. MEDIATION STAGE

At the mediation stage, the technology and user engage in a relationship to perform some action. At this stage, the user evaluates the features and affordances of the artefact, and begins to decide how to use the technology.

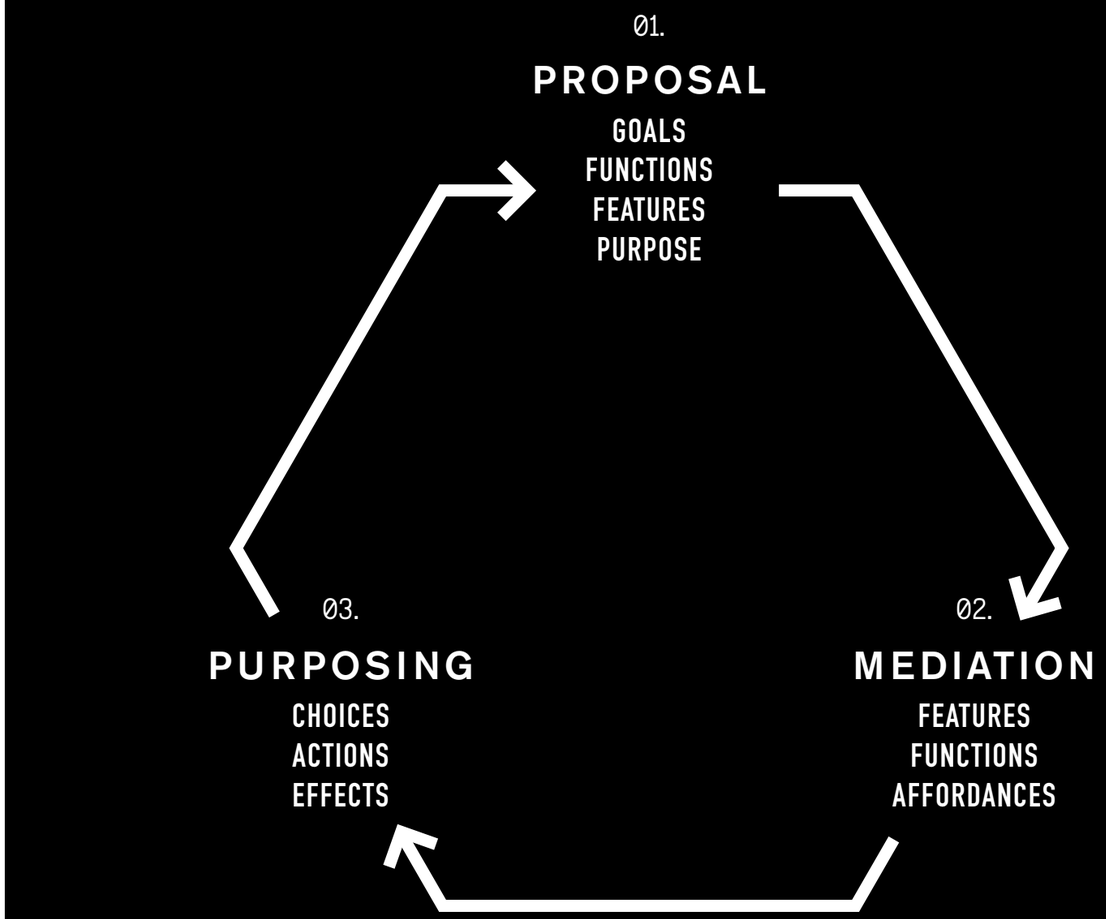
03. PURPOSING STAGE

Finally, the user makes a decision about how they will use the technology, and follows this up with action. At this stage, the user can accept the proposals offered by the designers, thus fulfilling the artefact's intended purpose. But they can also use it to achieve some other goal, thereby bringing to life an entirely different purpose (albeit within the boundaries set by the relevant affordances). Sometimes the purposing stage might be the beginning point of an entirely new artefact, at which point the design process begins anew with another set of proposals.

Thinking in this way reveals a few important points regarding the ethical purpose of technological artefacts:

+ **PURPOSE CANNOT BE FULLY DETERMINED BEFORE THE ARTEFACT IS USED.**
+ **PURPOSE IS INDEPENDENT OF THE INTENTION OF THE DESIGNER.**
+ **PURPOSE IS A PROPERTY THAT EMERGES THROUGH THE RELATIONSHIP BETWEEN DESIGNER, USER AND ARTEFACT.**

Each of these has important ethical implications we will explore later. It also reminds designers that their intention is only influential insofar as it is embedded in design decisions and design features of the technology. And even then, subject to the restraints of partial binding – design affordances – the user has the power to repurpose the artefact in another way, creating new areas of ethical concern.



VALUES

In ethics, values refer to the ends toward which our choices may be oriented.

They define what is good about whatever we're doing. For instance, if a law firm holds 'justice' to be one of its values, we should expect the firm to make decisions in favour of justice. It would also be conspicuous if a law firm did not have something like 'justice' as one of its stated values, given the kind of work it does.

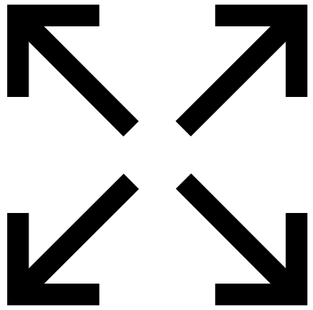
We don't pretend we could – or would want to – prescribe which values technology ought to serve. There is a vast array of goods – as in, things we value and orient our choices toward – in the world to which people might dedicate their lives. Any of them might be values advanced, protected and preserved by technology. For instance, medical technology will value health, whilst military technology is unlikely to do so to the same extent.

Artefacts can concurrently serve a variety of different values. Furthermore, technology design can, and often does, include hidden values through the adoption of inherent social or cultural norms. These values may not be know to designers or users, but reflect unconscious or inherent beliefs about the world that deserve more scrutiny than they receive. For this reason, we need to think of values in two ways:

01. EXPLICIT

02. IMPLICIT

Explicit Values



Explicit values are those goods a piece of technology openly endorses and pursues.

These are usually identified at the proposal stage and may be embedded into the design of the technology in a variety of ways. For example, ‘Connecting People’ is one of Facebook’s explicit values. Indeed, a widely-discussed internal ‘provocation’ written by Andrew Bosworth made clear how explicit Facebook’s interest in connecting people is:

We connect people.

That can be good if they make it positive. Maybe someone finds love. Maybe it even saves the life of someone on the brink of suicide. So we connect more people.

That can be bad if they make it negative. Maybe it costs a life by exposing someone to bullies. Maybe someone dies in a terrorist attack coordinated on our tools.

And still we connect people.

The ugly truth is that we believe in connecting people so deeply that anything that allows us to connect more people more often is *de facto* good.[9]

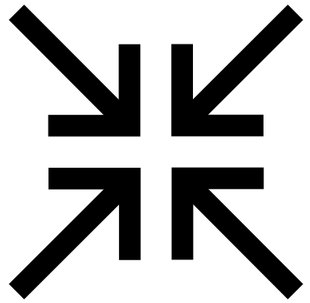
Whilst written as a provocation – and offering a range of ethical issues worthy of discussion – this is a good example of explicit values. It is clear what Facebook considers to be ‘good’ and what it’s ‘guiding star’ might be.

Other examples of explicit values would include:

+ **SECURITY** realised in home alarm systems
+ **KNOWLEDGE** realised in online education platforms
+ **EXCELLENCE** realised in the engineering of Formula One race cars

[9] Sheera Frenkel and Nellie Bowles, ‘Facebook Employees in an Uproar Over Executive’s Leaked Memo’, New York Times, 30 March 2018, https://www.nytimes.com/2018/03/30/technology/facebook-leaked-memo.html?smid=fb-nytimes&smtyp=cur Facebook memo

Implicit Values



Most – if not all – pieces of technology will also carry implicit values.

These values are embedded in or implied by an artefact’s purpose and design. They may not be condoned or recognised at the design stage. For example, the manufacturer of a high-end watch *explicitly* values accuracy, but they also *implicitly* value exclusivity.

Another set of implicit values exist in the form of techno-logic (p.26). Many technological artefacts implicitly value control, efficiency, effectiveness and measurement. And whilst these can be good, they are usually instrumental goods. Efficiency is good if it allows us to better obtain or distribute other things we value. Effectiveness is good when it facilitates justice, fairness or some other value. An efficient serial killer is in no way preferable to a bumbling, incompetent one.

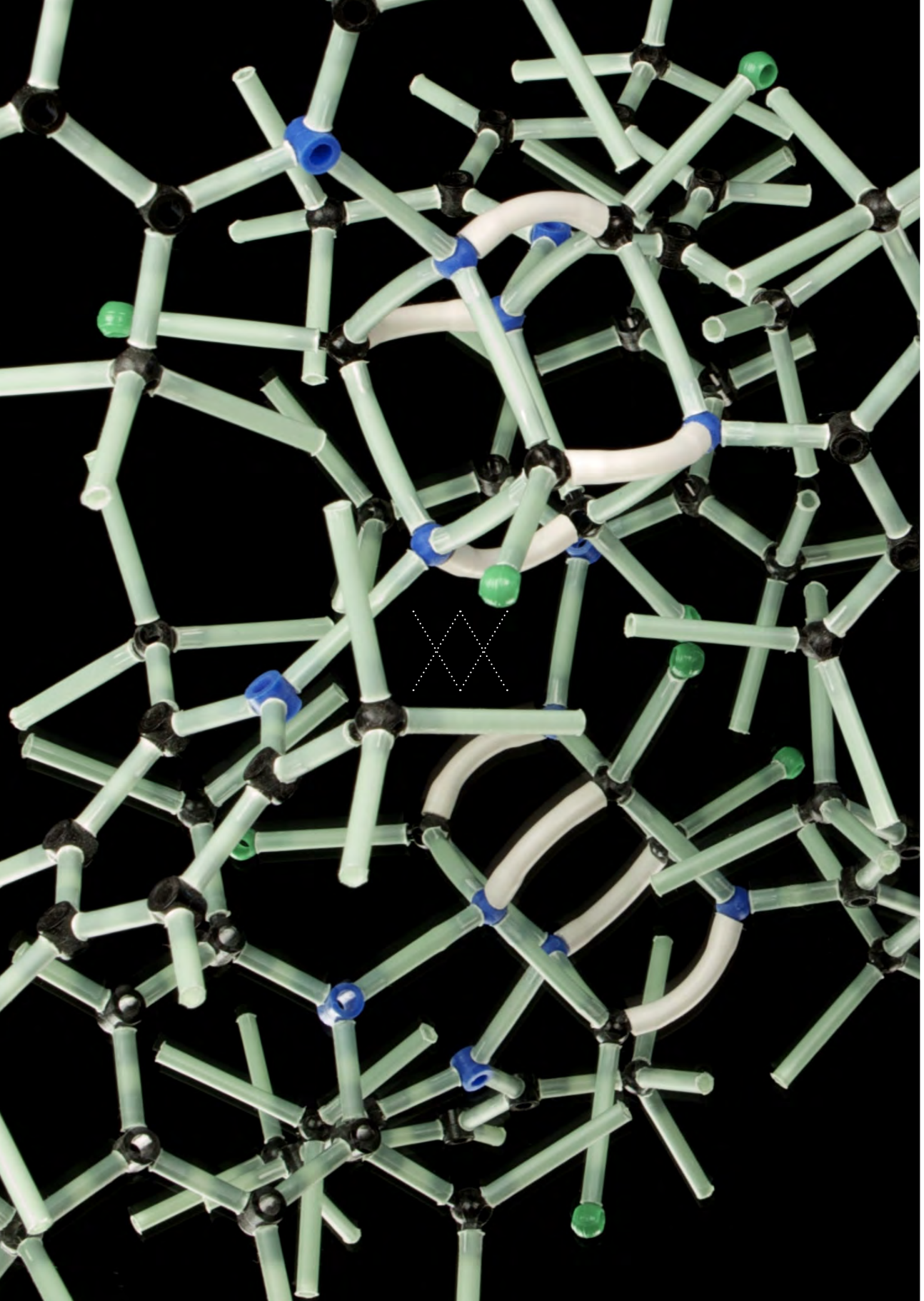
Values are also likely to reappear in the ‘purposing’ stage of technology usage. Although designers embed values into artefacts to help frame user choice, user values also play an important role. Values help determine the extent to which users respond to different affordances, accept the proposals of designers or produce positive effects.

Thus, when considering values, designers need to consider:

+ **THE WAYS IMPLICIT VALUES MIGHT BE INFLUENCING DESIGN DECISIONS**
+ **WHETHER DESIGN DECISIONS ARE PROJECTING VALUES UNINTENDED OR UNDESIRED BY DESIGNERS**
+ **HOW FUNCTIONS AND FEATURES CAN HELP EMBED VALUES INTO THE ARTEFACT**
+ **WHOSE VALUES DESERVE PRIORITY – DESIGNERS, USERS OR THOSE OF SOCIETY**
+ **WHICH OF A RANGE OF VALUES SHOULD BE EMBEDDED INTO THE ARTEFACT**
+ **WHETHER AND HOW TO ACCOMMODATE SOCIAL AND CULTURAL VALUES**
+ **HOW DIFFERENT VALUE SYSTEMS MIGHT LEAD TO THE REPURPOSING OF ARTEFACTS, FOR GOOD OR ILL**

[10] Jenny L Davis & James B Chouinard, ‘Theorizing Affordances: From Request to Refuse’, Bulletin of Science, Technology & Society, 2016, Vol. 36(4) 241-248

[11] Ibid., 242-244



PRINCIPLES

In ethics, principles help determine right from wrong.

Principles are general guides to action. Unlike rules, principles allow for a broad range of different values to be realised. As a result, they tend not to be overly prescriptive. The goal of a principle isn't conformity, but consistency.

For instance, the principle 'Do unto others as you would have them do unto you' does not tell us anything about how we should want other people to treat us. It can thus serve as a bridge for discussion between divergent schools of thought and value.

What follows are a set of principles designers can apply to all forms of technology.

These principles go 'all the way down'. They apply not only to every piece of technology, but to every element of that technology. It is possible for an artefact to be largely ethical, save for one feature that does not meet these ethical requirements. This is enough for us to consider the artefact as ethically unsatisfactory.

What follows are general ethical principles for technology design. We note several design questions and challenges that each principle gives rise to. We intend for these to serve as a guide to ethical restraint and enlightened progress in the field of technology.

Exemplary Functions

The characteristics of technological artefacts matter when we're making ethical judgements.

The features, functions and affordances of an artefact can determine the purpose of an artefact, the values it conveys and the ways people use it. For this reason, it is important to connect design features and functions to ethical principles.

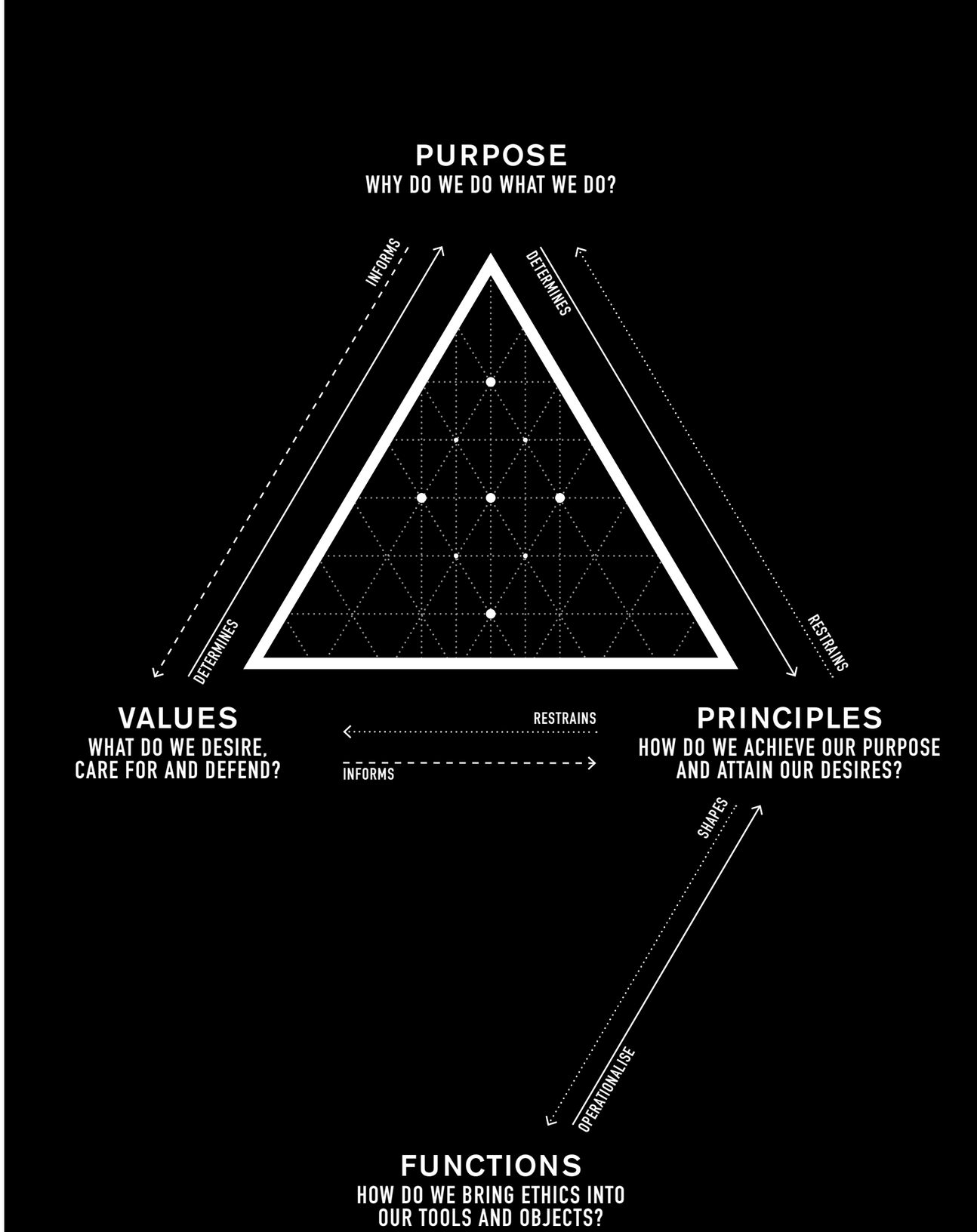
For the purposes of our discussion, **functions** are the activities an artefact performs. **Features** are the characteristics that determine how the artefact and user work together to perform a function. For example, one function of smart phones is the ability to browse the web. Features that enables this function include a mobile internet connection, web browser and so on.

