Conceptual Disruption and the Ethics of Technology

Hopster, J.; Brey, P.; Klenk, M.B.O.T.; Löhr, G.; Marchiori, S.; Lundgren, B.; Scharp, K.

DOI
10.11647/OBP.0366.06

Publication date
2023

Document Version
Final published version

Published in
Ethics of Socially Disruptive Technologies

Citation (APA)
Hopster, J., Brey, P., Klenk, M. B. O. T., Löhr, G., Marchiori, S., Lundgren, B., & Scharp, K. (2023). Conceptual Disruption and the Ethics of Technology. In L. van de Poel, L. Frank, J. Hermann, J. Hopster, D. Lenzi, S. Nyholm, B. Taebi, & E. Ziliotti (Eds.), *Ethics of Socially Disruptive Technologies: An Introduction* (pp. 141-162). Open Book Publishers. https://doi.org/10.11647/OBP.0366.06

Ethics of *Socially Disruptive Technologies*

An Introduction

Edited by
Ibo van de Poel,
Lily Frank, Julia Hermann,
Jeroen Hopster, Dominic Lenzi,
Sven Nyholm, Behnam Taebi, and
Elena Ziliotti

https://www.openbookpublishers.com/

ISBN Paperback: 978-1-80511-016-3
ISBN Hardback: 978-1-80511-017-0
ISBN Digital (PDF): 978-1-80511-057-6
ISBN Digital ebook (EPUB): 978-1-78374-789-4
ISBN Digital ebook (XML): 978-1-80511-050-7
ISBN Digital ebook (HTML): 978-1-80064-987-3
DOI: 10.11647/OBP.0366

Cover image: Blue Bright Lights, Pixabay, 6th April 2017, https://www.pexels.com/photo/blue-bright-lights-373543/
Cover design: Jeevanjot Kaur Nagpal

6. Conceptual Disruption and the Ethics of Technology

Lead author: Jeroen Hopster[1]
Contributing authors: *Philip Brey, Michael Klenk, Guido Löhr, Samuela Marchiori, Björn Lundgren, Kevin Scharp*

This chapter provides a theoretical lens on conceptual disruption. It offers a typology of conceptual disruption, discusses its relation to conceptual engineering, and sketches a programmatic view of the implications of conceptual disruption for the ethics of technology. We begin by distinguishing between three different types of conceptual disruptions: conceptual gaps, conceptual overlaps, and conceptual misalignments. Subsequently, we distinguish between different mechanisms of conceptual disruption and two modes of conceptual change. We point out that disruptions may be induced by technology but can also be triggered by intercultural exchanges. Conceptual disruptions frequently yield conceptual uncertainty and may call for conceptual and ethical inquiry. We argue that a useful approach to addressing conceptual disruptions is to engage in conceptual engineering. We outline what conceptual engineering involves and argue that discussions on conceptual disruption and conceptual engineering can benefit

1 All mentioned authors contributed in some way to this chapter and approved the final version. JH is the lead author of this chapter. He coordinated the contributions to this chapter, outlined its structure, and did the final editing. SM wrote a first version of Section 6.1. GL wrote a first version of Section 6.2 and Section 6.3. KS wrote a first version of Section 6.4. PB wrote a first version of Section 6.5. MK and BL commented on the chapter and modified it.

 https://doi.org/10.11647/OBP.0366.06

from closer integration. In closing, we discuss the relevance of studying conceptual disruption for the field of technology ethics, and point to the promise of this line of research to innovate practical philosophy at large.

Fig. 6.1 Conceptual engineering. Credit: Menah Wellen

6.1 Introduction

The aim of this final chapter is to provide a theoretical lens on a core theme of the preceding chapters: conceptual disruption and the need for conceptual change. What kinds of conceptual disruptions can be distinguished? How can philosophers and ethicists address them? And what is the relevance of studying conceptual disruption for ethical theory and for practical philosophy at large? We start by recounting some of the conceptual disruptions that have been discussed in the previous chapters and offer further leads to theorize about them. Next, we point to some of the different causal triggers of conceptual disruption, which include not only technologies, but also intercultural dialogue. Thereafter, we introduce a recent philosophical approach that can help in addressing conceptual disruptions: conceptual engineering. We conclude by discussing the relevance of conceptual disruption

for technology ethics and by stating its promise as a future research program that can benefit practical philosophy at large.

Before we get into the topic of this chapter, we should emphasize that conceptual disruption is not the only aspect of socially disruptive technologies worthy of ethicists' attention. So are 'social disruptions' more generally. Many of the examples that have been discussed throughout the previous chapters concern 'social disruptions' (Hopster, 2021a), i.e. social dynamics, often fostered by emerging technologies, whereby important aspects of human society are prevented from continuing as before, provoking normative disorientation, and giving rise to a variety of ethical and social challenges. Social disruptions may also involve the disruption of concepts, but social disruptions are not limited to conceptual disruptions. Social change occurs at many levels, and the conceptual level may not always be the most salient or interesting one.[2]

Yet, one reason for focusing specifically on conceptual disruption in this chapter is that, up until recently, this has been a relatively neglected topic of inquiry in the ethics of technology, and more so than the topic of social disruption. It is an explicit ambition of the ESDiT program to put conceptual disruption on the map of academic scholarship, and this concluding chapter provides several leads to develop that ambition. But in stating this emphasis, we do not wish to downplay the importance of social disruption as a distinct and relevant topic of ethical inquiry. Spelling out the precise nature of social disruptions, as well as its ethical implications (e.g. Hopster, 2021b; O'Neill, 2022), remains a core focus within ESDiT. This chapter simply has a different focus.

