Edmond, Jennifer, Nicola Horsley, Jörg Lehmann, and Mike Priddy. "Expatriates in the land of data: Software tensions as a clash of culture." *The Trouble With Big Data: How Datafication Displaces Cultural Practices*. London,: Bloomsbury Academic, 2022. 159–177. *Bloomsbury Collections*. Web. 26 Apr. 2023. <<http://dx.doi.org/10.5040/9781350239654.ch-8>>.

Downloaded from Bloomsbury Collections, www.bloomsburycollections.com, 26 April 2023, 12:46 UTC.

Copyright © Jennifer Edmond, Nicola Horsley, Jörg Lehmann and Mike Priddy 2022. Released under a CC BY-NC-ND licence (<https://creativecommons.org/licenses/by-nc-nd/3.0/>). You may share this work for non-commercial purposes only, provided you give attribution to the copyright holder and the publisher, and provide a link to the Creative Commons licence.

Expatriates in the land of data: Software tensions as a clash of culture

More questions than answers?

The preceding chapters have presented myriad examples and evidence of the tensions that result from the intersection between common cultural practices and datafication's whispered promises of progressing the efficiency and connection for which we depend upon them. We have shown how power relations are being rebuilt according to the voices that get heard and the capital that is accumulated via 'surveillance capitalism'.¹ We have shown how histories (or perhaps futures) become rewritten, and how identities can be shaped by overarching imaginaries, like the nation states of old. We have explored the promises of knowledge creation paradigms based upon big data and shown how words we think we understand come to accrete rich new fields of meaning that short circuit rather than facilitate mutual understanding. We have shown how trusted institutions, from legal contracts to library collections, become divorced from the values and purposes they have traditionally held.

The discussion throughout this book allows us to reframe questions of the future development of technologies based on big data away from the very narrow issues such as data protection and the prevention of direct abuse towards questions that are far more holistic and centred of the human experience, such as:

- § *Do current data regulation platforms go far enough, and should algorithms be regulated?*
- § *How can the needs of the wider society be taken into account within an ecosystem in which software is generally designed to meet the needs of a client or a user?*

- § *How can the tension between what a client wants to see and the requirements for decision making be resolved against a backdrop of software as a product which needs to be user friendly, to be competitive?*
- § *How can the tendency of those currently wielding power over and through the digital to consolidate access, power and wealth be shifted towards a more equitable, redistributed benefit? As data inequality emerges as a major theme in data-driven society, what are the consequences of this inequality in terms of technology, infrastructure and employment?*
- § *Is there a potential to facilitate redistribution between the data-rich and data-poor and other kinds of economic inequality?*
- § *How do our digital practices evidence the gendered, classed and racialized bias that are relayed via everyday practices and exclusions?*
- § *How are hierarchies and everyday discrimination being sublimated into our data and platforms, how do the implications of this process infringe upon human dignity and diversity, and how can big data systems empower rather than marginalize under these conditions?*
- § *What does 'narrative' mean within the context of the data-focused computer sciences, and how does this tacit definition effect human sensemaking in the digital age?*
- § *The relationship between narrative and data has received much attention from philosophers of science, being variously presented as antagonistic, antithetical or even symbiotic; existing in relationships that can pose ethical and epistemological challenges for the researcher or software engineer. Less has been said about how this relationship is perceived within the computer science community. How can one make data comprehensible if you define it a priori as patternless, objective input?*
- § *How does the human need to make sense through storytelling become recast in an era that seems to fetishize data as having an almost god-like claim to truth?*
- § *With the growing push for Government regulation of AI, the 'Driverless Dilemma' has emerged as a uniquely computational development on the well-established 'trolley problem' wherein a driverless car must choose between the lesser of two evils. Rather than asking whether it is possible to design something that can make machines more humane, should we consider whether it is possible to design machines that make people more humane?*
- § *If human happiness requires human agency, why therefore are we so concerned about machines either taking on human characteristics (such as creativity or imagination) or not having them (like ethics)?*