As principles give rise to design issues and challenges, functions and features must resolve them. We will give some examples of functions and features that might provide some 'design solutions' to help an artefact to align with the ethical principles below. However, it is important to note here that we do not prescribe a comprehensive set of functions and features that arise out of the application of each principle.

The obligation to realise each principle – by way of function and feature – falls on each and every designer. The 'burden of proof' falls onto the technologist to demonstrate how (by what means) each principle has been given effect. That demonstration must meet certain standards. It must be:

+ **UNDERSTANDABLE** to a reasonable, independent and informed person
+ **ASSURABLE** by a third party
+ **ACCESSIBLE** to anyone with an interest in the artefact's use and effects
+ **COMPLETE** in demonstrating its application to all the relevant ethical principles

Ethical design challenges may not apply in every case. However, designers should be bound by an 'if not, why not, and what else?' approach. If a particular principle is said not to apply, or if a design solution isn't suitable or possible, designers should be ready to explain why, and give evidence of other measures that have been taken to operationalise the principles in question.



PRINCIPLES

PR00. OUGHT BEFORE CAN

The fact that we can do something does not mean that we should.

PR01. NON-INSTRUMENTALISM

Never design technology in which people are merely a part of the machine.

PR02. SELF-DETERMINATION

Maximise the freedom of those affected by your design.

PR03. RESPONSIBILITY

Anticipate and design for all possible uses.

PR04. NET BENEFIT

Maximise good, minimise bad.

PR05. FAIRNESS

Treat like cases in a like manner; different cases differently.

PR06. ACCESSIBILITY

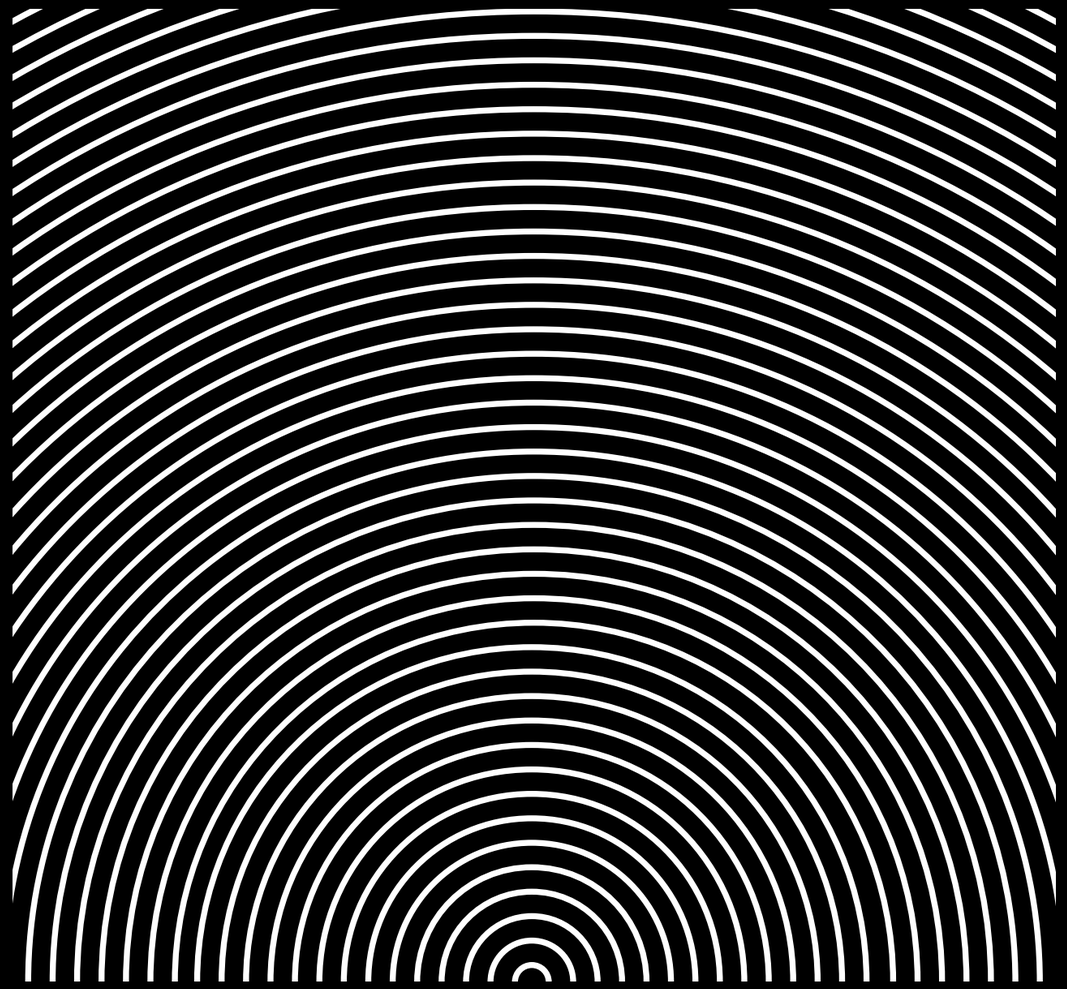
Design to include the most vulnerable user.

PR07. PURPOSE

Design with honesty, clarity and fitness of purpose.

GOVERNING PRINCIPLE

OUGHT BEFORE CAN



The fact that we can do something does not mean that we should.

There are lots of possible worlds out there – lots of things that could be made or built. Ethical design is about ensuring that what we build helps create the best possible world. Before we ask whether it's possible to build something, we need to ask why we would want to build it at all.

Philosophers working in ethics have long accepted that "ought implies can". It's accepted that we don't have an obligation to do something that cannot be done. In the technology we design, we should acknowledge a similar principle: just because something *can* exist doesn't mean it *should* exist.

Philosopher Richard Buchanan says, design is "an art of forethought".[13] This forethought should include asking whether or not an artefact contributes to the good.

In an industry based on creativity, innovation and originality, pushing the envelope is valuable. But problems await us if we make technology for reasons unrelated to making the world a better place.

Adherence to ethical principles should trump all other opportunities and concerns. No technology is useful or ground-breaking enough to justify ignoring ethical concerns.

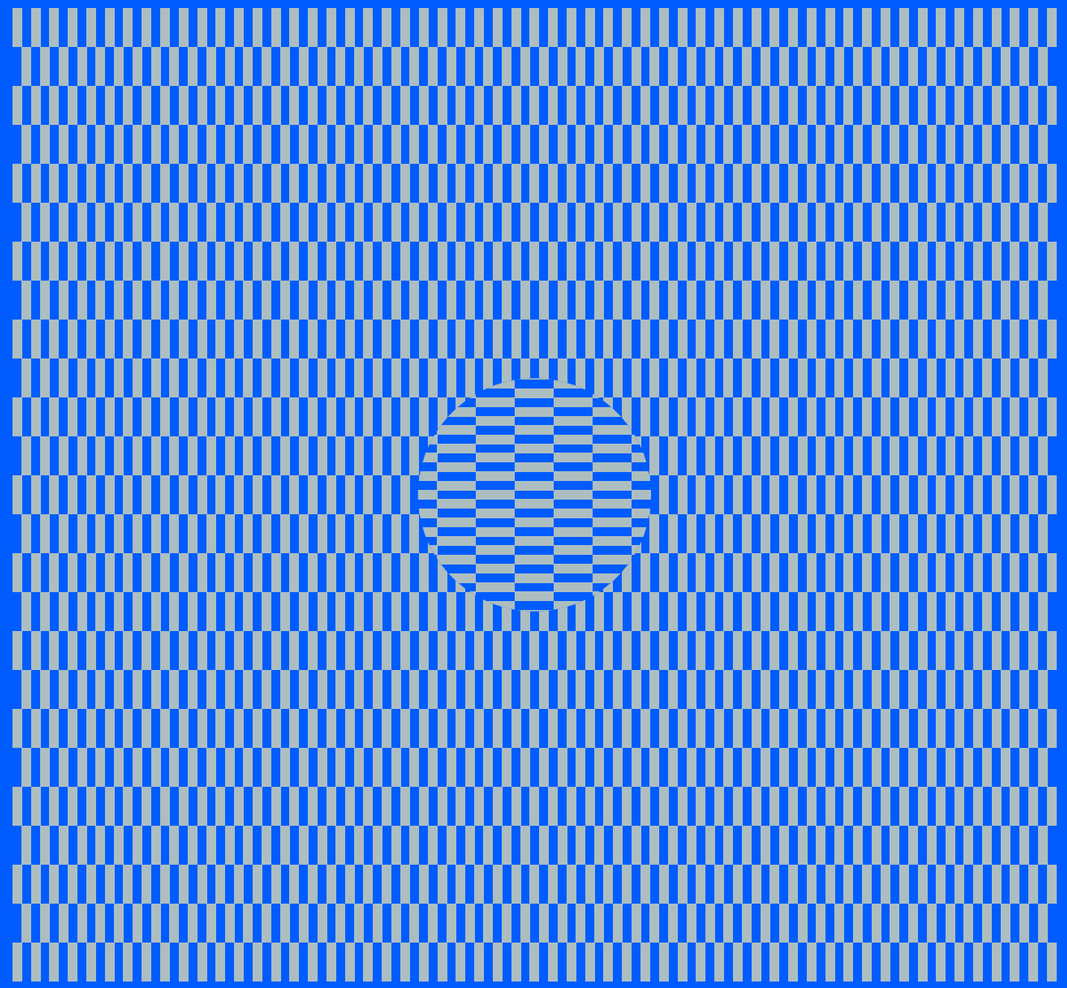
In *Jurassic Park*, mathematician Ian Malcom contemplates the technological feat of resurrecting dinosaurs. He concludes: "Your scientists were so preoccupied with whether or not they could, they didn't stop to think if they *should*."

Technology designers should see ethical responsibility and reflection as preconditions for their work. Otherwise, the dinosaurs may be let loose in the park.

[13] Richard Buchanan, 'Design and the New Rhetoric: Productive Arts in the Philosophy of Culture', Philosophy and Rhetoric, Vol. 34(3), 2001, pp. 183-206

PR01.

NON-INSTRUMENTALISM



Never design technology in which people are merely a part of the machine.

Some things matter in ways that can't be measured or reduced to their utility value. People, ecosystems, some kinds of animal life and political communities shouldn't be used as tools that can be incorporated into design. They must be the beneficiaries of your design, not elements of a machine or design system.

Lots of things in life appear to us only as tools. They are *instrumental* goods, whose value to us lies in their usefulness. There are many examples. Perhaps the most salient one is money. No reasonable person desires money for its own sake. Its value lies in what it allows us to do: to buy the things we want or need. Currency which is no longer accepted legal tender – like one cent coins in Australia – holds no value. Other things are valuable only because we say they are – like the value a toddler's drawing has to her parents. This kind of value is *relative* value.

Some things have value beyond their use or the personal feelings of any one person. These things, even if they are literally useless or don't change in the world in any measurable way, still hold value. Such things are said to be *intrinsically* good. An example is friendship. C.S. Lewis once wrote that "Friendship is unnecessary, like philosophy, like art, like the universe itself… It has no survival value; rather it is one of those things which give value to survival." For Lewis, friendship isn't good because it's useful; it's good in and of itself.

What Lewis recognised – and what technology must also acknowledge – is that there are certain things of basic value. If we treat them as though they are *mere* tools, we mistreat them. This is not to say intrinsic goods can't be instrumentally useful – there *is* survival value in friendship. But to assume this is *all* that makes friendship – or art, the environment, people or political communities – valuable is to miss the point.

Teleological ethics argues that we ought to treat things in a manner consistent with their nature and purpose. Furthermore, deontology argues that if something has dignity we must treat it as an 'end in itself'. We cannot pretend it was a mere tool (means to our ends).

Technological artefacts aren't immune from these moral prescriptions. Technological design should recognise and treat what is intrinsically valuable as such. Although it might be practical to treat everything as quantifiable, tradeable or fungible for the purposes of technology design, ethics requires us to do more. It asks us to measure our actions by reference to what's unconditionally good.

One way to restrain our actions toward what's inherently good is by showing respect. Respect in this context refers to acknowledging the intrinsic value certain things have. Philosophers have used several terms to express this value. For many years, Christians argued that human value stemmed from the *imago dei* – the likeness of God in which humans were designed. In secular spaces, human dignity and rights have often been the basis of these kinds of value claims.[14] What is at the core of the value of respect is that we treat things in a way that reflects their moral status.

Often, the objects of intrinsic value will be people. In other cases, we should consider higher order animals, ecosystems, great works of art or natural wonders as such. For example, the Whanganui River, in New Zealand, has been granted the same rights as a human being. A dam that instrumentalised a priceless river, like the Whanganui, is open to being considered unethical.

Non-instrumentalism demands we avoid using things of intrinsic value as tools. We must consider things of intrinsic value as the *beneficiaries* of technological design rather than as 'cogs in the wheel'.[15] This means refraining from using things of intrinsic value as functional or networked elements. Design should avoid the temptation to include humanity and other things of limitless value within the purview of techno-logic.

[14] The German philosopher Immanuel Kant distinguished between different ways of assigning value to an object. Things we use to meet our needs have a market price; things that entertain and delight us have a fancy price; and things that have value in themselves – regardless of their effect in the world – have dignity. Things with prices can be exchanged with other things of equal value. By contrast, things with dignity are irreplaceable. They are of infinite and intrinsic value. As a result of this, if something has dignity we cannot use it as a means to achieving our ends. We must treat it as an "end in itself". C.f. Immanuel Kant, Groundwork of the Metaphysics of Morals, (New York: Cambridge University Press, 1997), 4:434-4:435

[15] A variant on this principle can be found in the German Federal Ministry of Transport and Digital Infrastructure Ethics Commission, 'Automated and Connected Driving', June 2017, Accessed: http://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

Rules of Thumb

01. DON'T REDUCE PEOPLE TO THE STATUS OF MERE 'THINGS'

Design should value people – and other possessors of intrinsic value – more highly than instrumental goods. Technology shouldn't treat intrinsic goods as transactional or instrumental. It should not make trade-offs in favour of instrumental goods if it means causing harm to something of intrinsic value.

02. SACRIFICIAL TECH

A decision-making system may at times have to weigh its own value against something of intrinsic value. At these times, the technology should prefer to protect what has intrinsic value.

03. PEOPLE AREN'T PARTS

Technology must not reduce something that is intrinsically valuable to the status of being a mere element within the technology. Things with intrinsic value aren't *part* of technology, they are part of the *purpose* for which technology exists.

04. DON'T MANIPULATE

When we manipulate people, we treat them as a tool to achieve our goals. This immediately means we're failing to respect them as we should. Design that intends to manipulate or exploit its users or others it affects, is unethical.

05. PROMOTE DIGNITY

Technology should aim to help people be treated as they should. This may mean making prescriptive design decisions that partially limit people's liberty. We are not obliged to help people make undignified choices.[16] Technology should not let people treat themselves in ways inconsistent with their dignity. For example, an online platform that let people sell themselves into slavery would fail this rule.

[16] A famous test case is that of Bernd Brandes, who consented to being killed and cannibalised by Armin Meiwes. Despite wanting to be killed, most would argue that those expressing desires like Brandes – even if of sound mind – should not be permitting to carry through such an action.

Ethical Design Challenges[17]

[17] Throughout this document, we will list ethical design challenges and solutions that correspond to each of the principles. However, we don't expect these lists to be exhaustive. There will be other challenges and questions worthy of asking at the design stage, and other solutions we have not considered here. These questions and solutions can be used, but design teams and other subject matter experts will be able to identify others. For two examples among many, see: The Markkula Centre, 'Ethics in Technology Practice, Accessed: https://www.scu.edu/ethics-in-technology-practice/ and Artefact Group, 'The Tarot Cards of Tech', https://www.artefactgroup.com/the-tarot-cards-of-tech/

Ethical Design Solutions

01. VALUE HIERARCHIES

Some systems assign value on a simple numerical basis. This risks failing to make important value distinctions. Designers could instead create a hierarchy of value classes, preventing trade-offs from being made 'up the ladder'. This would prevent a trade-off between an intrinsic and instrumental good being made in favour of an instrumental good, even if it offered greater overall benefit.