6.2 Types of conceptual disruptions

In the introduction, we defined conceptual disruption as a challenge to the meaning of a concept that prompts a possible future revision of it. This challenge may pertain to individual concepts, and also to our

2 Moreover, not all conceptual disruptions are entangled with social disruptions. There can be instances where concepts are challenged, but where this challenge does not emerge from societal disruptions (e.g. the introduction of the Archeae as an independent biological kingdom could be regarded as a significant conceptual disruption, but with no societal implications). Social and conceptual disruption are related, but distinct.

conceptual scheme as a whole. We argued that conceptual disruptions can be interpreted in three ways, and that technology typically plays a prominent role in each of them (Hopster and Löhr, 2023). First, we may be faced with a 'conceptual gap'. That is, we lack the concepts needed to describe a novel technological artifact, or to normatively evaluate the new impacts and affordances to which it gives rise. Second, we may be faced with a 'conceptual overlap'. That is, more than one of our existing concepts may be appropriate to describe and evaluate a novel technology, but there is uncertainty as to which concept is most suitable. Third, there may be cases of 'conceptual misalignment'. In such cases, existing concepts *do* seem applicable to conceptualize a new technology and its impacts and affordances. However, this apparent good fit actually masks an underlying value misalignment: the concept and its use do not express the values that a community of concept-users, upon ethical reflection, would like it to express.

When thinking about conceptual disruption in these terms, one should be sensitive to various problems that lurk in the background. First, for any given example, the most appropriate framing may itself be contested: what may be regarded as a conceptual gap by some ('the problem is that we lack an appropriate concept of X!'), may be understood as an overlap by others ('no, the real problem is that we have conflicting concepts of X!'). Second, talk of 'conceptual disruption' may suggest a somewhat reified and monistic understanding of concepts: it may leave the impression that the meaning and extension of concepts is always clearcut, and that there is one dominant concept or conception that gets disrupted. Such an understanding might assume more conceptual agreement than actually exists and ignore a great deal of disagreement and diversity. Furthermore, one might worry that cases of conceptual overdetermination can sometimes be unproblematic and may even be fruitful: conceptual overlaps create room for a plurality of plausible interpretation that can be tailored to specific contexts or domains. We think these are legitimate concerns. Endorsing a reified and monistic understanding of conceptual disruption constitutes a pitfall that a plausible account should expressly avoid.

Yet, when keeping these pitfalls in mind, conceptual disruption can be a fruitful concept for philosophical inquiry, and it is helpfully understood in terms of gaps, overlaps, and misalignments. Can we identify instances

of these three types of conceptual disruption in the examples discussed in the preceding chapters? Consider the examples of social robotics, outlined in Chapter 3. We observed that social robots are blurring the line between 'alive' and 'lifelike': we intuitively perceive social robots as being alive in some sense, although we are aware that they are not (Carpinella et al., 2017; Spatola and Chaminade, 2022). This could be interpreted in terms of a conceptual overlap: both sides of dichotomous concepts like 'alive' and 'lifeless', or 'animate' and 'inanimate', seem applicable to social robots. At the same time, progress in social robotics arguably gives rise to conceptual gaps. Consider the binary distinction between 'moral agency' and 'moral patiency'. Arguably, there is reason to ascribe some positive moral status to intelligent machines. But neither paradigm examples of entities having moral agent-status (reflective humans), nor examples of entities having moral patient-status (sentient animals), provide a solid model for such ascriptions. Perhaps we need to articulate different notions of moral status such as the 'relational status' advocated by Gunkel (2018) and Coeckelbergh (2010), or distinguish between different gradations of moral status. Differently put: arguably the emergence of social robots points to a conceptual gap which calls for making changes to our existing conceptual framework, with regard to a cluster of important moral concepts (moral status; moral agency; moral patiency; associated notions of responsibility; *etc*).[3]

The third type of conceptual disruption we identified are conceptual misalignments. Does social robotics also give rise to this type of conceptual disruption? Recall that social robots often take on a humanlike form, which may come with certain advantages, but also engenders certain risks. Authors who emphasize the downsides of humanoid robots have argued, for instance, that anthropomorphizing robots encourages unwanted disruptions to our moral system, which, in turn, 'could seriously disrupt our ability to govern, as well as our economy' (Bryson, 2018: 22). One way to interpret this worry is in terms

3 This example could also be understood as a case of 'moral disruption' (Baker, 2013; Nickel, 2020; Rueda et al., 2022), since the concepts at issue — moral agency and patiency — are quintessential concepts of moral theorizing. To the extent that conceptual disruptions call for rethinking these foundational moral concepts, or their precise extension, conceptual disruptions have direct implications for ethical theory and moral practice, and conceptual disruptions are entangled with moral disruptions.

of conceptual misalignment. If we were to extend our concept of moral patiency to social robots, Bryson worries, then we are conceptualizing a novel technology using a concept that *prima facie* appears to have a good fit, but that upon reflection actually involves a misfit with other concepts and values. Our concept of 'moral patient' may be naturally extended to robots, but on further consideration, the moral implications of doing so are contentious. Assuming this view, then, we should conceptualize novel technologies in such a way that conceptual misalignments do not transpire.

We discussed another example of conceptual misalignment in chapter two, when considering the relation between the concepts of 'demos', 'democracy', and 'public sphere'. By giving citizens and non-citizens equal substantive access to online political debates that shape the political agenda, social media has severed the conceptual relationship between the 'demos' and 'public sphere', giving rise to a conceptual misalignment. Note that such a misalignment may only become apparent upon ethical reflection. This is what sets conceptual misalignments apart from conceptual gaps and overlaps: in the case of misalignment, our conceptual scheme continues to function fluently, yet in a way that is ethically problematic, as the functioning does not reflect how concepts *should* be aligned.