- § *Why has complexity become coterminous with stress and frustration, or viewed as a risk to be managed?*
- § *How do we invest trust and delegate decision making to algorithms when we do not understand fully how they function (e.g. black box, deep learning) or know what it means to 'trust' a machine?*
- § *Do we believe that the decomposition of complex tasks into executable subroutines perhaps changes the nature of both the question and the answer (or do we care)?*
- § *If we accept the proposition that attention should be viewed as a human right and privacy as a public good, how do we draw lines between selves and others in data environments where we cannot see what (or whom) we expose, and how (for public and private are not clear binaries)?*
- § *How are the pressing issues of the current age bigger and more complex than earlier questions of artisanship versus mass production?*
- § *Are we now tasked with moving towards a post-human algorithm, one that is free of the biases and prejudices (conscious or unconscious) that currently delimit human drives and human(e) choices?*
- § *Finally, with all of its jargon, block boxes and immateriality, how can we look critically at how we talk about digital technology? What lessons can we learn from this exercise?*

The answers to these questions are beyond the scope of this book, but they do illustrate the range of issues that can be opened up to new approaches through a more fundamental incorporation of humanistic and cultural perspectives into the development and critique of big data and indeed software-driven practices, and how necessary such approaches are. A few things we can conclude are, however, that:

- recommender systems based on highly effective algorithms preform individual aesthetic choices and decisions, which can be understood as a massive intervention into consumption habits with yet unknown consequences, leading to their normalization in culture
- the digitization of cultural commons secures the status of the Big Tech corporations and serves the end of predictive analytics based on the 'behavioral surplus' exploited by companies such as Google Books, Amazon Kindle, and Netflix. This is an exploitation of cultural heritage for the profit of private companies and can be termed a 'tragedy of the commons'.²
- Big Tech companies have begun to establish content and assume the

role of producers and distributors of cultural products, a development that threatens to displace human cultural production (which, it has been pointed out, serves social functions as processes first, and economic ones as products, second³). This trend towards algorithmic cultural production is a form of cultural imperialism or colonialism on the part of the big technology companies.

- the analysis of user behaviour allows the monopolization of knowledge creation processes by private corporations (Amazon, Google Books), which in turn supports a de-differentiation of knowledge production and its uncoupling from the customary social embedding into society.
- what was once the manageable realm of ‘discourse’ has now become a strange, detached area of information/disinformation/fake news/alternative facts; we have lost our sense of the social dimension of information processing in a world which seemingly has become reigned by algorithms which are not understood by anybody. After all, if the organization responsible for creating a system can’t clearly define what they mean when they speak of our data’s place in it, how are we supposed to understand any risks we take?

What this book can also do is to provide the framework by which to deliver this more holistic approach, enabling a methodology for the kinds of further studies indicated above. To do this, we need to get beyond viewing culture and datafication as either one and the same or as neutrally disconnected, but perhaps instead as manifestations of different values, hierarchies and beliefs – that is, different cultures.

Is software production also a culture?

These are broad and important considerations, and probably not the ones at the front of your mind as you download a new app or click to agree to its terms of service. Such questions are not individual, so much as cultural, and the problems they raise are only visible at the level where our collectives inform our actions and identities. As such, if you have ever travelled abroad, leaving your home country to experience the language, norms, practices and beliefs of another culture, you may have already experienced the best-possible analog for the state of excitement and expansion of your understanding of the world, but also alienation and tension, that these kinds of technologies place us in. Big data, for all the benefits

it brings, has also given rise to tensions and perversions not only of things we think we know, but who we think we are, disrupting practices so intrinsic to individual and social relationships and so sublimated in cultural practices that we can sense, but not necessarily verbalize the change. To address challenges that are hard to perceive, because of our limitations, position, biases or lack of understanding, we need tools to enhance our 'vision': but such instruments, be they microscopes or clean rooms, need to be closely adapted to the precise object of study and the effect it has. This will be no simple achievement. Even the 'digital natives' of the 'Google generation' are not immune to this effect: as one journalist wrote, 'as a typical millennial constantly glued to my phone, my virtual life has fully merged with my real life. There is no difference any more. Tinder is how I meet people, so this is my reality. It is a reality that is constantly being shaped by others – but good luck trying to find out how.'⁴

Part of the insidious and troubling nature of this effect is the manner in which technology has encroached upon our free will in ways that we often cannot even see. We can rebel against censorship when we note something is missing, but how do we seek out perspectives we do not even know we are being denied? Technologies of a certain sort, those that gate not our interaction with the world of objects, but with the inner world of our interpretation and understanding, of the social and the cultural, have always had this effect, to a certain extent. Indeed, it was Plato who gave us the first argument concerning how a new knowledge technology (in Plato's case, writing) gave rise to concerns about the impact it would have on 'memory and wit'. And if writing was the first such case, it was by no means the last: technologies as diverse as the printing press, open library shelves and television have all been heralded as agents to degrade our existing capacities.