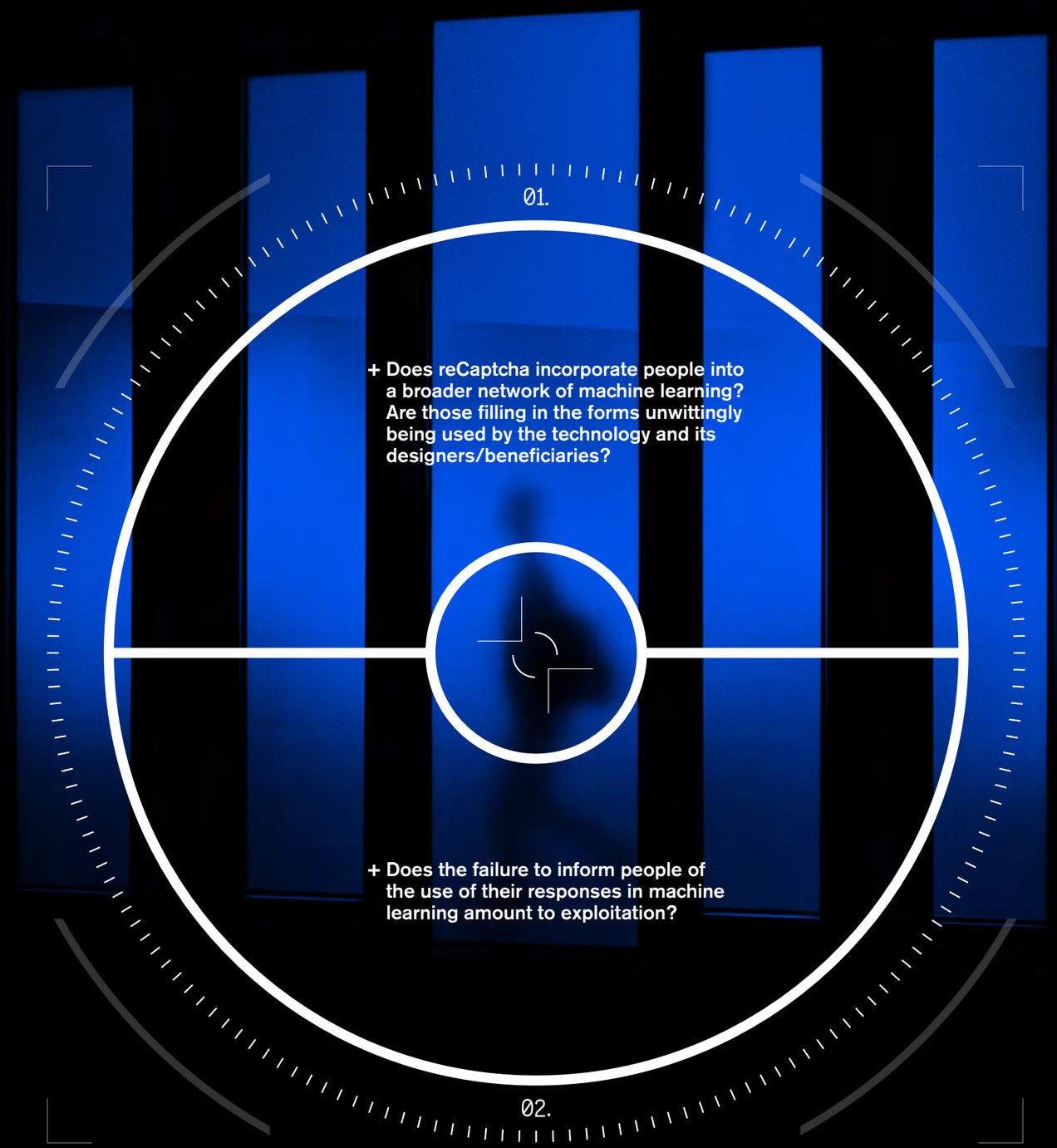
02. EXPLICIT, ENTHUSIASTIC CONSENT

Design should make it clear to users to what they are consenting. At routine intervals, it should also reconfirm their consent. This consent is only valid if it is provided in light of all the relevant information.

03. PROTECTED BY DESIGN

Some goods might be so important that their protection is explicitly designed into the technology. For instance, an augmented reality (AR) game might recognise the value of a particular environmental landmark and design it as a 'no fly' zone, where the AR does not function at all, driving users away or encouraging them to engage with the environment as it is – something to be appreciated, not modified or used.

Ethical Design Challenges



Ethical Design Solutions

Case Study: Recaptcha

Since its release in 2007 and acquisition by Google in 2009, reCaptcha has become a common tool for human verification.

It is used to ensure an online form is being completed by a human rather than a piece of software. The technology involves retyping hard-to-read words or selecting amongst a series of images to identify a common element ('click the images with a lion in them'). Unbeknownst to most users, their responses were being used to train AI in script and visual analysis.

01. NO CAPTCHA RECAPTCHA

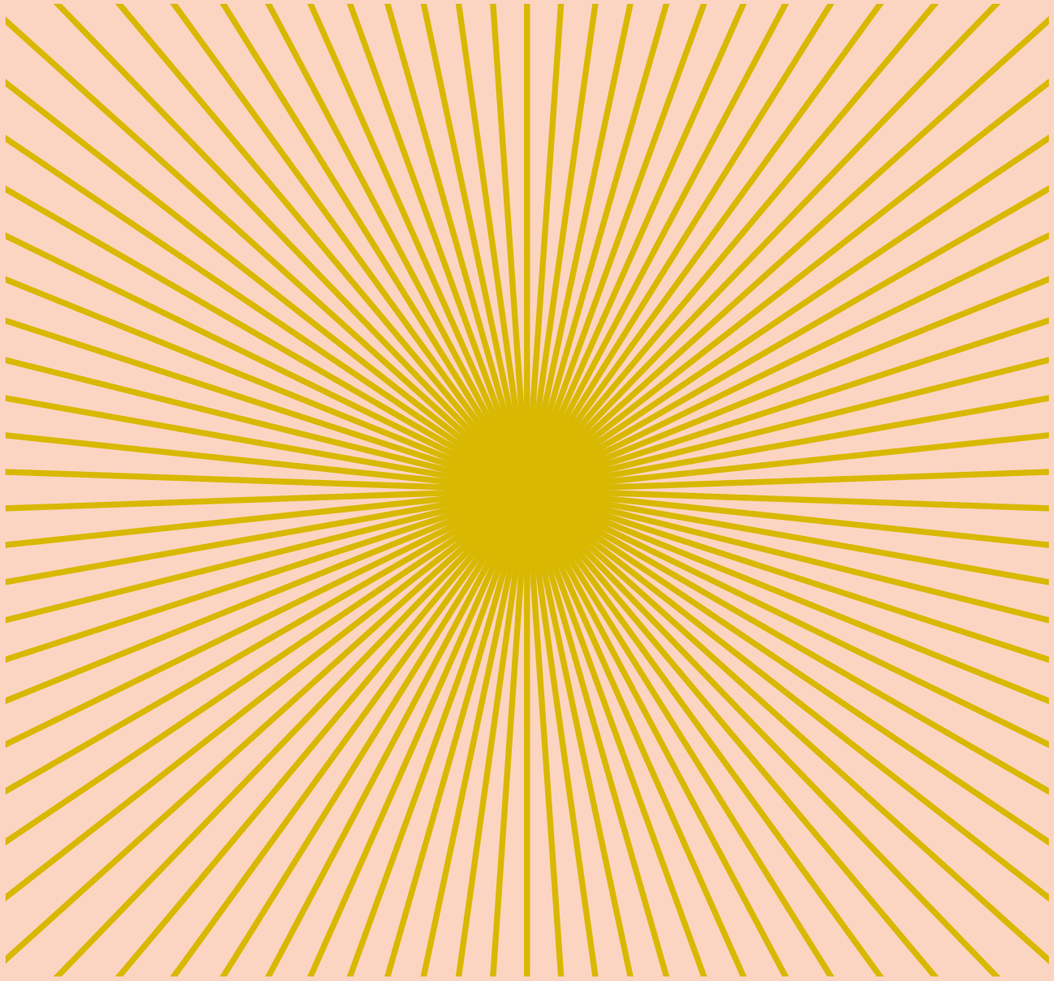
Google has now released a 'No Captcha reCaptcha', which does not require users to answer questions to verify their existence. Instead, they click a box saying, 'I'm not a robot'. This triggers a verification process based – as far as is known – on mouse movements, browsing history, IP address and other behaviour. There is still a question of whether this information is being collected or used, and how aware users are of this. However, insofar as users are no longer being used to train AI, this update may mean reCaptcha is better aligned to non-instrumentalism.

02. INFORM USERS

ReCaptcha could have informed users about the use of their responses. They might also have permitted users to opt-out. They would still be able to verify themselves as human, without having their responses used to train AI. This would prevent the abuse and would probably still provide plenty of data for machine learning purposes, without instrumentalising people.

PR02.

SELF-DETERMINATION



Maximise the freedom of those affected by your design.

Technology is meant to be an extension of human will. It's meant to empower us to achieve goals we otherwise couldn't. Technology can't achieve this goal if it interferes with our freedom. We need to make design choices that support people's ability to make free choices about how they want to live and engage with technology. But remember: maximising freedom doesn't always mean maximising choice – sometimes too much choice can be paralysing.

Almost all traditions within philosophy and the humanities include freedom as a key aspect of what it means to be human.

There are several different reasons for this. Autonomy is often argued to be the basis for dignity. Freedom and control over our lives is important for happiness and human flourishing. Free will – the ability to choose between alternative courses of action – is a cornerstone of ethics and moral responsibility.

All this leads to the conclusion that part of what makes humans unique is that we are not machines. We are free; able to act contrary to our instincts and do things people couldn't predict. Without this spark of freedom, the entire field of ethics would be redundant.

Thus, almost all ethical traditions give the autonomy and liberty of people significant weight. Indeed, our freedom is so important most agree we cannot give it up – even voluntarily. For instance, we cannot sell ourselves into slavery.

These observations around human agency are especially pertinent for technology. Insofar as i) technological artefacts are designed to assist humanity and ii) autonomy is a key aspect of humanity, then iii) artefacts should not negatively interfere with human freedom.

Instead, technology should support and empower human freedom. This includes, but is not limited to, political freedom – and raises ethical 'red flags' for technology that restricts choices, manipulates on an unconscious level or reframes social, cultural or moral norms. Technology that undermines our ability to decide for ourselves how we should act, what choices we make and so on, is in many ways as totalitarian as that which undermines a political regime.

Rules of Thumb

01. NON-USER IMMUNITY

Allow people to choose which technologies will shape their lives and to what extent. Design that fails to preserve the agency of non-users, fails to protect self-determination. For instance, there are ethical questions to be answered about technologies that change how people use community spaces or include non-users within a network without consent[18]. For example, are there ever good public policy reasons for restricting autonomy – for example, when managing the spread of a lethal epidemic (like Ebola)? What then, might be the justifiable exceptions to this rule of thumb?

02. USER SOVEREIGNTY

The users of technology hold the rights to any products that emerge from its use. These may include data, artwork, intellectual property or any range of other goods. Ownership rights acknowledge the time, labour and imagination users invest in using technology. Users may transfer ownership through ethically and legally acceptable processes. These may include readable, transparent End User Licensing Agreements, payment or any number of other arrangements.

03. NUDGING

Nudges are design choices that consciously influence the decisions people make. The use of nudging is growing more common in technology. Some of these nudges are transparent and obvious, such as when a fitness app sends you a notification telling you to go for a walk. Others are more subtle, as when a video streaming service autocues the next video immediately following the end of the last one. This reframes the nature of your choice about whether to keep watching or not.

The ethical considerations around nudging are myriad. Trade-offs between freedom, benefit and efficiency will need to be assessed on a case-by-case basis in order to determine whether self-determination is meaningfully enhanced or detracted from by a particular nudge. However, a clear understanding of the purpose, beneficiary and effects of the nudge will help determine its ethical character.

[18] This principle does not govern whether the state might compel citizens to use a certain piece of technology. This question is important, but it pertains more to political philosophy and ethics than the ethics of technology per se.

Ethical Nudging

Nudging is a clear challenge for self-determination. It targets the non-rational aspects of human decision-making to achieve desired outcomes. It thus risks treating people not as humans but as “conditioned animals” – a term coined by philosopher Hannah Arendt.

This risk is manageable. To an extent, we are conditioned animals. Human beings are not purely rational. Our choices, whether mediated by technology or not, are always framed in non-neutral ways – intentionally or otherwise. It would thus be impractical to attempt to banish nudges. It would also be exhausting on decision-makers to always make them aware of every way a nudge was acting on them. We could hardly get through the day constantly being informed of the ways, say, urinal design influences our toilet habits, how nutritional information encourages us to eat more healthily or whatever. This information is unlikely to meaningfully contribute to free decision-making.

Still, before using nudges, designers should consider the following rules of thumb:

01. NUDGE TOWARD RATIONALITY

Many philosophers believe non-rational elements of human thinking to be ‘pathological’. This frames the subconscious and non-rational parts of us as a blight on good thinking. Nudges that aim to reduce non-rational, unethical influences on thinking are justifiable. For instance, nudging against biases that cause discrimination help us act more rationally, not less.

02. NUDGE TOWARD VIRTUE

Some nudges may aid a person or community in becoming the ideal version of themselves. For instance, nudges that encourage organ donation may push a community toward living solidarity, care for others and generosity.

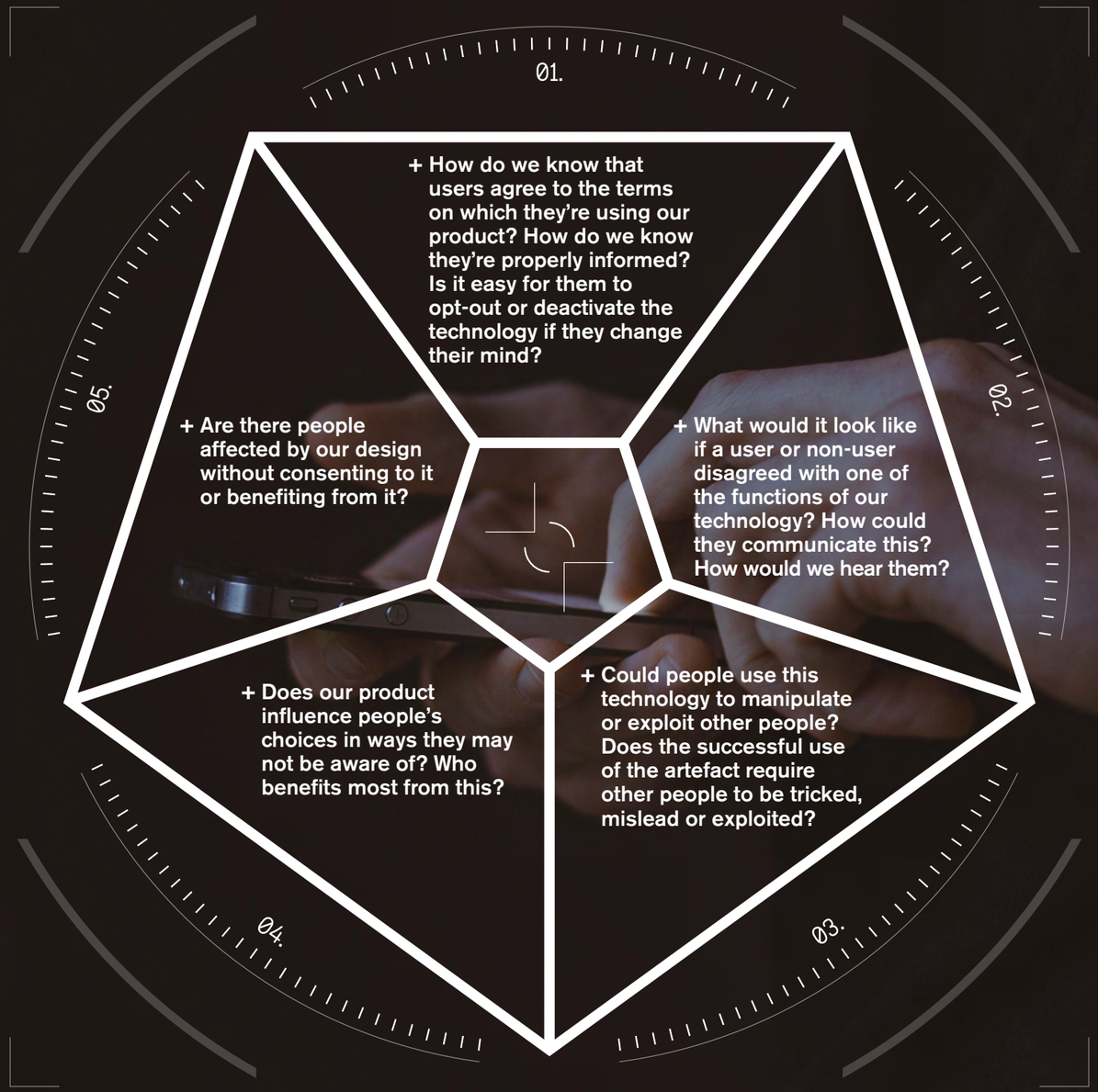
03. NUDGE TOWARD A ‘PLAN OF LIFE’

A ‘plan of life’ outlines how a person will harmoniously meet their various goals. It focusses on balancing various needs, goals and objectives across a lifespan. At times, short-term needs and desires can undermine a plan of life. Nudges might help to ‘right the ship’, directing people toward the goals we know they care about, even if their actions do not align with them in the moment. Note, designing these nudges requires designers to know a user’s plan of life. We should not assume to know someone else’s plan of life in abstraction.