Conceptual gaps, overlaps and misalignments are useful terms for studying conceptual disruption. But not all the case studies discussed in this book straightforwardly adhere to this tripartite distinction. As noted, conceptual changes are not always disruptive, in the sense that they do not always overturn a well-articulated conceptual status quo. Conceptual changes can also (and often do) occur in contexts of uncertainty, where norms of conceptual application are contested, or vague. Robots challenge our understanding of the human, but what the concept of 'a human' and 'human nature' (Hannon and Lewens, 2018) amounts to has itself been contested throughout intellectual history. These are 'essentially contested concepts' (Gallie, 1955), which are continuously being disputed when it comes to their interpretation. Novel technologies can give powerful impetus to rethinking and conceptualizing them anew, but in doing so they do not always disrupt a clearly established conceptual status quo. Consider the distinction between what is internal and external to the human body, which is

challenged and blurred by various technological artifacts, such as ventricular assists, or the artificial womb (Chapter 5). Arguably, the boundary between what is external and internal to the body (and the associated social and moral norms) was never clearly established. New disruptive technologies, however, prompt us to clarify this distinction, such that new norms can be established (e.g. is shutting off an artificial womb by a third party a violation of bodily integrity?). Here, conceptual disruption involves a call for clarification and conceptual articulation, where conceptual frameworks were previously vague or indeterminate.

We noted that conceptual specification should not always be regarded as a *desideratum*. The ambiguity and indeterminacy of conceptual frameworks may serve some functions, for instance as it allows for flexibility and context-specificity. Yet there are conditions in which the articulation of clear conceptual norms is called for. Consider the challenges of ascribing responsibility in the context of climate change (Chapter 4). For global geo-engineering technologies, we observed that ascribing responsibility is a very daunting task. Yet at the same time, articulating applicable notions of responsibility also seems of the utmost importance, given the major significance of the challenge at hand and the need to adequately respond to it (e.g. Jamieson, 2015). According to some, we need an adequate concept of responsibility to maintain a moral community; according to others, we need the concept of responsibility to uphold moral agency, or to help to steer actions in a desirable way.

Revising the concept of 'responsibility' in the face of new technological pressures might initially appear to be an isolated conceptual change. But as many of the examples we have discussed in this book showcase, conceptual disruptions are typically not limited to single concepts. Instead, they challenge clusters of interrelated concepts (Löhr, 2023). Consider the conceptual disruptions brought about by reproductive technologies as introduced in Chapter 5, which challenge our concepts of 'mother', 'father', and 'parent', as well as our concepts of 'birth', 'beginning of life', and 'personhood'. Similarly, social media challenge our concepts of 'demos', 'democratic public sphere', and 'self-rule' (Cf. Section 2.3), social robots challenge our concepts of 'agency' and 'moral patiency' (Cf. Section 1.4 and Section 3.3). The same holds for climate engineering technologies: these do not only challenge

our concept of 'responsibility', but also the associated notions of 'agency' and 'control' (Cf. Section 4.3).

Conceptual disruptions may give rise to conceptual changes, but this need not always be the case (Löhr, 2022). In fact, technological pressures may also give impetus to conceptual preservation (Lindauer, 2020). Consider once again the concept of 'democratic public sphere', addressed in Chapter 2. The advent of social media appears to have called into question the necessity and usefulness of referring to geopolitical factors to identify the democratic public sphere. In doing so, it has made it arduous to pinpoint exactly *where* such a public sphere exists. However, this conceptual disruption has not challenged the concept of 'democratic public sphere' as such. Instead, it has opened the door to multiple possible conceptions of such a democratic public sphere. As such, one may argue that the concept of 'democratic public sphere' would seem to be ultimately adequate and should be preserved.

When referring to conceptual disruption, it is important to bear in mind this distinction between concepts and conceptions. Rawls (1999) famously stated that the concept of 'justice' allows for various conceptions, i.e., specific interpretations of the concept, such as his own 'justice as fairness'. Accordingly, conceptions may be understood as different interpretations that give precision to a concept, which are often contested or in some sense indeterminate (cf., Veluwenkamp et al., 2022). Some of the cases of conceptual disruption we have discussed in this book are similarly best understood as cases where new conceptions are advanced and discussed.

Frequently, however, conceptual disruptions do call for changing *concepts*, or enriching our conceptual schemes. Building on the discussion of Chapter 5, one might think that once a pregnancy can occur in an artificial womb, we also need another notion for the removal of the foetus after forty weeks of development from the artificial womb. We currently call this 'giving birth', but this concept does not seem entirely appropriate — for one, it involves two living beings, whereas the removal of a baby from the artificial womb involves only one. It may benefit us to introduce a distinct concept for this type of 'event'.

We submit, as an interim conclusion, that the study of conceptual disruption will benefit from clear criteria as to what counts as a conceptual disruption. Building on the previous discussion, we propose

that the presence of a conceptual gap, overlap or misalignment can be taken as indicative of conceptual disruption. Challenges to specific *conceptions* may also qualify as conceptual disruptions, though it should be kept in mind that this is not the same as the disruption of *concepts*. We grant that, even when these criteria are further fleshed out, there might still be disagreement as to whether they apply in any particular case.