Big data and AI, its processing-facing counterpart in the ecosystem of technology assemblages that are growing so exponentially at the moment, can be differentiated from their many predecessors in a number of ways, however. The opacity of these technologies is one reason for this. Big data is big precisely because it is not human-readable; deep fakes are a threat because they are so true to life. Big data and AI lurk silently behind many of our most comfortable spaces, be that the public square where facial recognition is applied to the data streams from its security cameras, the app through which we get to know our future partner or the advertisement targeting serving us credible-looking fake news via our personal social media feeds.

More concerning still, however, is the question of power discussed in Chapter 7, concerns that lurk but one layer further below these algorithms

using our personal data to target us. In some cases, this power may indeed be in the hands of a foreign government: the EU High Level Expert Group on fake news and online disinformation⁵ recognizes the role of foreign governments in instigating disinformation without naming any particular names, while the rise in cyberterrorism activities by Russia in particular has been highlighted in security committee reports of the US and UK governments.⁶ Equally, however, the underlying power may be that of a skilled individual rather than an institutional programme, motivated perhaps by economic need (as in the case of the clickbait mills that arose in the former Yugoslavia in the 2010s⁷), youthful mischief (as in recent hacking of the Twitter accounts of prominent individuals) or a vigilante streak (as in the case of the Anonymous hacktivist collective). But most often, the motivations behind the algorithmic manipulation of social behaviour is not so much a failure of the social and technical systems, but a success, as the greatest power behind the big data systems that control us is of course the companies that in turn control them. In the year 2000, the Fortune 500 list of America's largest companies included only one technology company: IBM. The rest of the list was populated with the usual mix of manufacturing (Ford, Boeing), services (Citigroup, AT&T) and retail (Walmart).⁸ Although the manner in which you define 'largest' here matters, by 2020, international rankings – in terms of market capitalization at least – tell a very different story, with only three of the top ten companies not being the high-tech giants, and with Chinese entries like Tencent and Alibaba gaining ground quickly on Microsoft, Apple, Alphabet, Facebook and Amazon.⁹

These companies, valued in the billions, if not trillions, of dollars, have a scale and economic power that eclipses that of many countries. So, if they have a similar scale to countries, do they also have their own culture? If we define culture as the 'knowledge, beliefs, expectations, values, practices, and material objects by means of which we craft meaningful experiences for ourselves and with each other',¹⁰ the answer may seem to be yes. While one might be tempted to view many of these at least as promoters of a very American form of economic and cultural corporate imperialism, there are enough distinct properties common to these subcultures that one can perhaps also speak of a culture of software production, manifested in a view of the world and a sense of place in it that is promoted by and within these collectives, disciplines and organizations. To understand this, we can draw not only on theories of how national cultures manifest in organizational contexts, but also on the many accounts that have been written about the prevailing norms, hierarchies, biases, beliefs, habits of speech, narratives of origin and identity and tolerances within

interpersonal interaction, in particular as pertains to issues such as moral and ethical boundaries, perceptions of privacy and so on that have been observed and documented within software companies.

The foundational work regarding cultural approaches to workplace identities and interactions of Geert Hofstede maps in very intriguing ways on to the questions of the work-based habits and cultures of software developers, in spite of the different frame of reference underlying his conclusions (the impact of national cultures on multicultural workplaces and business interactions).¹¹ Although not all of Hofstede's dimensions map equally precisely on to what can be observed in software development teams, and although one does not want to risk falling back on stereotypes, reflecting on them in this context provides an interesting window for beginning to view software development as a work context with distinct cultural overtones. In particular, the following discussion focuses on five of Hofstede's parameters, in an attempt to at least get a bearing on the centre of gravity for the values and norms underpinning software production.

An obvious place to start is with the dynamic of MASCULINITY VERSUS FEMININITY. This is relevant in spite of the fact that the early workforce of the tech industry, if one can even apply that term to the ecosystem of the 1940s and 1950s around the large main frame systems such as the ENIAC, was strongly feminized. This is clearly not the case today, as studies like Hardey's¹² make all too clear. At that time, coding was viewed as manual labour, 'low on the intellectual and professional status hierarchy'.¹³ As the centrality and complexity of the coding process became recognized, however, these tasks were more and more taken over by men. Hand in hand with this shift also came the increasing association of the attributes of a good coder with traditional values of masculinity: an early attempt to develop a psychological profile of programmers concluded that as a group they were avid problem solvers, liked to learn new things and didn't like people.¹⁴ Though one does not want to take gender differences too far towards essentialism, one can see how coding became defined as a space away from the emotional labour so often associated with women in Western societies, 'a refuge from the unpredictability of humans, from their greyscale emotions and needs'.¹⁵ Expanded from a psychological profile to a cultural one, the presence of gender biases in data, code and software products but also a certain engrained tension, if not outright hostility, towards end users of software systems and products can be seen as manifestations of two distinct sets of prevailing attitudes about the world coming into unresolved conflict.