04. NUDGE WITHOUT BENEFIT

The political philosopher John Rawls said inequality could only be justified if it offered the most benefit to the least well-off. Great wealth mustn’t ‘trickle down’, it must close the gap between the rich and poor. Benevolent nudging means nudging to benefit the subjects of nudging rather than the designers, owners or those in power. Thus, nudges that are justified by commercial instincts would become difficult to defend.

Ethical Design Challenges



Ethical Design Solutions

01. OPT-IN PROCESSES

Seeking explicit, informed consent from those affected by the design of an artefact will take significant steps to address issues of self-determination.

02. ACCESSIBLE EULAS

Self-determination requires informed choice. Rights and responsibilities, as outlined in End User Licensing Agreements (EULAs) must be presented in an accessible way that creates genuinely informed consent.

03. CONSENT-BASED NUDGING

Some forms of nudges, such as life-plan nudges, will be best preserved self-determination by seeking user consent.

04. BEST INTEREST NUDGING

When explicit consent is impractical or impossible, nudging should be made in the best interests of the people being nudged.

05. USER CONTROL

User sovereignty can be preserved by providing users full control over the outputs of their use of the technology – data or otherwise – as well as the means to make transactions and exchanges of ownership of these outputs.

Ethical Design Challenges

Ethical Design Solutions

Case Study: Pokémon GO

Pokémon Go is a mobile-based augmented reality game that allowed people to roam the physical world catching Pokémon on their devices.

In 2016, Pokémon Go drew huge groups of people to prominent landmarks, which were featured in the game. Algorithms would also randomly spawn Pokémon in a variety of locations – some of which were private properties. This disrupted traffic, businesses and religious services. Whilst the administrators or owners of landmarks could 'opt out', they were included in the technology without their consent.

01. OPT-IN

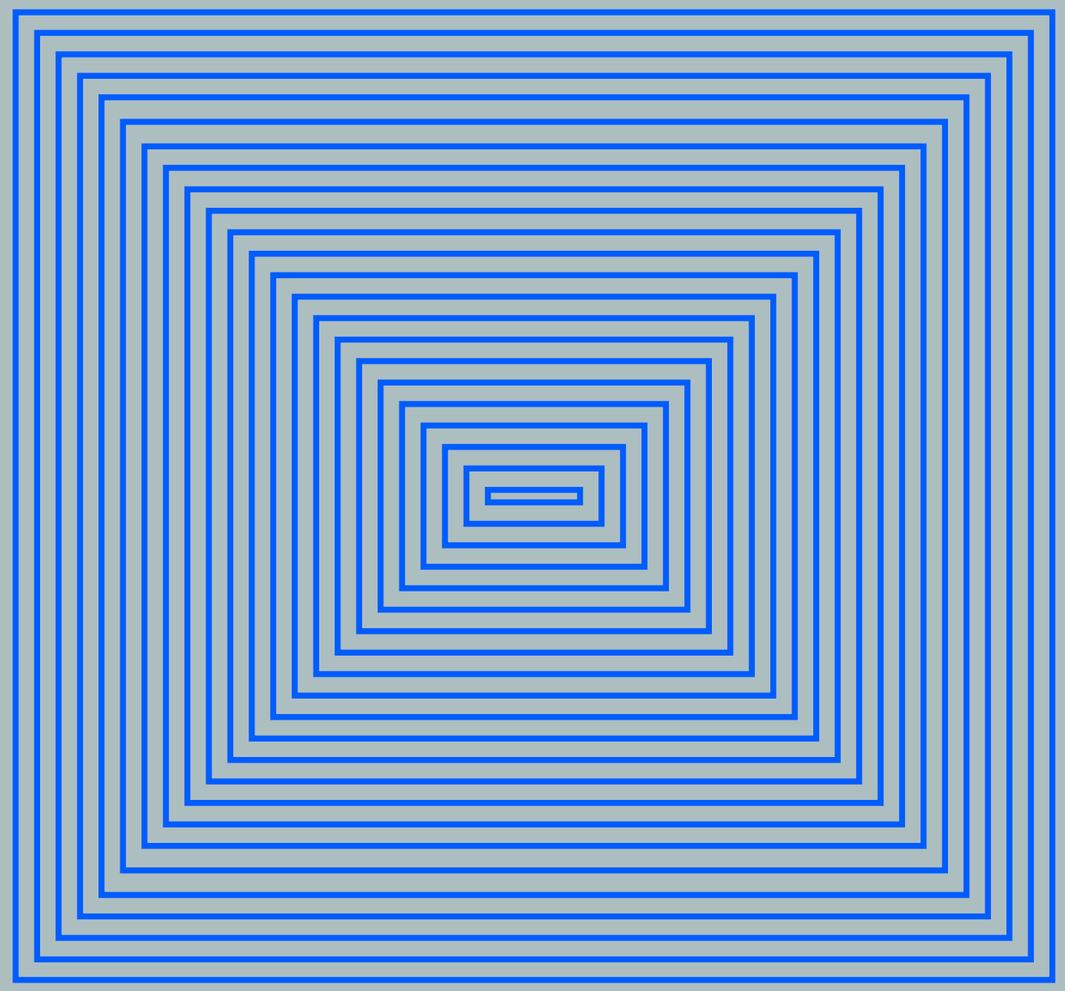
Pokémon Go implemented an opt-out for key locations like Pokémon gyms. However, this process required business owners to provide a range of information – proof of ownership, photos, latitude and longitude. The process was time-consuming and difficult for some to access – putting the onus on non-users to earn their right against inclusion. An opt-in process may have been more preserving of self-determination.

02. ENCOURAGING HEALTH

The gamification of Pokémon Go rewarded users who walked whilst using the game – it enabled them to hatch 'eggs' and nudged users toward physical movement and health. However, the app needed to be open when these steps were being taken, making it difficult to walk quickly or safely, but helping keep people on the app. Best interest nudging might have permitted the app to keep tracking steps with a locked screen to more fully encourage health and engagement with the world.

PR03.

RESPONSIBILITY



Anticipate and design for all possible uses.

Technology is usually designed with a specific use case – or set of use cases in mind. Problems often arise when users deviate from the intended use case. Often, it's entirely possible to predict the different ways people will use our designs, if we take the time to think it through. Failing to imagine alternate uses and their implications is risky and unethical. Doing so can alert us to potentially harmful uses we can safeguard against, or potential benefits we can maximise through good design.

Just as designers and users have shared custodianship of the purpose of a technological artefact through the process of *proposing-mediating-purposing*, responsibility is a shared concept in the ethics of technology. Unlike the instrumentalist, who puts sole responsibility on the user, or the determinist, who holds the designer entirely to account, we argue that assigning responsibility is a more fluid and complex – though no less important – process.

Designer Responsibility

Designers are responsible for the aspects of technology developed at the proposing stage of design. These include:

+ **INTENDED PURPOSE:** Designers have some end in mind when they design artefacts. The extent to which this end is an ethically defensible one is the responsibility of designers.
+ **NUDGES:** Nudges operate by undercutting the agency of the user. This means the designer must take additional responsibility – for good or ill – for what happens as a result.
+ **AFFORDANCES:** Designs can be more or less open to particular kinds of use by way of affordances. The extent and ways affordances facilitate ethical or unethical behaviour is the responsibility of designers.
+ **REPURPOSING:** Whilst designers are not responsible for the ways users might repurpose their work, they are responsible for the openness of their designs to repurposing. In some cases, they may be accountable for having their designs be too open, in other cases for being too closed.[19]
+ **VALUES:** Both explicit and implicit values will be embedded into the design of an artefact. Designers are responsible for these values, the ways they are prioritised, their effect on the communities in which they will be used, and so on.[20]

[19] The question of openness to repurposing has been widely explored in research ethics, where it is known as the 'dual use' problem. Dual use problem occurs when scientific research can be "used in both morally desirable and undesirable ways… and the risk of undesirable use is sufficiently high that it is not clear that the agent may permissibly pursue the project." Dual-use technologies are usually seen as problems to be managed. The questions concern how to best minimise the risks of abuse whilst also enabling the research to be used as intended. See: Thomas Douglas, 'The dual-use problem, scientific isolationism and the division of moral labour', Monash Bioethics Review, 2014, Vol. 32(1-2), 86-105 at p. 86

[20] A more extensive list of values-based design considerations can be found on p. 23

User Responsibility

Users are responsible for aspects of the mediating and purposing stages of technology. Areas of user responsibility include:

+ **AFFORDANCE RESPONSES:** Affordances invite particular modes of use. For instance, an affordance can refuse particular kinds of use. However, a skilled user may be able to bypass some refusals, transforming the affordance into a discouragement (for instance, people who know how to pick locks turn the act of opening the door without a key from a refusal to a discouragement). The extent to which a user heeds and responds to affordances is their responsibility.
+ **ACCEPTING PURPOSE:** Technological artefacts can be designed with an illegitimate purpose in mind. This is something for which designers are held accountable. However, if a user knowingly uses the technology to realise a goal that is inherently wrongful, they should be held responsible for this.
+ **REPURPOSING:** The open nature of technological purpose means it is often possible for users to repurpose technology. The ways in which an artefact is repurposed are primarily the responsibility of the users who do so. When this is done well, credit should flow to the user;[21] when it is done badly, responsibility and accountability should do so as well.
+ **PERSONAL IMPLICATIONS:** The extent to which technology use benefits or detracts from a user's life will vary. Designers cannot account for every possible implication – nor should they be held responsible for them. Users are responsible for their choices to use technology in certain contexts or situations – for instance, a designer is not responsible for making a video game so compelling a user lost track of time and missed a job interview. The only caveat to this concerns nudging and affordances: if designers aim subconsciously to influence decision-making, they may be responsible to the extent that they have subverted agency (partially or entirely).

Shared Responsibility

There are some elements of technology use where it is not possible to lay blame solely at the feet of one party. In these cases, responsibility needs to be shared. These include:

+ **DEMAND:** Designers will often claim that market demand justifies their producing a certain product or for doing so in a particular way. In these cases, the users are indeed responsible for the demands they make; however, designers are responsible for the extent to which they accede to those expectations. They can, and perhaps should, say 'no'.
+ **IMPACT ON NON-USERS:** Designers may fail adequately to protect non-users. In such cases, they are primarily responsible for this design decision. However, if users knowingly use the technology, despite the effect it has on non-users, they share in the harms that may result.

There are limits to responsibility. First, designers cannot be blamed for genuinely unforeseeable uses or effects of technology. Second, if developers can foresee potential abuses of their technology but have no way of preventing them from occurring (think of those who produce hammers that are used as a weapon to kill), they may still be justified in developing the technology, assuming that:

A. The benefits outweigh the foreseeable harms, and they do not intend those harms to occur and

B. They have taken all reasonable steps to prevent those harms from occurring.[22]

[21] This is one reason for defending user ownership rights, see above.
[22] See the Net Benefit principle below for more on this.

Rules of Thumb

01. UPDATE AND RECALL

Designers should continue to improve their technology in order to progressively reduce the potential harms and maximise the benefits. If unanticipated harms occur, designers have an obligation to address these as quickly as possible, whether through update or recall.

In the event that the original designers are unable to serve this role any longer, they must transfer this responsibility to another capable person or group. If they are unable to do so, they must make users aware that the product is no longer being supervised and encourage them to stop using it (the same way lifeguards notify swimmers of an unpatrolled beach and urge them not to swim) or shut it down (as when a beach is closed for reasons of safety). If a user continues to use technology, despite either effective warning or attempted shut-down, the user bears full responsibility for the consequences of their use.

02. RIGHTS CUSTODIANSHIP

Where the use of technology requires users to hand over some element of their moral rights, designers must respect those 'borrowed rights' in the same way they would want their own rights to be respected. Technology design must respect user sovereignty and thereby accept custodianship of user property and data.

03. TRANSPARENT PROVENANCE

The complete history of artefacts and devices, including the identities of all those who have designed, manufactured, serviced and owned the item, should be freely available to any current owner, custodian or user of the device. The provenance should include all details of all updates (or amendments) to the technology. The record should be subject to verification and be indestructible. In cases of data, the provenance, sources and metrics necessary for quality assessment should be made available.

04. MANAGE REPURPOSING

Designers should ensure they have taken all reasonable, ethically permissible steps to minimise harmful repurposing. This may include designing the technology in ways that 'nudge' against abuse or restricting the extent to which design affords repurposing.

Ethical Design Challenges

Ethical Design Solutions

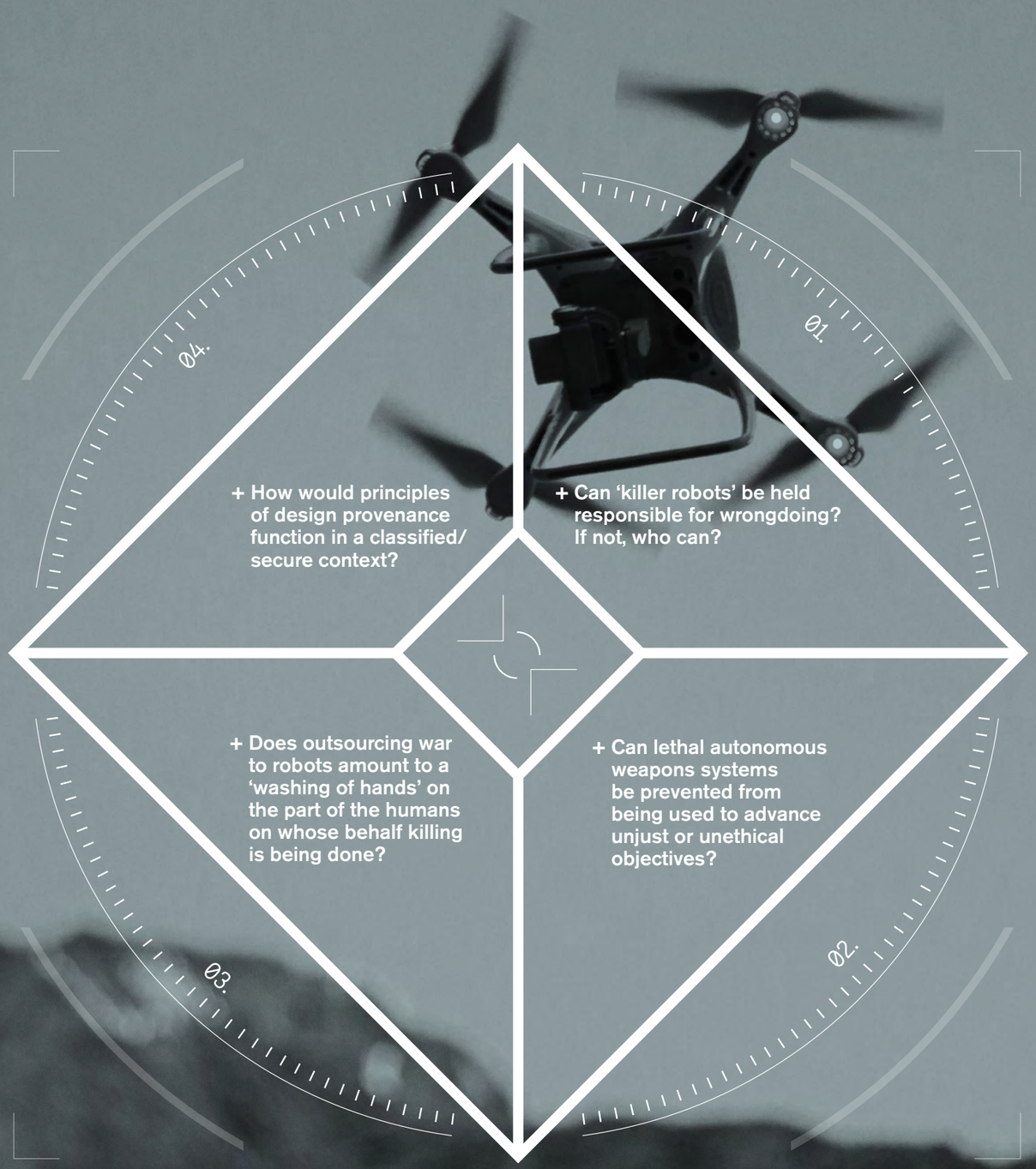
01. USAGE TRACKING

The ability to track and monitor the different ways a technology is being repurposed will allow designers more readily to manage harmful repurposing.

02. LEDGERS

A ledger that can accompany or be linked to specific artefacts (and classes of artefact), detailing: updates, ownership, data collection and other relevant information would provide users with a clear understanding of provenance.

Ethical Design Challenges



Ethical Design Solutions

Case Study: Killer Robots

The design of ‘killer robots’, for use in military operations, has been a subject of heated debate in both technology and military circles.

These machines might minimise combatant deaths and trauma by minimising the number of people on a battlefield. They might also reduce civilian casualties because of superior decision-making.

However, many worry these robots would be unable to account for the limitless variables of war. They might also create an ‘accountability gap’ wherein it becomes unclear who is responsible for ethical failures on the part of robots. This becomes especially concerning if lethal autonomous weapons fall into the hands of malevolent actors.

Another concern is whether certain acts, such as war and killing, are ones for which we ought to be fully responsible. The French existentialist, Albert Camus, argued that killing cannot be justified unless the killer is also willing to die. Is this what is required to take genuine responsibility for war?