6.3 Mechanisms of conceptual disruption and modes of conceptual change

In the introduction we noted that disruptions involve both a 'disruptor', i.e. the disruption instigator, as well as an object of disruption. In this section we focus on the disruptors, i.e. the causal mechanisms of conceptual disruption. What are triggers of conceptual disruption and change?

As we have argued throughout this book, technology often constitutes such a disruptor, for instance when the introduction of technological artifacts provokes new norms or re-classifications. One should keep in mind, however, that 'technology' consists of more than artifacts alone. Technological artifacts are often embedded in more encompassing sociotechnical systems. Consequently, while technology frequently plays a substantial causal role in triggering disruptions, the arrow of causality may be difficult to discern. Consider the global climate engineering technologies discussed in Chapter 4 and the associated conceptual and social disruptions of the Anthropocene. In this case, what is the cause of disruption? Many of the key technologies at issue here — such as Solar Radiation Management — have not (yet) materialized. Yet at the same time, these technologies are entangled with visions about human control over the Earth's climate system.

Furthermore, not all conceptual disruptions are caused by technologies. Indeed, there is a variety of other causal mechanisms that can do so. Here we highlight one such mechanism, which is particularly potent as a trigger of conceptual disruption: intercultural dialogues and interactions. Consider the Ubuntu framework, mentioned in Chapter 4. Ubuntu has a notion of community that is much broader than traditional Western conceptualizations. This notion allows the inclusion of ancestors as well as future generations in the moral community (Kelbessa, 2015).

As Wiredu (1994), quoted in Chapter 4, remarks: '[I]n this moral scheme the rights of the unborn play such a cardinal role that any traditional African would be nonplussed by the debate in Western philosophy as to the existence of such rights'. Now, once Western philosophy comes into contact with this very different ontological point of view, and once Ubuntu philosophy comes into contact with the very different starting point in the West, their respective conceptual frameworks are challenged and require rethinking. This is an example of a conceptual disruption that might occur through intercultural dialogue or confrontation. Such exchanges can provide inspiration for conceptual amelioration: they broaden the horizon of conceptual possibility and allow for criticism of possible shortcomings of conceptual frameworks that are taken for granted.

Conceptual disruptions are relative to the conceptual framework that is being disrupted — and here, too, intercultural differences are highly relevant. Cultural contexts affect whether technologies are socially and conceptually disruptive. We noted that robotics technologies can disrupt the distinction between what is animate and inanimate, but in a community which endorses animist beliefs and ascribes agency to artifacts to begin with, this would not constitute much of a conceptual disruption. Or consider proposals to grant legal rights to natural entities such as forests or rivers, which disrupts traditional Western conceptions of legal personhood. Yet it does not seem very disruptive relative to the conceptual scheme of the Māori people, which is much more sensitive to the importance of protecting socio-environmental relationships (the Māori concept of rāhui, for instance, places temporary constraints on human activities to ensure immediate responses to threats of serious harms) (Watene, 2022). Hence, conceptual disruptions are relative to conceptual frameworks, which may in turn be culturally relative.

In closing this section, let us point out that apart from different mechanisms of conceptual disruption, we might also distinguish between different modes of *conceptual change*. Attempts to overcome a conceptual disruption (e.g. through conceptual engineering; see Section 6.4) may lead to conceptual change. This can happen, for instance, when technology produces novel entities (new artifacts and new consequences of the use of technology) that do not make a good fit with our conceptual scheme, so much so that adaptations seem in

order. These can be superficial changes, such as the introduction of new concepts to designate the new technology and some of its uses, components and consequences. But they can also be more profound changes, which challenge fundamental philosophical concepts, like those of agency, organism, or mind. This happens, for instance, when a new technology produces hybrids that do not seem to fit existing fundamental concepts, such as intelligent robots, synthetic organisms, and brain-computer interfaces. These are conceptual changes that occur in a 'loud way', prompted by a conceptual disruption.

However, conceptual changes need not always be prompted by conceptual disruptions. We might call such an instance of conceptual change that occurs without conceptual disruption 'silent conceptual change'. One way in which such silent change can happen is when a technology generates new application domains for concepts. For example, many moral and philosophical concepts are currently reapplied in a digital context, leading to concepts such as 'digital well-being', 'digital democracy' and 'cybersecurity'. Similarly, the rise of genetic technologies now enables concepts like 'genetic privacy' and 'genetic equality'. These new technological manifestations may extend or change the meaning of the original concept. But this does not give rise to conceptual gaps, overlaps, or misalignments, at least not obviously so.

6.4 Conceptual engineering

What can we do in response to different types of conceptual disruptions? One general approach, which has attracted lots of attention in recent philosophy, seems particularly relevant for the task at hand: conceptual engineering (Scharp, 2013; Eklund, 2015; Cappelen, 2018; Burgess and Plunkett, 2013a; 2013b; Cappelen and Plunkett, 2021).

Conceptual engineering can be understood as a branch of philosophy dedicated to investigating how best to improve our concepts and other 'representational devices'. 'Representational devices' can be understood, roughly, as more or less accurate mental images of what the world is like. The central question for conceptual engineers is whether, when, why, and how we ought to change our concepts (and other representational devices). For example, should we strive to use concepts that are as accurate as possible, that 'carve nature at its joints', so to speak? Or are

there perhaps criteria other than accurate representation that should guide us in choosing the concepts we use? Thus, conceptual engineering covers questions about how to assess existing concepts, how to create new ones, and how to implement new conceptual proposals in actual populations of concept-users. Philosophers who work in the field of conceptual engineering are also interested in the *ethics* of changing our conceptual repertoires. In short, the field covers the philosophically relevant issues in the process of intentionally changing the concepts we use to think, or the meanings we use to communicate.