Another interesting point of reflection within Hofstede's taxonomy of national traits is that of LONG or SHORT TERM ORIENTATION. It is perhaps

not surprising that software development teams and indeed companies may have an inherent and engrained bias towards the short term. With technology changing so fast, historical or long-term perspectives are inevitably absent in coding circles in particular, as new standards and platforms can arise very quickly to displace existing expertise. In this respect, software engineers may be productively compared to civil, mechanical and other engineers, less than half of whom remain active in development roles for more than a decade past their initial entry into the field.¹⁶ The founding motto of Facebook, 'Move fast and break things', speaks very much to this prevailing attitude in a sector in which the workforce in all of the top tech companies has a median age under forty, with a handful of the leaders in the data revolution, such as Facebook and Google, having a median age below thirty.¹⁷ This leads to what one critic has called 'a world flush with cash and devoid of experience'.¹⁸ And this is surely the point of Facebook having a huge campus providing (controlling) leisure activities, arts, favouring particular aesthetics – demonstrating the culture of the Republic of Facebook, but also celebrating the transformation of technology companies from the conservative dark suit uniform of IBM to the playground-like atmosphere favoured by the successor generation.

The criteria around POWER DISTANCE and INDIVIDUALISM VERSUS COLLECTIVISM are reflected in the manner in which hierarchies and in groups are formed in software development teams. The same culture that has given us the hostility towards users mentioned above has also given us the open-source movement; the same ecosystem that spawned Facebook's 'move fast' motto gave us Google's founding motto of 'don't be evil' (though how much this still resonates with the values and decisions of the company is an open question). Indeed, the lack of diversity in these circles clashes with the industry view of itself as principled, open and meritocratic. The field can appear this way because self-taught amateurs work alongside people with PhDs and because of a certain hacker ethic that brought '... a shift in who became a coder and why. For the first time, programmers were emerging in living rooms, as teenagers, propelled by the culture of making, acquiring, and sharing software.'¹⁹ The curiosity, problem orientation and perhaps youthful idealism of the founding generation of software company CEOs have not scaled well as the companies intended to embody these values have acquired trappings that instead befit their size and economic power. Perhaps as a result, coders seem to maintain a stronger loyalty to their tribe, rather than to any particular company or employer.²⁰

UNCERTAINTY AVOIDANCE also seems rather baked into the practice of coding, which is highly unforgiving of irregularities or uncertainties. Jeff

Atwood, the founder of programmer's Q&A resource, Stack Overflow, captured this aspect in a reflection on how the interactions on his site represented the best and worst of coder behaviour, with the computer itself as a toxic, abusive colleague: 'The reason a programmer is pedantic is because they work with the ultimate pedant. All this libertarianism, all this "meritocracy", it comes from the computer. I don't think it's actually healthy for people to have that mind-set. It's an occupational hazard. Not everyone is like this. But on average it's correct.'²¹ This aspect of software culture became rather a leitmotiv for the KPLEX project, in fact, as so many of the conflicts we observed came down to a desire to preserve the 'hidden treasures' of complexity and the traces of uncertainty in systems with very little tolerance for this.²²

Finally, Hofstede's categories differentiate between INDULGENCE versus RESTRAINT. This dynamic, so resonant of the preceding discussion of Rousseau's critique of the Enlightenment, may again be visible in the stereotype of the corporate office of the software age (Google, Apple) rather than the hardware age (IBM). Interestingly, however, while this indulgence seems to apply to the trappings of the Fortune 500 lifestyle, it certainly does not reflect the rigour of coding itself, of the methodologies for managing the work processes of software development (such as Agile), or indeed the tendency noted above to put the requirements of the system ahead of those of the user.