01. HUMAN ARBITER

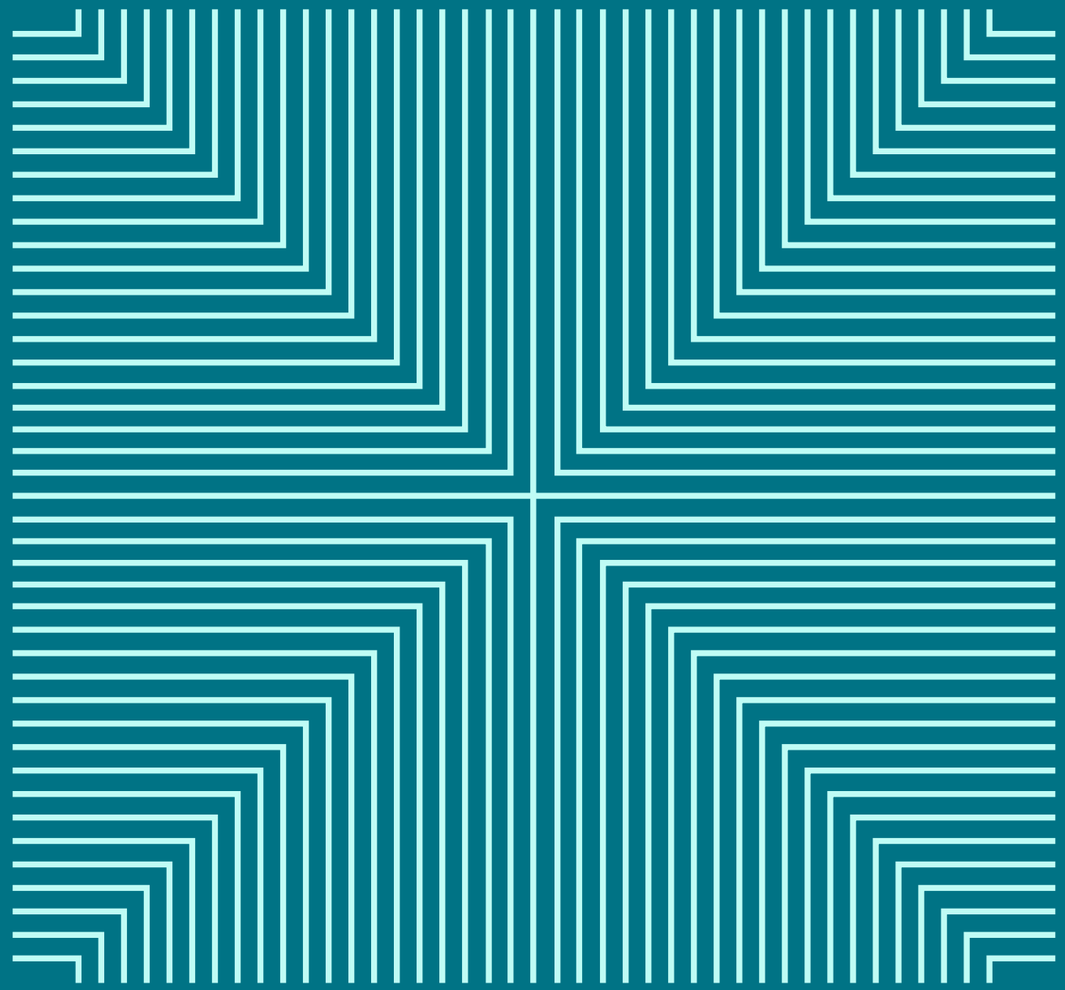
The most frequent solution proposed is to retain a final human arbiter of any lethal action taken by a robot, as is the case with armed UAV’s, which still bear a pilot who delivers the ‘kill shot’. This makes it clear who is accountable for the action taken.

02. IMPLEMENTING GOVERNING PRINCIPLE

Implementing the ‘should before could’ principle: some have argued against creating lethal autonomous weapons systems at all, thus removing the ethical risks involved.

PR04.

NET BENEFIT



Maximise good, minimise bad.

The things we build should make a positive contribution to the world – they should make it better. But more than this, we should also be mindful of the potentially harmful side-effects of our technology. Even if it does more good than bad, ethical design requires us to reduce the negative effects as much as possible.

Given the potential for technology to achieve a variety of ends, designers must be confident their work is of some demonstrable benefit.

"Benefit" is a broad term, and should be measured in relation to the following criteria:

+ **HUMAN WELLBEING**
+ **INDIVIDUAL FREEDOM**
+ **IMPROVEMENT OF THE MORAL CHARACTER OF PEOPLE**
+ **THE COMMON GOOD**
+ **INSPIRING WONDER AND HOPE FOR THE FUTURE**
+ **ACCESS TO BASIC NEEDS**
+ **PROTECTION OF HUMAN RIGHTS AND DIGNITY**
+ **THE STATE OF THE ENVIRONMENT**

Measuring benefit can be complex.

Designers should consider their impact on each of these criteria – there may be benefits located in some and harms in others. Trade-offs may be justifiable, but only if all the effects are known.

Although outcomes aren't the only things that matter, this principle acknowledges them as one important factor among many. The mere fact a piece of technology achieves some benefit is not enough. It must also meet the other requirements outlined in this paper.

This principle also requires makers of technology to actively attempt to reduce harms as much as reasonably possible – even if in doing so there is a corresponding decline in the amount of benefit available.

Where harms are inescapable, these may be justified under a philosophical concept known as **double-effect theory**, which suggests unavoidable harms may acceptable if:

+ **ONLY THE GOOD OUTCOMES ARE INTENDED.**
+ **THE GOOD OUTWEIGHS THE CORRESPONDING HARM.**
+ **THE HARM IS NOT INSTRUMENTAL TO THE GOOD BEING ACHIEVED (E.G. VIA MASSIVE INVASIONS OF PRIVACY).**
+ **THE CHOSEN MEANS ARE NOT INTRINSICALLY WRONG.**
+ **ATTEMPTS HAVE BEEN MADE – AS MUCH AS POSSIBLE WITHOUT UNDERMINING THE GOOD – TO MINIMISE THE HARM.**

Rules of Thumb

01. ACKNOWLEDGE OPPORTUNITY COSTS

Recognising net benefits as ethically serious also means acknowledging the time spent developing, marketing and using one technology is time that cannot be spent doing the same for other technology. A philosophical movement known as Effective Altruism suggests we should do the most good we can, meaning energy should be dedicated to technological development in proportion to the potential benefits each development offers.

02. ANTICIPATE SIDE-EFFECTS

Whilst the good outcomes should be the primary focus of the makers of technology, due attention must be given to foreseeable side-effects and ways to minimise or avoid them.

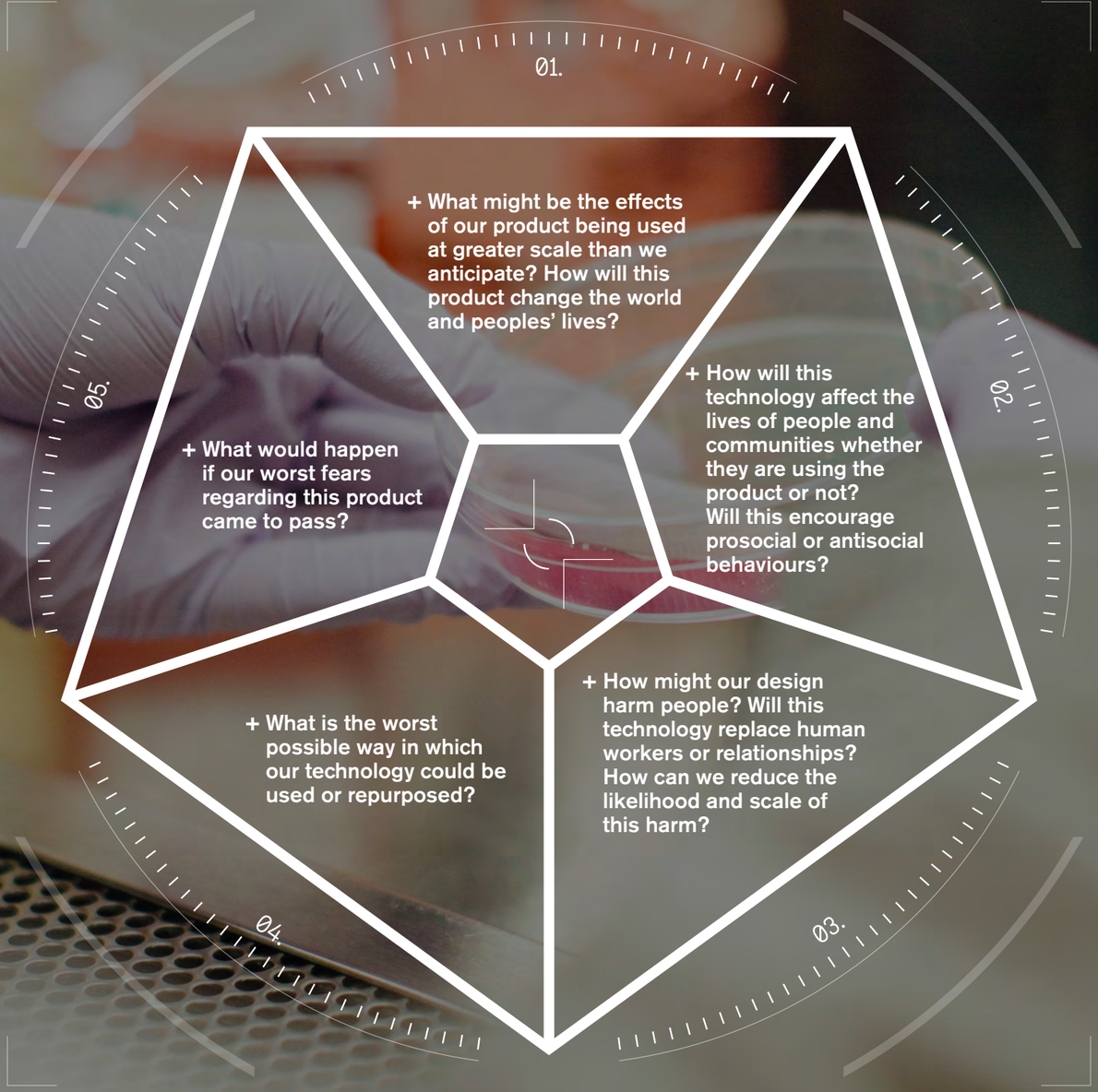
03. MINIMISE HARM

Within the purview of achieving the core purpose and benefits anticipated by the technology, potential harms must be minimised as much as possible – even if this is costly to the efficiency or effectiveness of the technology.

04. THE ENDS DON'T JUSTIFY THE MEANS

Technology has to do the right thing in the right way. It doesn't matter how much good a piece of technology might cause, it should not be designed unless it can satisfy each of the ethical principles listed in this paper.

Ethical Design Challenges



Ethical Design Solutions

01. HARM METRICS

Technology must not only measure what it *intends* to do, it must have clear ways of knowing when, and to what extent, it is causing harm. Without this, decisions to intervene or withdraw products cannot be made as quickly as possible, allowing greater harm to occur.

02. HARM QUARANTINES

Wherever possible, technology should include safeguards to stop harms from snowballing if the technology is misused. For instance, a 'circuit breaker' that disconnects a device from technology. Or, where a biological agent could be used for harm, limiting the capacity for theft or transportation by restricting access to within a laboratory setting.

Ethical Design Challenges

Ethical Design Solutions

Case Study: Mousepox

In 2001, Australian researchers attempted to create a “contraceptive vaccine” to be used in pest control.

The vaccine would stimulate mouse antibodies against their own eggs, effectively rendering the mice infertile. They chose to use the mousepox virus as a way of delivering the vaccine. However, they quickly discovered that the inclusion of the gene triggered the creation of large amounts of IL-4, a molecule that does boost antibody production but also kills the cells that fight off the mousepox virus. In lab tests, the virus, which usually only causes minor symptoms in mice, killed them all in nine days. They also discovered the virus was particularly resistant to vaccinations against mousepox.

Worryingly, mousepox is closely related to smallpox – a disease which has no treatment and against which we rely on vaccination. Were someone to use the learnings from this mousepox vaccine to pair smallpox and IL-4, they would pair one of the largest biorisks in the world with the means for destroying the body’s ability to fight off the disease. The results could be catastrophic.

01. HARM QUARANTINE

Some propose that only viruses that cannot self-replicate should be used in experimental programs such as the mousepox case, thus generating a kind of ‘harm quarantine’, allowing for the ability to control and minimise harm.

02. ENSURE AN ANTIDOTE

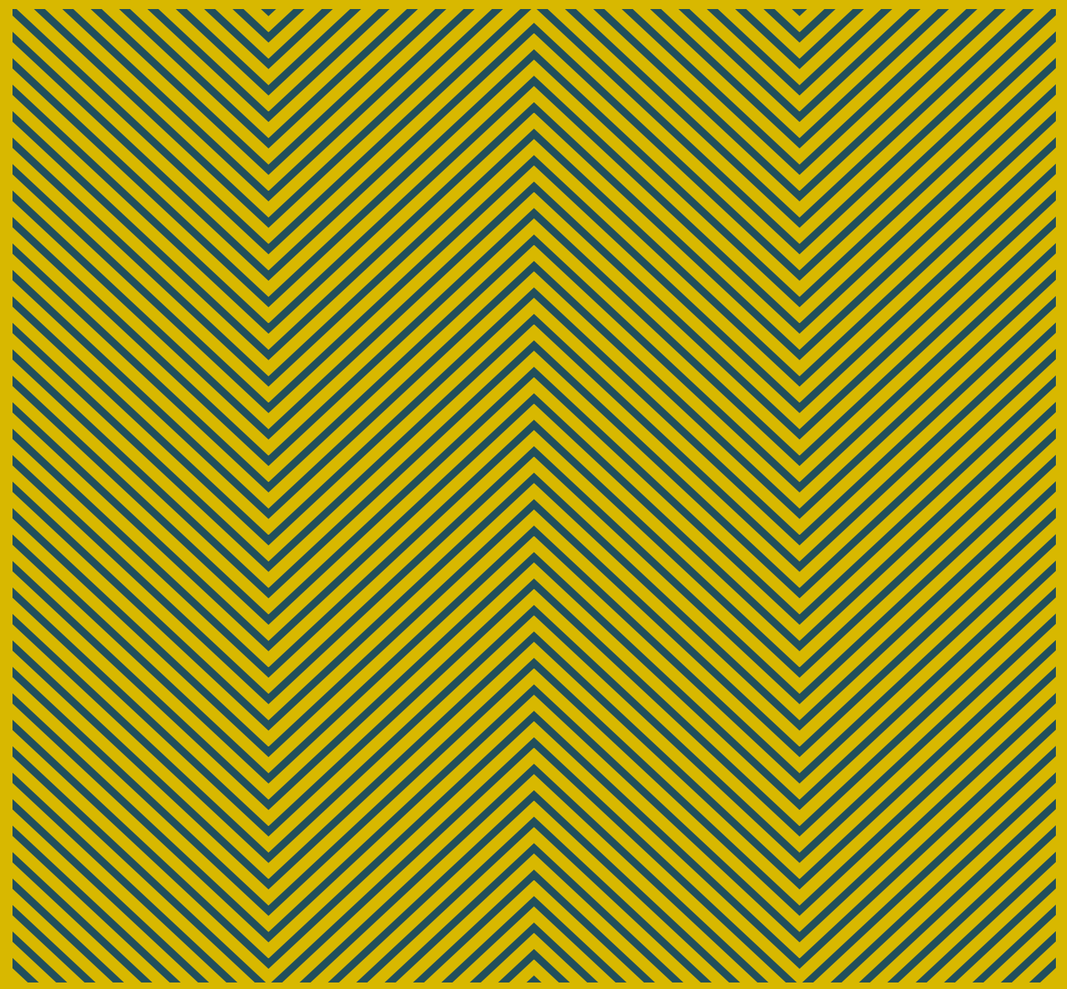
Potential harm might be offset by the adoption of a policy that required, prior to release, developers be confident of there being an antidote to their new creation – or at least that an antidote was possible. Whilst this would delay the enjoyment of the benefits for pest control, it may be morally preferable given the potential for harm.

03. OBSTRUCT BAD ACTORS

The risks inherent in this study triggered a range of reflections on how to manage sensitive scientific research. Recommendations included publishing the research without including the methodology to prevent malicious actors from accessing the designs. However, there were concerns this would diminish the reliability of the study, given critical peer review would be difficult to obtain. This is an example of the kind of reflection required by the net benefit criterion – considering how best to minimise harm without compromising on important moral opportunities.

PR05.

FAIRNESS



Treat like cases in a like manner; different cases differently.

Technology designs can carry biases, reflect the status quo or generate blind spots that mean some groups of people are treated negatively on the basis of irrelevant or arbitrary factors such as: race, age, gender, ethnicity or any number of unjustifiable considerations. Fairness requires us to present justifications for any differences in the ways our design treats different user groups. If some groups experience greater harm or less benefit than others, why is this the case? Are our reasons defensible?

The United States of America's *Declaration of Independence* begins with a series of 'self-evident truths', which serve as ethical building blocks that underpin the rest of the document. First among them is the belief that every person is of equal moral worth.

"We hold these truths to be self-evident, that all men are created equal…"

Since well before American independence, that idea has been a staple of political, ethical, theological and legal reflection. No person should receive different treatment because of unchosen or unearned characteristics. They must be treated impartially.