Conceptual engineering is frequently contrasted with conceptual analysis: rather than unveiling the meaning of our concepts, the aim of the former is to *change* concepts, on the basis of moral, epistemic, or other considerations. Thus, *conceptual engineering* is strongly associated (and sometimes identified) with *conceptual ethics* — the study of which concepts we should choose — rather than the study of which concepts we have already chosen in our public language. Conceptual engineering, then, is a way of intentionally engineering or changing our conceptual repertoire.

Although intentionally changing concepts has been a prominent feature of Western philosophy and science for centuries, it has recently become a major area of philosophical inquiry itself (e.g. Haslanger, 2000; Scharp, 2013; Burgess et al., 2020; Cappellen, 2018). That is, conceptual engineering is an approach to doing philosophy as much as it is an area of philosophical inquiry. There are prominent debates in the history of philosophy that are clearly about conceptual engineering, such as the debate between Carnap and Strawson about the method of explication, which we discuss below. However, the field as a whole is relatively young, with major works published only in the early 2000s.

Conceptual engineering can be undertaken for many different reasons and in different ways. Chalmers (2020) distinguishes between *de novo engineering* and *re-engineering*: de novo engineering consists of the construction of new concepts, whereas re-engineering is 'fixing' or 'replacing' existing concepts. More broadly, one might distinguish between three ways of engineering our conceptual schemes: (i) changing an existing concept in a way that retains that very concept through the change; (ii) replacing an existing concept with a new one that is intended to perform better than the old one; (iii) introducing a

totally new concept that has no ancestors. Each of these kinds of projects brings unique desiderata and success conditions. It is crucial to keep these distinctions in mind when characterizing or evaluating a given conceptual engineering proposal.

For purposes of this book, we are particularly interested in projects of conceptual engineering that arise in response to technological disruptions. Not all projects of conceptual engineering fit these parameters. Indeed, the 'engineering' metaphor notwithstanding, until recently, mainstream work in conceptual engineering has not focused much on conceptual engineering in response to technological developments (Hopster and Löhr, 2023). However, as examples from the previous chapters illustrate, technology often plays a prominent role in conceptual disruption and conceptual change, in many more instances than previously noticed and discussed. Those interested in the debate about conceptual engineering are natural allies of those who investigate technological and moral change.

Consider the definition of a 'planet' by the International Astronomical Union (IAU). Prior to the late twentieth century, it was commonly believed by scientists and the public that there were a relatively small number of planets, certainly fewer than one hundred, and perhaps a large number of smaller objects orbiting the sun. With advances in astronomical technology, scientists discovered a large number of objects in the Kuiper belt that are similar in size to Pluto. Moreover, they expected to find hundreds or even thousands of these objects. The new astronomical technology led to a conceptual disruption: (i) most people believed there were a small number of planets orbiting the sun, but (ii) scientists discovered using the new technology that there are hundreds of objects like Pluto orbiting the sun. Claim (i) seems to be something like a conceptual truth about the idea of a 'planet'. Claim (ii) is the scientific discovery that came from technological advancement, leading to the conclusion that there are hundreds of planets. Scientists felt it was urgent to address this conceptual disruption. They could have embraced the claim that there are hundreds of planets in our solar system. This option, however, would have required major changes in our concept of 'planet' since being a planet would no longer be a special category with only a few members. The other option was to redefine the term 'planet'

so as to exclude the hundreds of Kuiper belt objects, and this is the line the IAU actually pursued.

This account of the new definition of 'planet' and the uproar about Pluto illustrates the three stages typically involved in processes of conceptual engineering. First, a conceptual challenge arises, which is often brought about by new technologies. Second, conceptual engineers question what should be done about this disruption. A key issue that arises at this second stage, with an eye to procedural justice, is the question of who is involved in arriving at this verdict. Is it a call for a certain group of experts to make, such as the IAU in the case of Pluto? Or should others be involved? The third stage is that of implementation: how do conceptual engineers go about spreading the word and ensure uptake in relevant communities of concept users?

By what standards should we decide on the aptness of our concepts? This is the question about appropriate criteria for the conceptual engineering process. One possibility, advocated by Sally Haslanger, is to connect the project of conceptual amelioration — improving our concepts — specifically with social and political aims. For example, Haslanger (2000) argues that we should drop the terms 'mother' and 'father' and only use the term 'parent', to facilitate the fight for gender equality. But the success conditions of conceptual engineering might also be understood in different terms, which need not be explicitly social or political.

The questions we are asking here, to use an analogy from the field of engineering and design, concern the appropriate 'design requirements' for engineering concepts. Related questions include the following: When ought one spend the time to evaluate a concept and decide whether it is effective enough to keep as is? When ought one investigate possible changes when one identifies a conceptual disruption? What is the best way to decide on a course of action to implement a conceptual engineering project? One kind of answer appeals to the idea that concepts have a function, and that the need for conceptual engineering arises insofar as we need to improve that function, ensure its continuation, or prevent it from failure. This perspective naturally leads to a question about the nature of conceptual functions, which is a debated issue in the current conceptual engineering literature (Queloz, 2019; Klenk, 2021). For present purposes, we need not take any stance in that particular

debate; what matters is that concepts serve *some* function which we deem desirable, such that conceptual changes can potentially be regarded as adaptations or improvements — changes which make it the case that concepts better serve this function. Still other scholars wonder whether concepts have any functions that can be specified in a substantive way.

In closing this section, it is important to bear in mind that the field of conceptual engineering is not without dispute. There are various problems that critics of conceptual engineering have raised. Let us consider one such problem: the challenge that changing concepts amounts to changing the subject. On Carnap's understanding, conceptual engineering can be understood as a method of explication which advises philosophers to provide determinate, scientifically rigorous definitions for important philosophical terms that hitherto had fuzzy or merely intuitive definitions. Strawson's objection to explication has come to be seen as a basic issue that almost any conceptual engineering effort ought to be able to address (Carnap, 1950; Strawson, 1963). Strawson said essentially that explication is merely changing the subject. It does not address the original philosophical issues associated with the term in question. As such, explication ought not be seen as a legitimate philosophical methodology, because it leaves all the important philosophical problems untouched.

While there are many responses to Strawson's objection, it might be insightful to consider the planet example discussed above. This is a clear case of explication since it involved rejecting an intuitive meaning for 'planet' and adopting a more rigorous and scientifically acceptable meaning. It ought to be clear that these are two different meanings since they have two distinct extensions. Pluto is a member of the extension for the old meaning, but Pluto is not in the extension for the new meaning, and the same goes for the rest of the dwarf planets. Did the IAU merely change the subject, as an advocate of Strawson's objection would contend? They certainly did change what the word 'planet' is about. It was about a particular property and now it is about a slightly different property. So it seems like the answer is yes: the IAU did change the subject of the word 'planet', and thereby changed the subject of discussions using the word 'planet'. Was this pointless, as the Strawsonian suggests? Not at all. It is better for science to have discussions about the new subject

matter, rather than about the old one. Sometimes changing the subject matter is exactly what is needed to address a problem.

6.5 Implications for ethics of technology

The aim of the ESDiT program is not just to study how fundamental and moral concepts are disrupted by emerging technologies, nor is it just to propose improved and new concepts in response. It is also to innovate ethics and political philosophy generally, and ethics and political philosophy of technology specifically. Due to social and conceptual disruptions, ethics and philosophy may not always be using an optimal conceptual framework. Improving this framework may contribute to better theories and methods in the field, which will yield better results.

This outlook rests on the view that many of the key concepts of philosophy, fundamental and moral concepts like 'nature', 'agency', 'mind', 'justice' and 'liberty' are historically contingent. They do not exist in every society and culture, and have not always existed in past epochs. Also, even when these key concepts stably exist at different times and places, their meaning often varies. This is just an historical observation, which does not entail any normative claims on whether some concepts are better than others, or whether some concepts have universal validity. But it does prompt serious inquiry into the nature of conceptual change, disruption, and modification.

This outlook also rests on the view that, as contingent constructs, concepts always have flaws and limitations and can be improved upon. Moreover, as societies change, whether these are changes in natural conditions, social and economic structures, or cultural practices and beliefs, changes in fundamental and moral concepts may be desirable so as to ensure their usefulness under new conditions. For example, with growing secularization, individualization and availability of contraceptives in the 1960s, most people have concluded that concepts like purity, chastity and temperance are no longer useful, and should be replaced by concepts more fitting to modern sexual relationships, like openness, trust, commitment and respect. Similarly, many moral and fundamental concepts of philosophy are centuries old, and may not make a great fit with our twenty-first century world. An attempt at

conceptual engineering, that takes recent major changes in the world into account, would therefore be helpful.

Conceptual engineering is not new in philosophy. The field of environmental philosophy was made possible in large part because of the introduction of new concepts and changes in the meaning and scope of existing concepts. In particular, this includes the expansion of the concepts of moral patienthood and intrinsic value to include environmental entities and the introduction of the concepts of sustainability and sustainable development. Similarly, the introduction of the philosophical notion of natural equality in the seventeenth century by Hobbes and Locke, which was then translated into moral and political equality and equality before the law, has enabled the whole social contract tradition of political philosophy as well as the tradition of liberalism and its conception of individual rights. The introduction of the notion of privacy in the late nineteenth century has enabled philosophers to articulate and study a dimension of human autonomy and well-being that they might not have been able to discern and study as well otherwise.

What these examples show is that conceptual engineering, involving both the introduction of new concepts and the modification and improvement of existing ones, has historically been common in philosophy, and has helped the field to progress. We have argued in this book that unprecedented recent and forthcoming changes in society, brought about in large part by socially disruptive technologies of the recent past, present and future, require conceptual engineering of many of our fundamental philosophical and moral concepts, and that this may change our methods of doing ethics and philosophy.

Exciting challenges lie ahead. More needs to be understood about appropriate criteria and legitimate ways for conceptual engineering. That is, we should understand more about potential deficiencies that our concepts may suffer from, and legitimate goals to which they may be put to use. Insofar as the goal is to create effective change in how entire groups and societies use concepts, there is an important practical and normative question about how to implement such changes. The 'politics of implementation' (Queloz and Bieber, 2022) will be an urgent and important area for further inquiry.

Another challenge for ethics of technology is how to best study conceptual disruption and conceptual change through technology. This book has offered some building blocks for doing so, both theoretically as well through practical examples of technologies that (allegedly) disrupt existing conceptual schemes. But this leaves ample room for further refinements and raises various further questions. Consider the awareness we have raised that conceptual disruption always occurs relative to a given conceptual framework, which urges us to take a more intercultural perspective than has been common in academic philosophy. Yet how, exactly, should the conceptual schemes of a given cultural community be delineated?