Technology must reflect this deeply-held belief. The technologies we utilise carry with them implicit values and shape the dimensions of our choices in various ways. It is crucial they preserve the principle of impartiality as a basic condition of the way people relate to one another.

Given the powerful role design plays in shaping user choices, designers must understand the role bias can play in their decisions. Robust beta testing and user engagement can assist in identifying and addressing biases. However, more is needed. Bias can exist even in the act of determining who is, or is not, intended to be a user of a technology.

The notion of fairness and equality goes beyond what is captured by the *Declaration of Independence*. Impartiality is a negative duty - it tells us what we shouldn't do: that is, we should not confer unfair advantage. Besides this, technology design requires a substantive account of fairness and justice. It should identify fair processes for dealing with appeals to preserve the rights of relatively powerless users against administrators who can be seen to hold all the cards. Designers of technology should seek to do good, rather than merely avoid evil.

Rules of Thumb

01. TREAT LIKE CASES ALIKE

The principle of treating like cases alike asks us to ensure the basis for distinctions is non-arbitrary and based in real and significant differences between cases. This requires the makers of technology to look beyond correlation in determining how to treat different people and ensure people are treated in a just manner.

02. TREAT DIFFERENT CASES DIFFERENTLY

The reverse is also true: relevant differences should be treated differently. The principle of impartiality also requires technology not to make false equivalences. For instance, if one person has a disability that inhibits their ability to use technology, the principle of impartiality could not justify their being excluded for the reason that they are being 'treated the same' as all other users. There is a relevant difference that needs to be considered.

03. NATURAL JUSTICE

Users seeking to complain about their experiences with an artefact must have clear avenues for doing so in a way that ensures their complaints will be heard. They must also feel confident their complaint will be heard impartially, which may require an independent appeals process, transparency of decisions, access to the standards of evidence involved and so on.

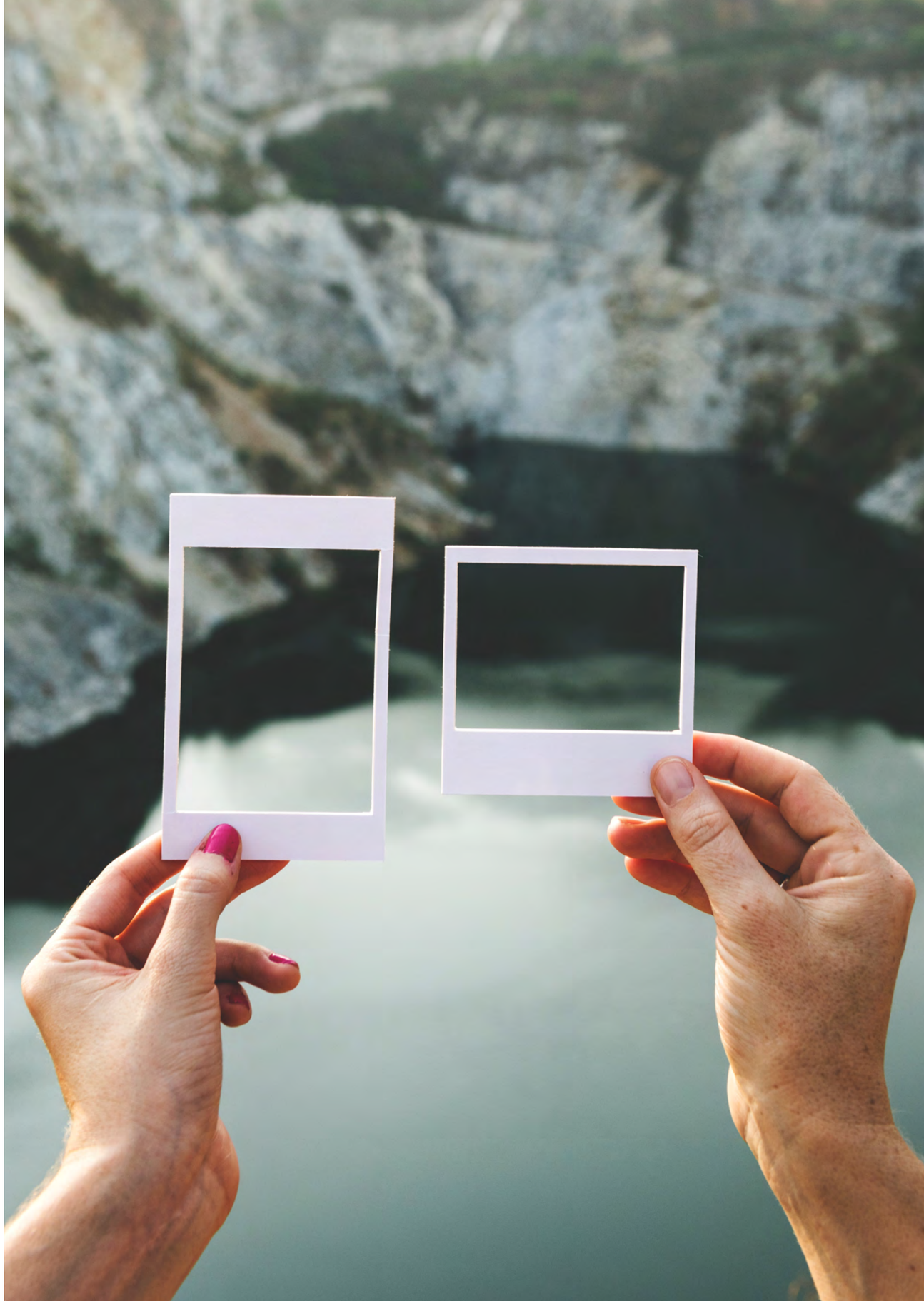
04. MINIMISE MORAL HAZARD

Moral hazard is a financial term referring to investment decisions where the person standing to benefit financially has no risk if the investment fails. They shield themselves from risk, leaving it to others who may potentially suffer.

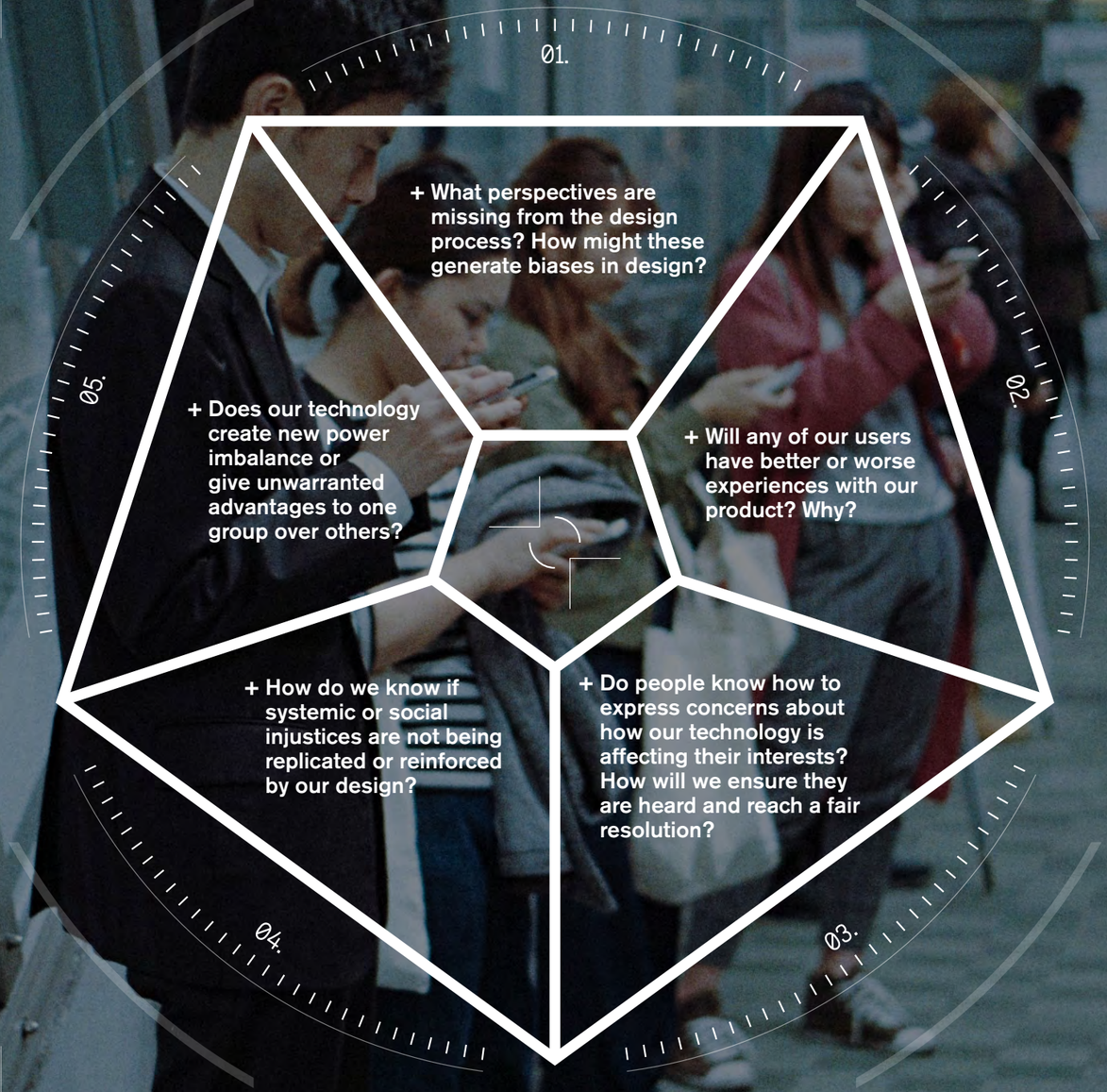
Technology can also give rise to moral hazards. It can generate advantages for one group by generating risks in another population. Such distributions of risk are unfair. If technology generates ethical risks, as a matter of design, these risks must be borne by those who are benefiting from the technology.

05. DIVERSITY IN DESIGN

Unconscious biases thrive in homogenous thinking spaces. Designers can reduce the likelihood of biases being embedded into technology by including diverse teams in the design process. This will ideally include diversity of: gender, race, ability, class, culture and ethical profile.



Ethical Design Challenges



Ethical Design Solutions

01. AUDITABLE DECISION-MAKING

Bias can sneak into technology without any malice or ill intent. For this reason, any decision-making processes, algorithms or data that help determine how to treat different users of technology should be able to be globally audited by an independent third party to determine the basis on which decisions are being made, and whether that basis is ethical. It will also be necessary for technology to have the ability to be locally audited to check whether the data inputs are representative or give rise to unfair, biased or inaccurate decisions. In cases where algorithms or data are proprietary, security protocols may not be so extensive as to prevent the possibility of an independent audit.

02. FAIRNESS MODELING

Whilst it is important to ensure technology is not explicitly designed to make arbitrary or unfair distinctions, this may not be sufficient to guarantee the technology is fair. Given the complexity of many systems, it will be equally necessary to test the outputs to determine whether results are skewed in ways that seem unjust, misrepresentative or out of line with commonly-held values.

Ethical Design Challenges

Ethical Design Solutions

Case Study: COMPAS

In some US states, algorithms and artificial intelligence are used to help decide prison sentences.

One such program is COMPAS, designed by Northpointe. Evidence has emerged of several cases where the accuracy of predictions skews on the basis of race. Black offenders are more likely to be deemed 'high risk' than white offenders also applying for parole. Although race is not one of the metrics COMPAS is coded for, the end result is racially skewed. Black people tend to receive longer punishments than white people for the same offenses.

COMPAS was 'fair' on another measure – it was equally accurate (approximately 60%) for all subjects, regardless of race. However, the inaccuracies were skewed. When COMPAS was in error regarding white subjects, it tended to underestimate the risk of reoffence; when it was in error regarding BLACK subjects, it overestimated the risk. In a further complicating factor, the two different kinds of fairness in competition here – equal accuracy for all parties vs equal severity of error – are mathematically incompatible; the COMPAS algorithm could utilise one model, but not both.[23]

COMPAS does address a significant social need. Human decision-making has proven inadequate in making parole and sentencing decisions accurately and impartially. For instance, an Australian Law Reform Commission report from 2006 found "compelling evidence of inconsistency in the sentencing of federal offenders across Australia." Thus, machine decision-making, even if not perfectly accurate, may be an improvement on current arrangements. However, these net benefits need to be proven – one study suggests human assessors were slightly better at determining recidivism than the algorithm[24] – and even if improvements are possible, they must be considered alongside very real concerns about fairness.

01. PRESENT LIMITATIONS

Judges in the Supreme Court of Wisconsin recommended that any risk assessments made by sentencing algorithms be presented alongside a discussion of the limitations of the assessment and a summary of the process by which decisions are made. Adoption of this recommendation would enable the broader justice system to identify relevant differences not accounted for by the machine process.

02. DATA ANALYSIS

Analysis of the data might reveal how many of the data sets being used by COMPAS were strongly correlated with race, thus identifying the ways in which unfairness could still be built into the process without explicitly screening for race. Then, it could be determined whether an effective program could still be designed without including these data sets.

03. FAIRNESS MODELING

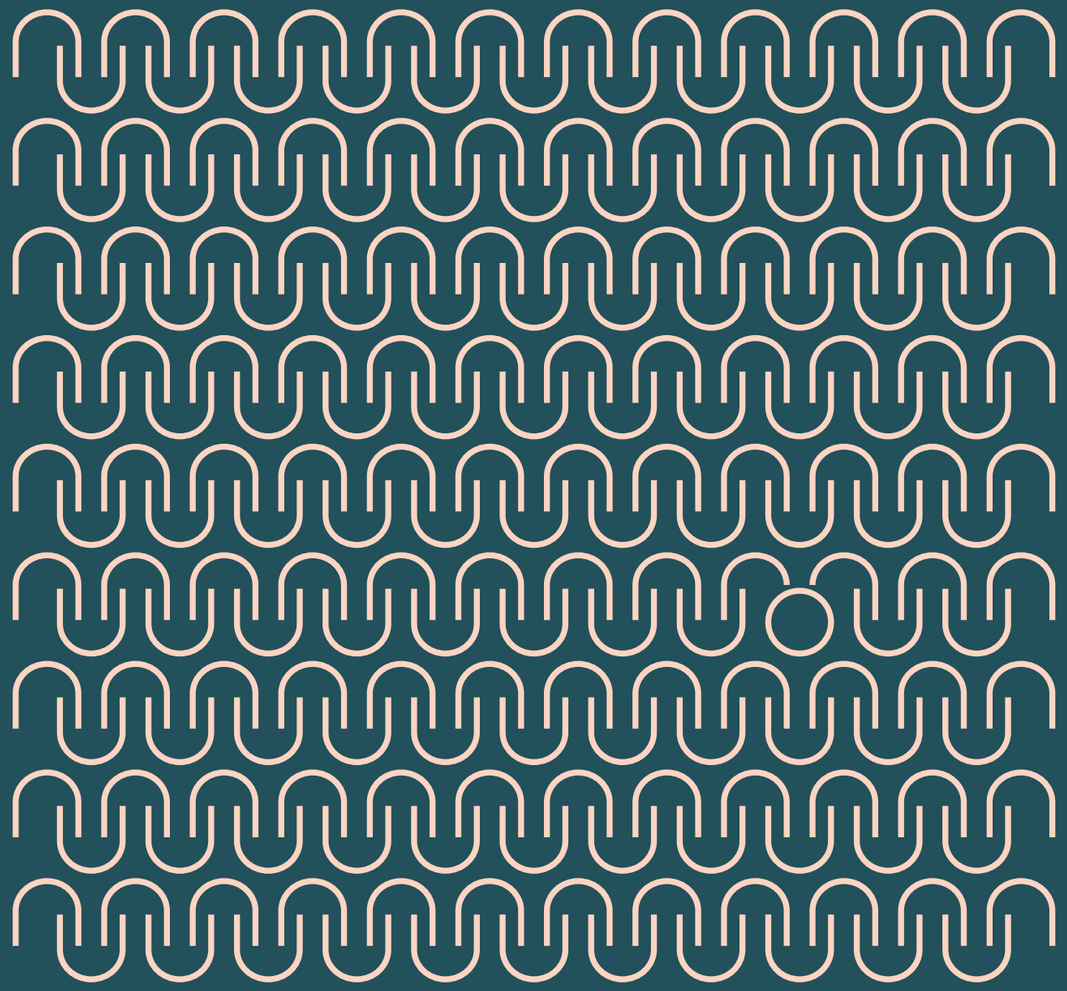
More comprehensive modelling in sandbox scenarios may have helped designers identify any biases or blind spots within their design.

[23] Ellen Broad, Made By Humans: The AI Condition, (Melbourne: Melbourne University Press, 2018)

[24] Julia Dressel, Hany Farid, 'The Accuracy, Fiirness and Limits of Predicting Recidivism', Science Advances, Vol 4(1), 2018

PR06.

ACCESSIBILITY



Design to include the most vulnerable user.

Whenever you have intended users and use cases, you also have people who you don't intend as users of the technology. This is a risk when design excludes people who might benefit from your design, if you'd only thought of them in the process. Design can reinforce social disadvantage, or it can help people overcome it. But it can only do this if we bear in mind all the possible users, without dismissing some groups as 'edge cases'.

We start with the assumption that each person has an equal right to access any piece of technology.