Another question concerns the methods to (empirically) study conceptual disruption and change. One potential method might be the analysis of corpus linguistics. Arguably, changes in concepts, including conceptual disruption, have some correlation with changes in language and words used, for example for describing certain phenomena or asking ethical questions about them. There are various existing methods for detecting semantic change in text corpora that might also be promising for studying conceptual disruption and change (e.g. Hamilton et al., 2016). Similarly, methods are being developed to use text corpora to study value change in relation to technologies (De Wildt et al., 2022).

As we saw in the introduction, philosophy and ethics of technology do not just want to understand technologies' impacts and disruptive potential, but also seek to contribute to better technologies in a better society. They do so by closely interacting with scientists and engineers and also with policy makers. Such a contribution requires attention not just to conceptual disruption and conceptual engineering, but also to how technologies are socially disruptive in cases that do not disrupt concepts. Such a contribution may also require new methods and approaches. For example, it may require new methods for studying and developing technologies in living labs, in which their disruptive potential is both studied and addressed (e.g. through design choices) in close collaboration with designers, engineers, artists and relevant stakeholders. It may also require adaptation of existing methods for responsibly developing new technology, such as value-sensitive design (VSD) (Friedman and Hendry, 2019). For example, Veluwenkamp and

Van den Hoven (2023) have proposed an approach to integrate insights from conceptual engineering into the VSD approach.

Ultimately, addressing the challenges brought by twenty-first century technologies — such as social media, social robots, climate engineering, and the artificial womb — requires not just the engineering of technology (as traditional engineers and designers have done) or just conceptual engineering (as proposed in philosophy), but a synthetic combination of both, with an eye to the fundamental values we want to uphold.

Further listening and watching

Readers who would like to learn more about the topics discussed in this chapter might be interested in listening to these episodes of the ESDiT podcast (https://anchor.fm/esdit) and other videos:

Guido Löhr on 'Do socially disruptive technologies really change our concepts or just our conceptions?': https://podcasters.spotify.com/pod/show/esdit/episodes/Guido-Lhr-on-Do-socially-disruptive-technologies-really-change-our-concepts-or-just-our-conceptions-e1uhlj2/a-a99r808

Arché *Conceptual Engineering* series on YouTube, set up by Kevin Scharp: https://www.youtube.com/c/ConceptualEngineering

References

Babushkina, Dina, and Athanasios Votsis. 2022. 'Disruption, technology and the question of (artificial) identity', *AI & Ethics*, 2: 611–22, https://doi.org/10.1007/s43681-021-00110-y

Baker, Robert. 2013. *Before Bioethics: A History of American Medical Ethics from the Colonial Period to the Bioethics Revolution* (New York: Oxford University Press)

Bryson, Joanna. 2018. 'Patiency is not a virtue: The design of intelligent systems and systems of ethics', *Ethics and Information Technology*, 20: 15–26, https://doi.org/10.1007/s10676-018-9448-6

Burgess, Alexis, Herman Cappelen, and David Plunkett (eds). 2020. *Conceptual Engineering and Conceptual Ethics* (Oxford: Oxford University Press), https://doi.org/10.1093/oso/9780198801856.003.0001

Burgess, Alexis, and David Plunkett. 2013a. 'Conceptual ethics I', *Philosophical Compass*, 8(12): 1091–1101, https://doi.org/10.1111/phc3.12086

——. 2013b. 'Conceptual ethics II', *Philosophical Compass*, 8(12): 1102–10, https://doi.org/10.1111/phc3.12085

Cappellen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering* (Oxford: Oxford University Press)

Cappelen, Herman, and David Plunkett. 2021. 'A guided tour of conceptual engineering and conceptual ethics', in *Conceptual Engineering and Conceptual Ethics*, ed. by Alexis Burgess, Herman Cappelen, and David Plunkett (Oxford: Oxford University Press), https://doi.org/10.1093/oso/9780198801856.003.0001

Carnap, Rudolf. 1950. *Logical Foundations of Probability* (Chicago: The University of Chicago Press)

Carpinella, Colleen, Alisa Wyman, Michael Perez, and Steven Stroessner. 2017. 'The Robotic Social Attributes Scale (RoSAS): Development and validation', *HRI '17: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 254–62, https://doi.org/10.1145/2909824.3020208

Chalmers. David. 2020. 'What is conceptual engineering and what should it be?', *Inquiry*, https://doi.org/10.1080/0020174X.2020.1817141

Coeckelbergh, Mark. 2010. 'Robot rights? Towards a social-relational justification of moral consideration', *Ethics and Information Technology*, 12: 209–11, https://doi.org/10.1007/s10676-010-9235-5

De Wildt, Tristan, Ibo van de Poel, and Emile Chappin. 2022. 'Tracing long-term value change in (energy) technologies: Opportunities of probabilistic topic models using large data sets', *Science, Technology, & Human Values*, 47: 429–58, https://doi.org/10.1177/01622439211054439

Eklund, Matti. 2015. 'Intuitions, conceptual engineering, and conceptual fixed points', in *The Palgrave Handbook of Philosophical Methods*, ed. by Chris Daly (London: Palgrave), 363–85, https://doi.org/10.1057/9781137344557_15

Eriksen, Cecilie (2020). *Moral Change: Dynamics, Structure, and Normativity* (Cham: Palgrave Macmillan)

Friedman, Batya, and David Hendry. 2019. *Value Sensitive Design: Shaping Technology with Moral Imagination* (Cambridge: MIT Press)

Gallie, Walter. 1955. 'Essentially contested concepts', *Proceedings of the Aristotelian society*, 56: 167–98

Gunkel, David. 2018. *Robot Rights* (Cambridge: MIT Press)

Hamilton, William, Jure Leskovec, and Dan Jurafsky. 2016. 'Diachronic word embeddings reveal statistical laws of semantic change', in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (*Volume 1:*

Long Papers) (Berlin: Association for Computational Linguistics), 1489–501, https://doi.org/10.18653/v1/P16-1141

Haslanger, Sally. 2000. 'Gender and race: (What) are they? (What) do we want them to be?', *Noûs*, 34(1): 31–55 https://doi.org/10.1111/0029-4624.00201

Hannon, Elizabeth, and Tim Lewens (eds). 2018. *Why We Disagree About Human Nature* (Oxford: Oxford University Press)

Hopster, Jeroen. 2021a. 'What are socially disruptive technologies?', *Technology in Society*, 67: 101750, https://doi.org/10.1016/j.techsoc.2021.101750

——. 2021b. 'The ethics of disruptive technologies: Towards a general framework', in *Advances in Intelligent Systems and Computing*, ed. by Juan de Paz Santana, Daniel de la Iglesia and Alfonso José López Rivero (Cham: Springer), https://doi.org/10.1007/978-3-030-87687-6_14

Hopster, Jeroen, and Guido Löhr. 2023. 'Conceptual engineering and philosophy of technology: Amelioration or adaptation?', Unpublished manuscript

Jacobs, Naomi (2023). 'De kunstmatige baarmoeder', *Wijsgerig Perspectief*, 1.

Jamieson, Dale. 2015. 'Responsibility and climate change', *Global Justice: Theory, Practice, Rhetoric*, 8(2): 23–47, https://doi.org/10.21248/gjn.8.2.86

Kelbessa, Workineh. 2015. 'African environmental ethics, Indigenous knowledge, and environmental challenges', *Environmental Ethics*, 37(4): 387–410, https://doi.org/10.5840/enviroethics201537439

Klenk, Michael. 2021. 'Moral realism, disagreement, and conceptual ethics', *Inquiry*, https://doi.org/10.1080/0020174X.2021.1995483

Lindauer, Matthew. 2020. 'Conceptual engineering as concept preservation', *Ratio*, 33(3): 155–62, https://doi.org/10.1111/rati.12280

Löhr, Guido. 2022. 'Linguistic interventions and the ethics of conceptual disruption', *Ethical Theory and Moral Practice*, 25: 835–49, https://doi.org/10.1007/s10677-022-10321-9

——. 2023. 'Do socially disruptive technologies really change our concepts or just our conceptions?', *Technology in Society*, 72: 102160, https://doi.org/10.1016/j.techsoc.2022.102160

Nickel, Philip. 2020. 'Disruptive innovation and moral uncertainty', *NanoEthics*, 14(3): 259–69, https://doi.org/10.1007/s11569-020-00375-3

Nickel Philip, Olya Kudina, and Ibo van de Poel. 2022. 'Moral uncertainty in technomoral change: Bridging the explanatory gap', *Perspectives on Science*, 30(2): 260–83, https://doi.org/10.1162/posc_a_00414

O'Neill, Elizabeth. 2022. 'Contextual integrity as a general conceptual tool for evaluating technological change', *Philosophy & Technology*, 35: 79, https://doi.org/10.1007/s13347-022-00574-8

Queloz, Matthieu. 2019. 'The points of concepts: Their types, tensions, and connections', *Canadian Journal of Philosophy*, 49(8): 1122–45, https://doi.org/10.1080/00455091.2019.1584940

Queloz, Matthieu, and Friedemann Bieber. 2022. 'Conceptual engineering and the politics of implementation', *Pacific Philosophical Quarterly*, 103: 670–91, https://doi.org/10.1111/papq.12394

Rawls, John. 1999. *A Theory of Justice* (Cambridge: Belknap Press)

Rueda, Jon, Jonathan Pugh, and Julian Savulescu. 2022. 'The morally disruptive future of reprogenetic enhancement technologies', *Trends in Biotechnology*, 2258: 1–4, https://doi.org/10.1016/j.tibtech.2022.10.007

Scharp, Kevin. 2013. *Replacing Truth* (Oxford: Oxford University Press)

Spatola, Nicolas, and Thierry Chaminade. 2022. 'Cognitive load increases anthropomorphism of humanoid robots. The automatic path of anthropomorphism', *International Journal of Human-Computer Studies*, 167: 102884, https://doi.org/10.1016/j.ijhcs.2022.102884

Strawson, Peter. 1963. 'Carnap's views on constructed systems versus natural languages in analytic philosophy', in *The Philosophy of Rudolf Carnap*, ed. by Paul Schilpp (LaSalle: Open Court), 503–18

Veluwenkamp, Herman, Marianna Capasso, Jonne Maas, and Lavinia Marin. 2022. 'Technology as driver for morally motivated conceptual engineering', *Philosophy & Technology*, 35: 71, https://doi.org/10.1007/s13347-022-00565-9

Veluwenkamp, Herman, and Jeroen van den Hoven. 2023. 'Design for values and conceptual engineering', *Ethics and Information Technology*, 25: 2, https://doi.org/10.1007/s10676-022-09675-6

Watene, K. 2022. 'Indigenous philosophy and intergenerational justice', in *Reimagining the Human-Environment Relationship*. United Nations University, http://collections.unu.edu/eserv/UNU:8829/UNUUNEP_Watene_RHER.pdf

Wiredu, Kwasi. 1994. 'Philosophy, humankind and the environment', in *Philosophy, Humanity, and Ecology*, ed. by H. Odera Oruka (Nairobi: ACTS Press)