Of course, these rights are not absolute. Intellectual property rights, risk of abuse or any number of considerations could trump a person's right to access technology. However, the design choices of technology makers are not among them.

Technology must be designed such that any person with the right and need to use it may do so without unreasonable difficulty. If not, certain people or groups may be excluded, without good reason, from something that might benefit them.

The term 'edge case' is sometimes used to describe these kinds of people – those who used the technology but are not part of the target audience. Technologists should stop thinking about edge cases. They should not assume they have total say over who their audience is. It isn't enough to say "we didn't design this technology for X group". If X group has a presumed right to the technology, they ought not to be excluded by design.

As technology designer Mike Monteiro writes:

"For years we referred to people who weren't crucial to our products' success as "edge cases". We were marginalizing people. And we were making a decision that there were people in the world whose problems weren't worth solving."[25]

[25] Mike Monteiro, 'A Designer's Code of Ethics', Medium, Accessed: https://deardesignstudent.com/a-designers-code-of-ethics-f4a88aca9e95

Rules of Thumb

01. NO EDGE CASES

Design your technology with an eye to who would be likely to use it, not who you would want to have using it.

02. CLOSE THE GAP

If there are necessary and inescapable differences in accessibility, active steps must be taken to close the accessibility gap, as much as is reasonably possible.

03. KNOWLEDGEABLE USE

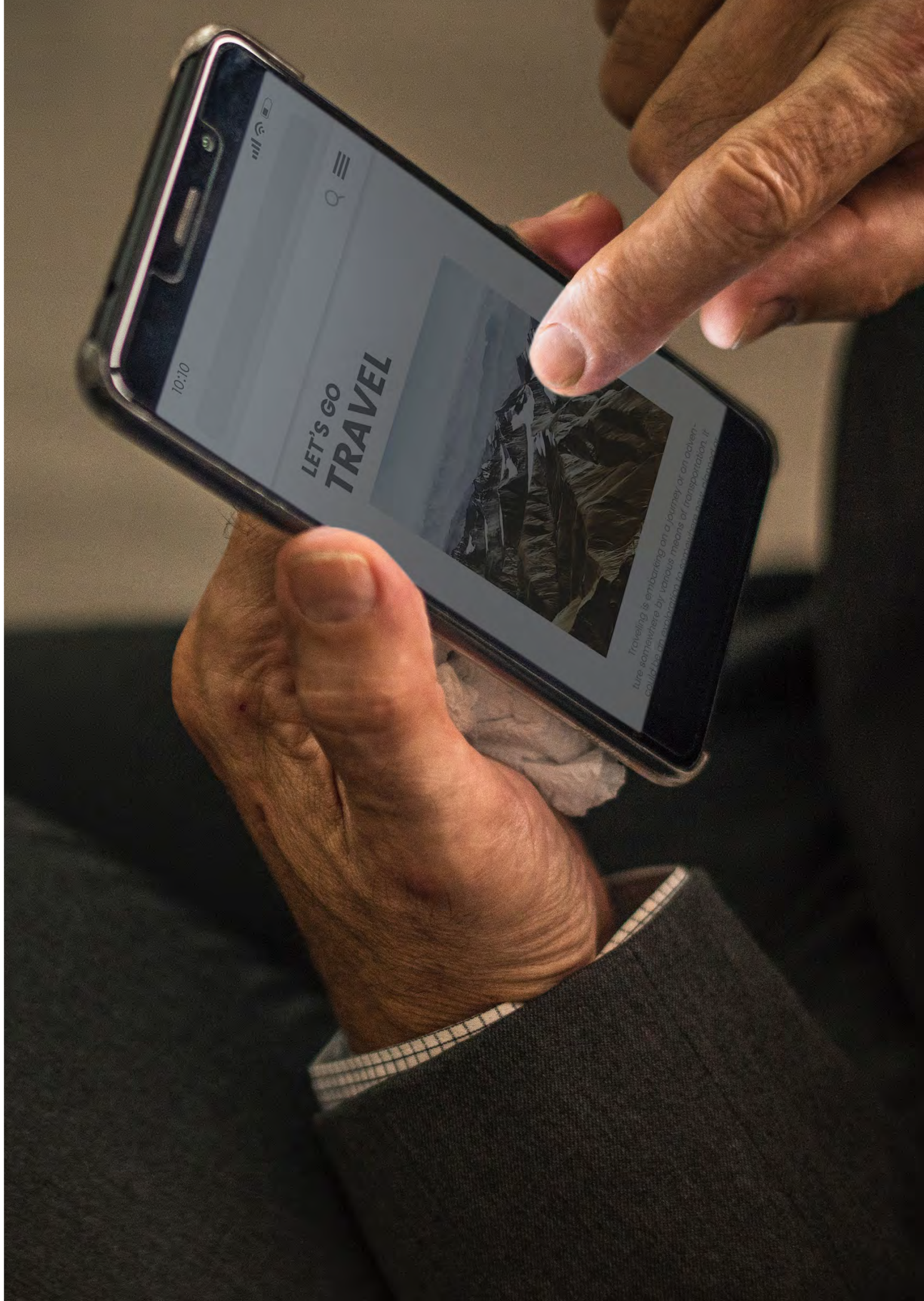
Access does not simply refer to the ability to pick up technology and use it. It also includes a person's ability to be informed about what they are using. Users should be able to access all the relevant information regarding trade-offs, compromises, implications for their civil liberties, rights and so on.

04. SHOW YOUR WORK

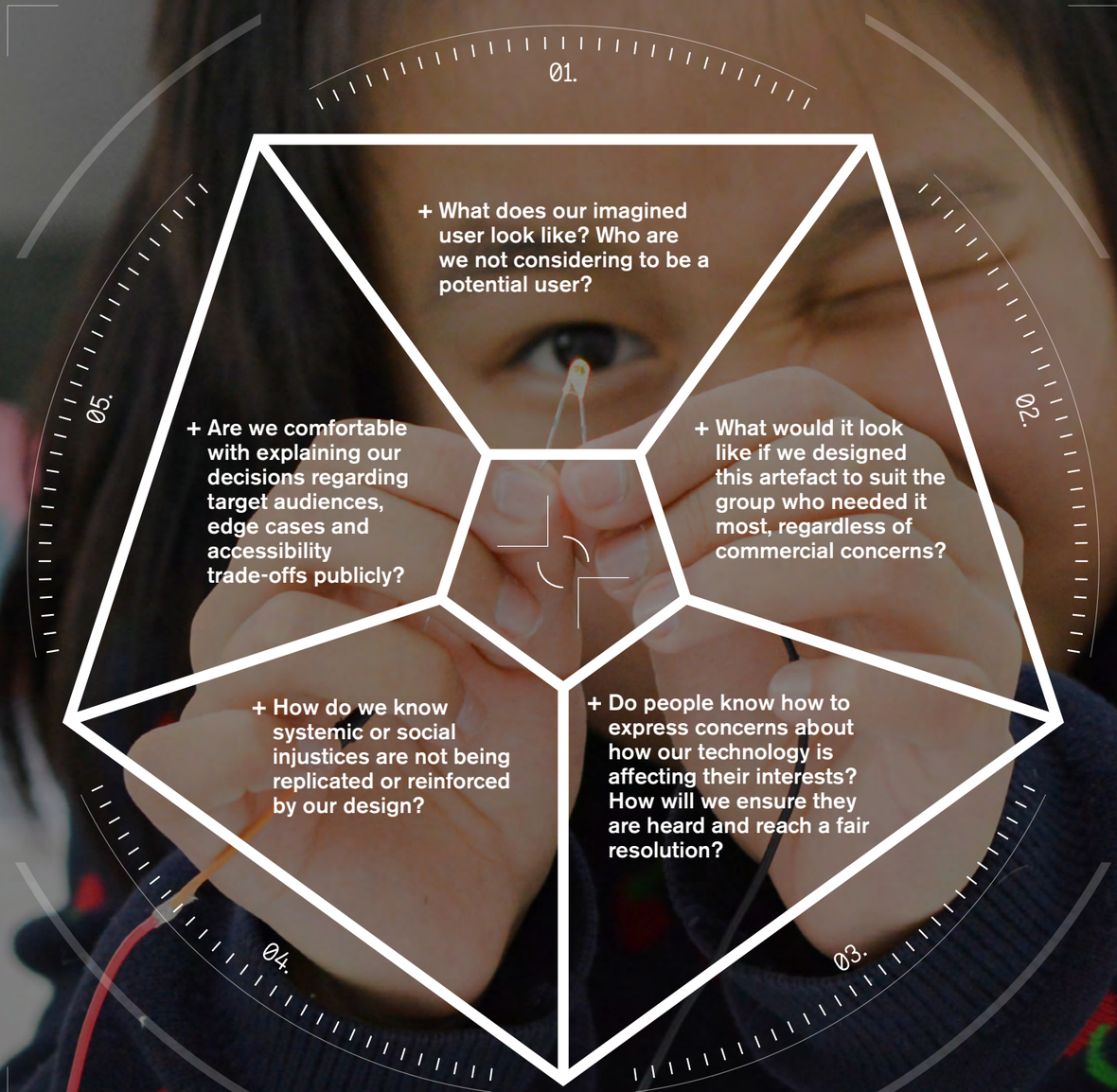
The ethical rationale for any piece of technology should be publicly available in an easily accessible form. This statement should explain how the technology has satisfied the principles outlined in this paper, what good the technology is going to provide the world, what the intended purpose of the technology is, why any potential harms are justified and what steps have been taken to mitigate them (as much as is possible without providing a guide to people wanting to subvert safeguards or cause mischief). In cases of classified technology, this should be available to an authorised, independent body.

05. SUCCEED SLOW

Technology is often released to marketing as a minimum viable product, with an eye to addressing bugs and issues as they arise. This mentality, and the related 'fail fast' philosophy of many products today, needs to be challenged. Rigour in testing, design and user consultation helps prevent issues of accessibility, preventing exclusive design decisions.



Ethical Design Challenges



Ethical Design Solutions

01. DIVERSE BETA TESTING

To identify any unintended access issues, new technologies must be tested across a diverse range of potential user groups. Processes that ask people to volunteer as beta testers may be insufficient to achieve this diversity – it must be explicitly sought to ensure groups who may previously have been relegated as 'edge cases' are not excluded by the very design of the technology.

02. USAGE OPTIONS

Inclusive design[26] is advanced by providing various options for how an artefact is used – including physical locations for use, input methods and so on.

[26] A series of other principles and solutions can be found in: The Center for Universal Design, 'The Principles of Universal Design v. 2.0' (Raleigh, NC: North Carolina State University, 1997).

Ethical Design Challenges

Ethical Design Solutions

Case Study: Carmat Artificial Heart

In 2014, French company Carmat successfully implanted an artificial heart into the chest of a 76 year-old man. Whilst he subsequently died, prematurely, due to a short circuit in the device, it heralded a revolution in healthcare. There would be no more need to find donors in order to conduct heart implantations.

However, due to its size, the Carmat heart is compatible with the bodies of only 86% of men and only 20% of women. The thoracic cavity needs to be a particular size to fit the heart, and because men tend to be physically larger than women, they are more likely to accommodate the heart at its current size. According to comments provided to *Motherboard*, in 2014, Carmat are not pursuing research into a smaller-sized artificial heart.[27]

01. CLOSE THE GAP

To satisfy the 'close the gap' principle, Carmat could continue to invest in research into smaller, more equitable designs for artificial hearts.

02. REDUCE THE SIZE OF THE EDGE CASES

A smaller-sized heart might have been stipulated at an earlier stage in the design brief; one that included a broader range of body-sizes and reduced the size of the 'edge cases'.

03. SHOW YOUR WORK

Carmat's discussions regarding the cost and practical limitations of their artificial heart have been a candid example of 'showing their work'.

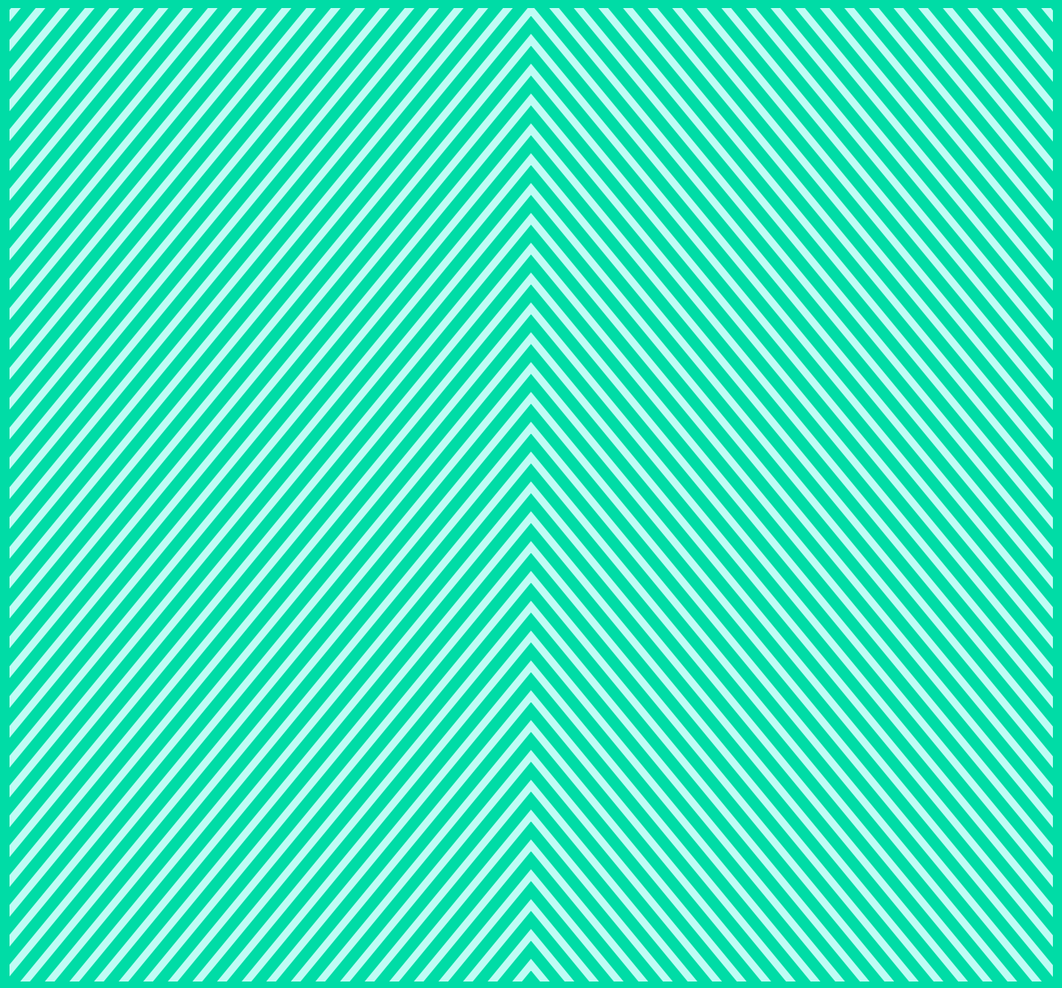
04. CONTINUE TO EXPLORE

A standing commitment to continue to explore this technology, when new developments permit it, whilst continually updating interested, excluded groups about developments would demonstrate a clear commitment to closing the gap.

[27] Victoria Turk, 'Technology isn't designed to fit women', Motherboard, Sept 13, 2014, https://motherboard.vice.com/en_us/article/mgb3yn/technology-isnt-designed-to-fit-women

PR07.

PURPOSE



Design with honesty, clarity and fitness of purpose.

Design is, in one sense, a promise. You are promising to solve a problem your users are experiencing. Like all promises, you should honour this promise. You must be honest and clear about the ability and limitations of your design. Moreover, your design should be tailored to the problem it's trying to solve – and be intended to solve a genuine problem. Good design serves an ethical purpose and does so in efficient and effective ways.

Given the importance of human agency and that one of the functions of technology is to assist and enhance the choices available to us and our ability to succeed in our choices, the effectiveness of technology takes on ethical importance. Technology that doesn't do its job invalidates the agency of the user the same way a failure to read someone's vote in a democracy undermines the voter's importance as a citizen.

As a result, technology must be suitable to achieve the goals it has been designed to achieve.

We have elsewhere discussed the ethical challenges that arise from the repurposing of technology. The following considerations don't address these directly. Here, we focus on the intended purposes of technology, from the perspective of designers.

Rules of Thumb

01. LEGITIMATE PURPOSE

Technology should be directed toward the common good of humanity. Technology that aims to achieve evil – or even aims at neutrality, offering nothing to the common good – cannot be justifiable, no matter how responsibly it has been designed or how much it facilitates human freedom.

02. CLARITY OF PURPOSE

Every piece of technology should be designed with a clear awareness of what it is for, so as to make users aware of the intended purpose.

03. HONESTY OF PURPOSE

Be honest about the capabilities and limits of an artefact. Don't oversell the scope of the technology in terms of what it can do, the problems it could solve or how distinctive it is from other products on the market.

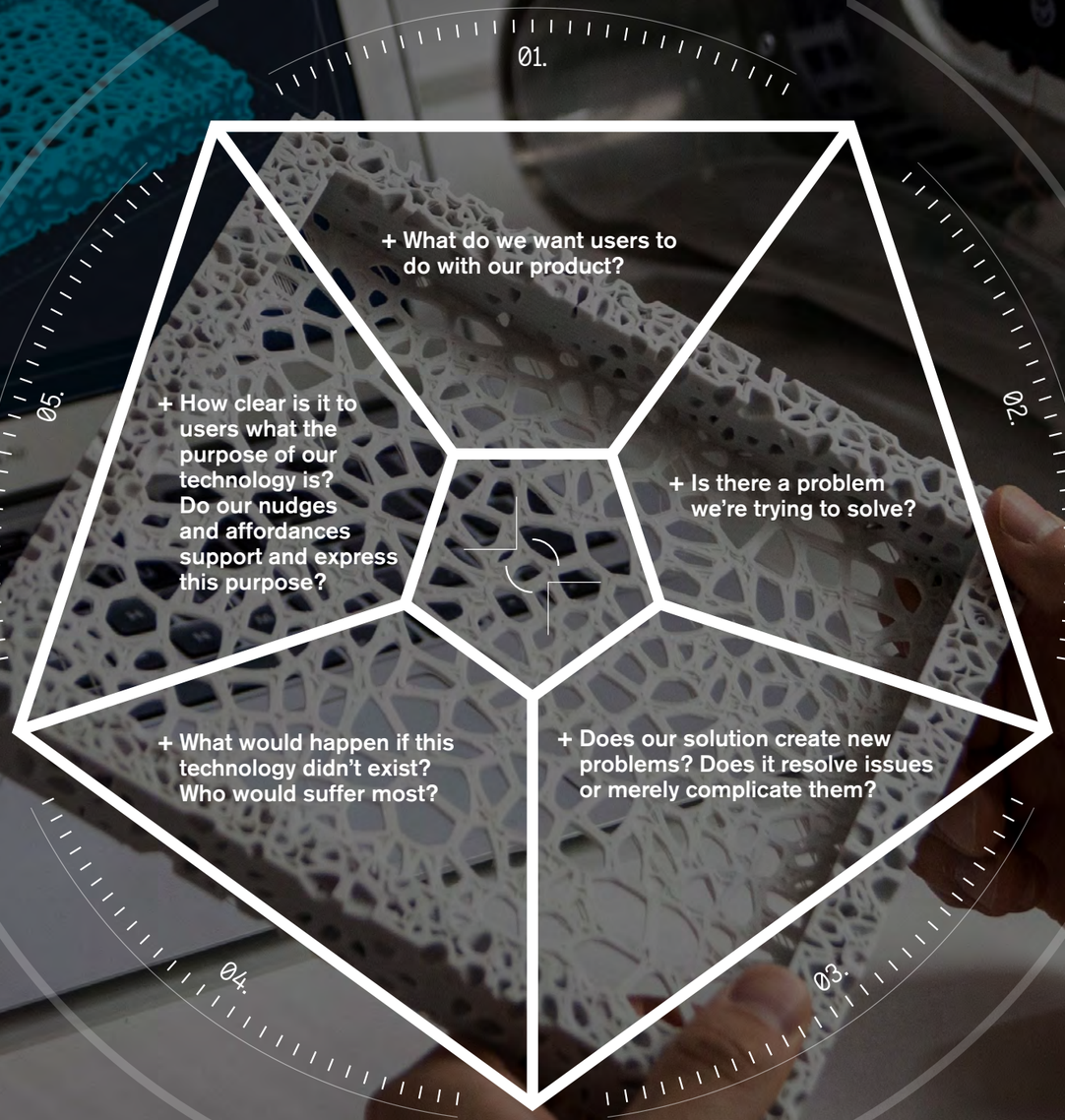
04. PRINCIPLED EFFECTIVENESS

Technology should be as effective as possible in achieving its purpose without undermining any of the other ethical principles outlined in this paper. If it is impossible to achieve a reasonable level of effectiveness without violating any other principle, the technology should not be made (with the exception of test models to try to enhance the effectiveness).

05. PRINCIPLED EFFICIENCY

To best aid and enhance human agency, technology should achieve its purpose as efficiently as possible within the constraints of the other means. The way technology achieves its goal must be both good and efficient. If it isn't possible to create technology that does the job efficiently and ethically, it shouldn't be made (with the exception of test models to try to enhance the efficiency).

Ethical Design Challenges



Ethical Design Solutions

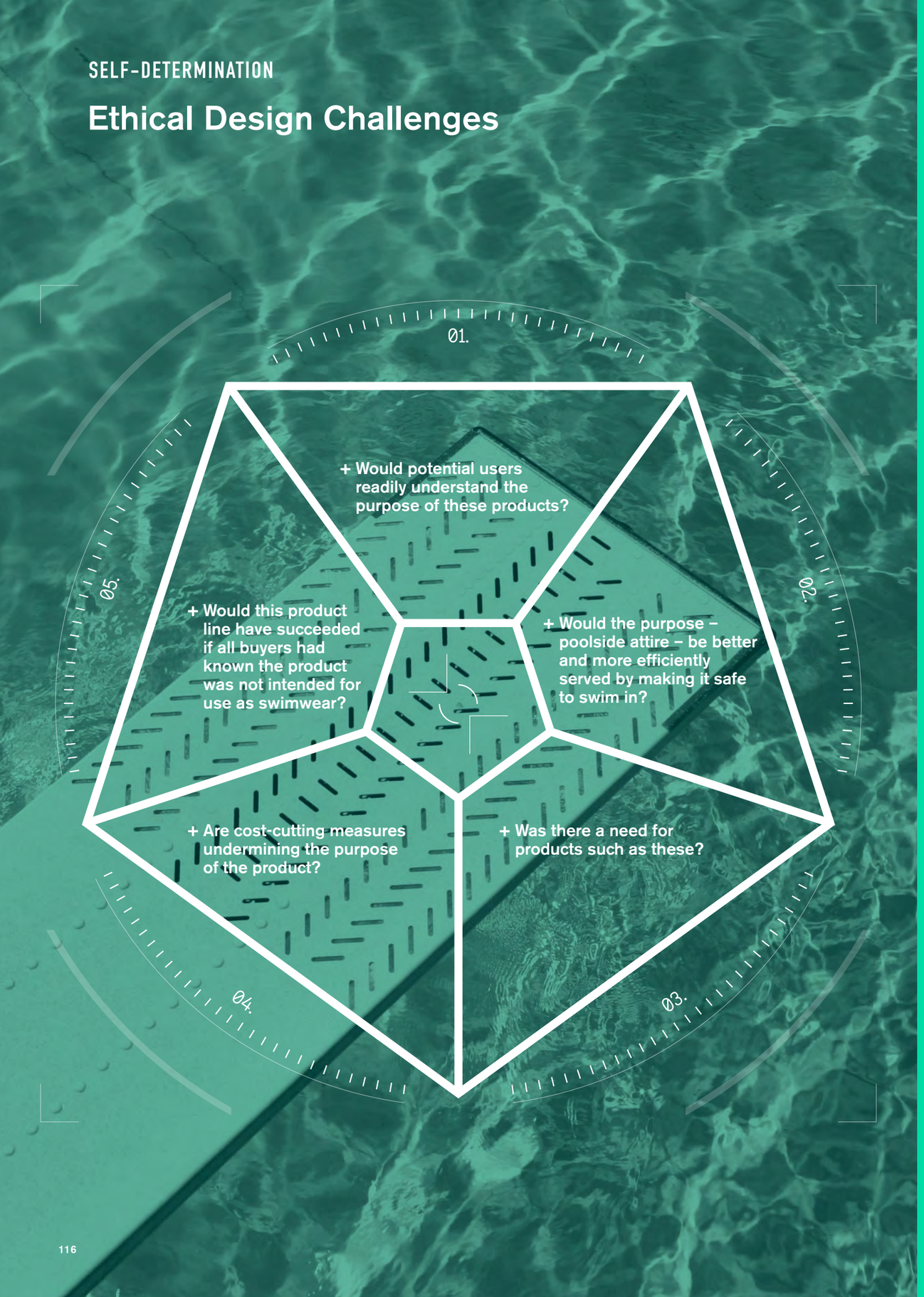
01. USER EXPECTATION TESTING

Given the risk in miscommunicating the intended purpose, scope and application of technology, companies should test their marketing and communications material to ensure users can and do understand the scope of intended use – what the product is (and isn't) for.

02. MINIMALIST DESIGN

Unnecessary, extraneous features can both distract from purpose and encourage greater repurposing – and the accompanying ethical risks. Ensure each function and feature of the artefact is purpose-driven, and avoid trying to serve too many purposes in a single design.

Ethical Design Challenges



Ethical Design Solutions

Case Study: Non-swimmable Swimwear

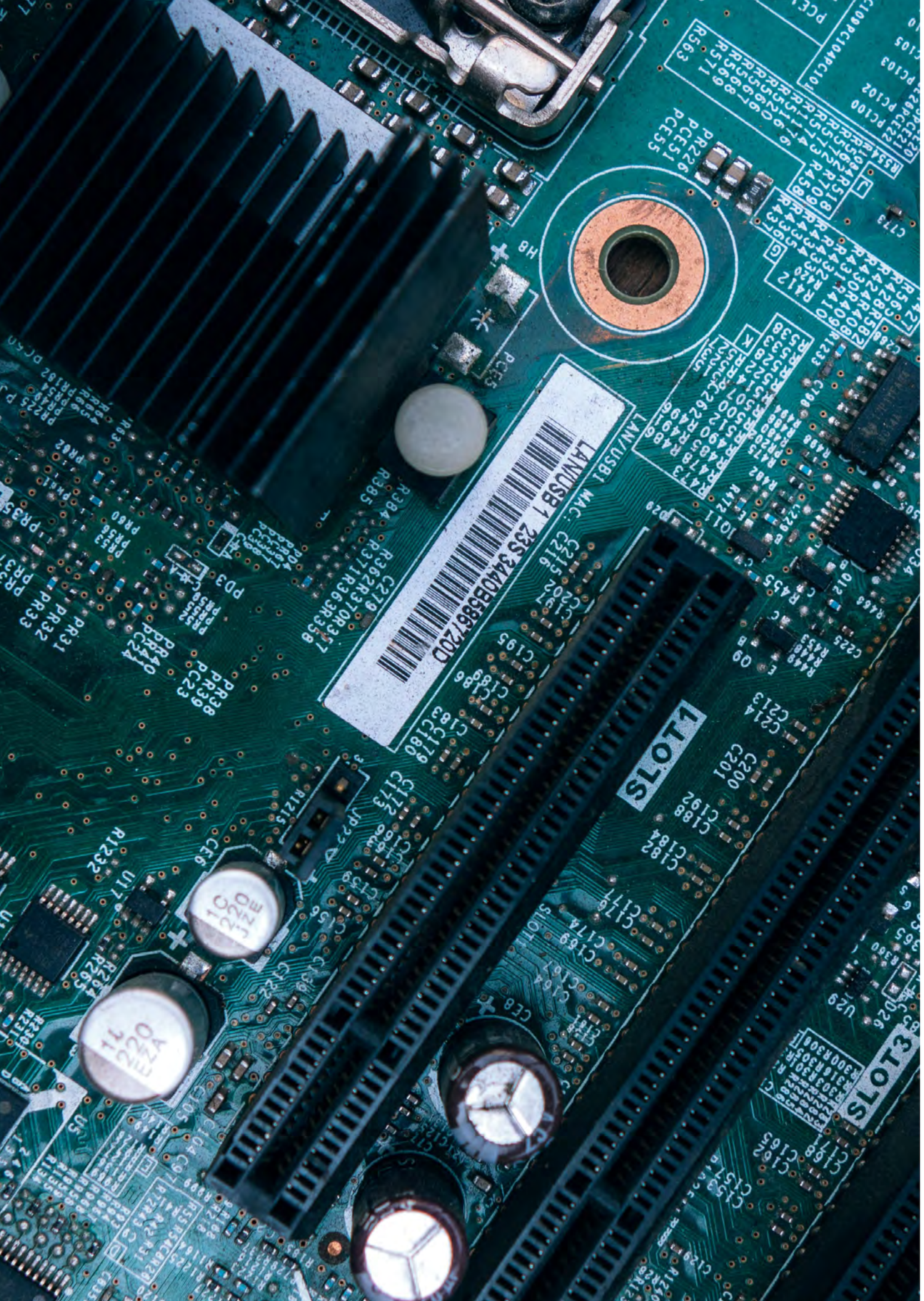
In 2018, a range of online fashion shoppers were left bemused by a series of swimwear options which were labelled with warnings such as, “not to be worn in water” or “may become transparent when wet”. Some fashion labels argued that their pieces were in fact ‘poolside attire’ rather than swimwear.

01. DESIGN FOR DUAL USE

Poolside attire could be designed for ‘dual use’ as both fashion and swimwear without detracting from purpose.

02. SPECIFY INTENTION OF USE

Clear advertisements specifying the intended use – and potentially embarrassing results of misuse – should have been prominently visible on the packaging to prevent misunderstandings.

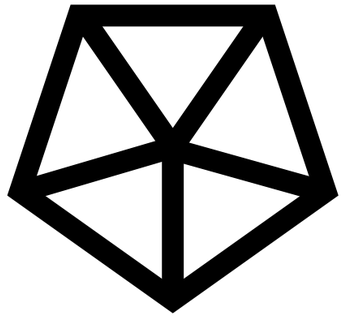


OTHER ISSUES

We believe the ethical framework outlined in this paper will go a considerable way to addressing a range of current and future issues in technology, but it isn't a silver bullet.

Simply outlining the ethical requirements for technological design will no more 'fix' technology ethics than similar approaches have in any other field. Medical ethics has clear guidelines, frameworks and case studies, yet malpractice, injustice and exploitation still occur. This suggests two related projects which require addressing in order to buttress the findings in this framework.

Ethics Education in Technology Design



Alongside a clear framework of values, principles and purpose, ethical conduct relies on people who care deeply and personally for these things.

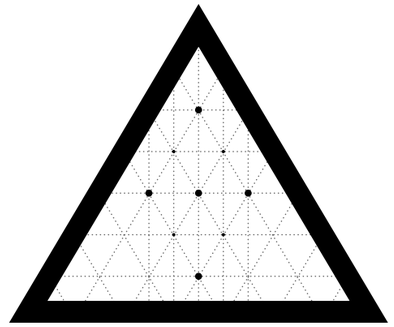
That is, alongside having the right ethical framework for technological design, we need the right education programs to ensure the framework is both understood and treated with the right level of seriousness.

To see the significance of this, it's helpful to consider an example from another profession:

A young enlisted marine in the Vietnam War's judgement concerning the distinction between combatants and non-combatants was compromised after he'd seen too many of his buddies 'blown away'. An officer found the youth with his rifle pointed at the head of a Vietnamese woman. The officer could have tried barking out the relevant provisions of military law. Instead, he just said "Marines don't do that." Jarred out of his berserk state and recalled to his place in a long-standing warrior tradition, the marine stepped back and lowered his weapon.

Without the right kinds of education and formation – leading those working in technology to see themselves as being a particular kind of person with particular ethical commitments – any framework (including the one outlined in this paper) is likely to be seen as just another document, policy or rule to be dismissed whenever it's convenient to do so. Technology design needs to be recognised as a form of ethical practice. This practice needs to be underpinned by dispositions of character that guarantee the integrity of the principled approach – even (or especially) in moments when regulation, management or oversight cannot bind behaviour.

Ethics & Technology Companies



Not all the issues facing technology today stem from technology itself.

Stories of workplace bullying, wage exploitation and tax avoidance are issues for technology companies, but they don't stem from the fact these companies deal in technology. They're issues for business ethics more broadly.

This is important because unless we recognise the broader ethical issues at play – matters of social justice, economics, governance and so on – technology companies will still face issues in securing legitimacy and trust, and their activities will be, at least to some extent, at odds with the purpose, values and principles outlined here. Can a technology company that is not paying a fair share of tax be said to be taking responsibility for itself? If staff are being underpaid or exploited, how can the principle of non-instrumentalism be said to apply?

For technology companies to get the ethics right, they need to get the ethics of technology right, but they also need to make sure they have the correct sense of what it means to be an ethical company.

FURTHER PROJECTS

This project is the first stage of a much longer endeavour to support ethical design in technology.

We envision this work to serve as the foundation for a range of other activities, including:

+ **INDUSTRY ENGAGEMENT, MEETUPS, INFORMAL TRAINING AND THE DEVELOPMENT OF COMMUNITIES OF INQUIRY AND HACKATHONS.**
+ **SPEAKING, CONSULTING AND ADVOCACY.**
+ **A SOPHISTICATED ETHICAL DESIGN RESOURCE, WHICH COULD BE UTILISED WITHIN DESIGN TEAMS AND ORGANISATIONS TO ENSURE THE ISSUES AND CHALLENGES WE'VE IDENTIFIED HERE ARE ADDRESSED.**
+ **TARGETED PROJECTS AIMED AT DEVELOPING SOME OF THE FEATURES IDENTIFIED HERE – FOR INSTANCE, HARM METRICS, ACCESSIBLE EULAS AND TRANSPARENT PROVENANCE.**
+ **SPECIFIC ETHICAL RECOMMENDATIONS CONCERNING TECHNOLOGIES OF MAJOR COMMUNITY INTEREST AND CONCERN – SUCH AS CRASH SCENARIOS FOR AUTONOMOUS VEHICLES.**

Beard, M and Longstaff, S A (2018) *Ethical Principles for Technology* The Ethics Centre, Sydney

THE ETHICS CENTRE
Level 2 Legion House, 161 Castlereagh Street, Sydney NSW 2000
+61 2 8267 5700 contactus@ethics.org.au

WWW.ETHICS.ORG.AU