Cross-cultural competencies for a Digital Age

This tacit cultural subtext that informs the behaviour of software companies and development teams has wide-ranging effects on the world that now relies on the services they provide, and the flows of how big data is collected and used are a flashpoint where miscommunication, unconscious biases and an underestimation of cultural differences can lead to disasters. At its more benign frontier, the technocratic drive for efficiency suffuses and disrupts our lives, with the fundamental value of optimization and scale being 'what has led to many collisions between software firms and life'.²³ The vulnerability of our attentional spheres²⁴ is exploited by the ubiquity of notifications and filters designed (in theory at least) to ensure we have the information we need when we need it, but in reality this undermines self-determination in where we place our focus and when, a capacity named as 'intellectual autonomy' by one of the authors of this book.²⁵ But this perception of humans as essentially incapable can also lead to abuses, in which greater trust is placed in the efficiency of a

black box and model behind a data and algorithm-driven system than in social responsibility of our fellow human beings. The perceived superior efficiency and neutrality of machines in the face of ethical and challenging decisions have led to a range of cases in which delegation of traditionally human decisions has led to abuses of not just social contracts but human rights. In particular the deployment of the Compas system,²⁶ in which the opaque projections of past inequalities were used to determine the future confinement or freedom of individuals, demonstrates how the tension between human social responsibility and their fallibility can become dangerously entangled with efficiency.

Many of the systems are of course open to abusive deployments as well, as can be seen in the case of the Cambridge Analytica role in swaying public opinion and/or manipulating (via suppression or stimulation) of voting via social media in the 2016 US election and UK referendum on leaving the EU. The fact that the mechanism for gathering the data to build the system behind this masqueraded as academic research makes this abuse all the more galling. But even if such systems can be regulated, used properly and held to high ethical standards, the centrality of the perception that machine efficiency trumps the slow and underinformed decision-making processes of the human being can still lead to perverse and disconcerting outcomes, such as the evidence we see that humans, from pilots to translators, relying on artificial intelligence in their professional lives can actually lose competency over time.²⁷ If ethical decision making, scientific discovery and the development of informed citizens also come to be seen as equally inefficient processes, one may fear where the biases and blinders of technology will leave us.

Our fear of inefficiency leaves us open the abuses of our focus in the so-called 'attention economy'. An infamous holy writ of early technology, Nir Eyal's *Hooked*, made much of the many ways that human psychological weaknesses could be exploited to meet some of the technological success metrics of 'stickiness' and repeat visits.²⁸ This approach positions behavioural engineering as a direct inlet port for technology into culture, a value also confirmed by former Google senior manager (now founder of the Centre for Humane Technology and their centrepiece initiative 'Time Well Spent') Tristan Harris: 'the job of these companies is to hook people, and they do that by hijacking our psychological vulnerabilities.'²⁹

Perhaps an even better example of precisely how disciplinary and corporate cultures can lead to assumptions and ultimately to disasters on a grand scale, in particular where the drivers of these disasters may be hidden within the sea of big data behind a proprietary system, is that of Facebook's entry into Myanmar.

In this case, the collision of data-centric floods of misinformation, such as were discussed in Chapter 2 of this book, power dynamics and a disregard for a receiving culture are very clearly illustrated. In essence, the situation in Myanmar was a sort of ‘perfect storm’, where underlying tensions between social groups and a perfectly timed change in telecoms regulation left the country vulnerable to a rise in unrest. Facebook took advantage of this latter shift, establishing itself via a loss-leading pricing for data access, as the dominant face of the internet in the country. It then however ignored warnings that their platform was being used to share and broadcast hate speech. According to a Reuters news report³⁰ as of 2014, they reportedly had only one (Dublin-based) content reviewer who spoke Burmese on staff: by 2015 this number had risen only to four, for a pool of 7.3 million active users. And, even as late as 2018, the company had no actual staff in Myanmar, with only outsourced contractors offshore to oversee the interactions they were facilitating. It is perhaps easy to point a finger at Facebook as a single company, but the evidence shows this gaping hole in perception to be more widespread than only theirs: former Google CEO Eric Schmidt commented in 2009 that studying global usage patterns of his company’s search engine convinced him that ‘people are the same everywhere’, and that ‘people still care about Britney Spears in these other countries.’³¹

Facebook not only underestimated the cultural challenges in Myanmar, of language but also of social cohesion, they also overestimated the power of their machine translation systems to be able to accurately flag dangerous content, and provide a technical fix to a cultural problem. But of course, machine translation has traditionally only been as good as the big data available to train the software, which means that results seen functioning well between languages like English, French and German can raise highly misleading expectations regarding languages like Burmese (or indeed Icelandic, as was discussed in Chapter 7). The arrogance associated with this particular corporate asset can also be seen in Mark Zuckerberg’s own post from May 2017, in which he announced the publication of the Facebook translation software code. The post concluded with the self-congratulatory, and highly misguided, claims:

Throughout human history, language has been a barrier to communication. It’s amazing we get to live in a time when technology can change that. Understanding someone’s language brings you closer to them, and I’m looking forward to making universal translation a reality.³²

Claiming that language is a barrier to communication utterly perverts the role that this subtle and varied human achievement has played throughout the history

of our species. It also greatly overestimates the capacity of machine translation to capture the essence of communication: indeed, the best neural network-driven translations can often hide more than they share by producing very plausible natural language outputs, which, however, utterly misrepresent the sentiments of the original utterance. In the case of Myanmar, however, the communication seemed to have been all too clear, while the engrained hubris of Facebook and its engineers in the tools they had proven valid according to their own cultural norms, were unable to see this, even as it unfolded in real time on their own servers and the streets of Mandalay. And, while Myanmar may be the most extreme example of this, it is far from exceptional: as when it comes to product design, culture matters (though perhaps it is very often only addressed in a very simplistic manner). Speaking of an example of a development team considering what protections might be incorporated into an early version of Twitter, Clive Thompson writes:

The team of young guys who made it were, demographically, far less likely to have experienced online abuse. They didn't prioritize it early on as an inevitable, looming problem they would need to address. On the contrary, one staff member dubbed their company "the free-speech wing of the free-speech party." They designed few safeguards against harassment, and years later, trolls and white supremacists discovered that Twitter was a fabulous way to harass targets.³³

The manner in which shared identities and groupthink at odds with common values and norms seem to pervade the development of software is reminiscent of Benedict Anderson's seminal work on the idea of imaginary communities, group-based labels representing masses of shared cultural capital, utterly intangible, and yet inspiring of loyalty even unto death. Anjuan Simmon's theory that software companies have not only begun to act like countries, but, indeed like the colonial powers of old, is a both compelling and frightening extension of this: "Technology companies today are increasingly colonial in their actions. This can be seen in the veneer of sovereignty they seek to cultivate, how they work across borders, their use of dominant culture as a weapon, and the clear belief that "superior" technology is a suitable excuse for lawlessness, exploitation and even violence."³⁴

But where do these identities arise from? As described above, a certain amount of the acculturation of software engineers may come from personal predilections, preferences that may be linked to psychology or early experiences of reward and competence. But if values and habits can be identified in what is essentially a culture of immigrants (there may be habits

of mind and thought that attract people to coding, but one would assume that few people are born into such an identity), then we may also usefully look at how young software developers are trained to absorb the values of their tribe, the incentives and penalties that exist, and indeed where the possible tops of the hierarchies may be found (e.g. in training and degree programmes), as these could be entry points for investigation or even intervention that might bring data-driven approaches to knowledge back in line with identities and cultural affordances (or at least buffer their effect to a speed at which norms can assimilate them).

We have explored examples of things that have happened, but what about things that might or could happen? In particular, the symbiotic relationship between big data and AI raise the stakes for what new risks might arise in the future. Making sense of the masses of individually captured records of actions and phenomena, from the exact temperature at a certain point on the globe at a certain moment, to the search terms entered in every Google search you have ever made, requires the patience, precision and speed of a machine applying an algorithmic model to make sense of. And, sometimes, the software parsing this data is not merely looking for patterns defined by its designer, but is using the data it has been given to 'learn' from patterns and correlations that may exist in it, but which even the software developer may not have recognized. It is here that we enter the realm of artificial intelligence and can see how it builds upon the potential of the big data corpora by learning and extracting patterns from it. The problem is, however, that at a certain point, even the engineer who designed the system may not understand how the conclusions were derived from the training data. It is from this phenomenon that we can find data biases leading to the development of systems that identify people of colour as gorillas, and even systems that have been found to 'cheat' (e.g. by using embedded metadata about an image rather than the visual content itself as a way to identify the object of a photograph).³⁵ So problematic have some of these data sources been found to be when big data is used to train AI that in one case MIT had to permanently decommission the popular Tiny Images dataset and urge users to delete off-line copies they might be using, after the data was found to regularly produce racist and sexist labelling of the images.³⁶

As technology becomes more complex, the threats it poses escalate to become overarching, targeting democracy, security, the individual right to self-determination and collective claims on maintaining cultural practices undisturbed by the intrusion of companies with a distinct set of practices that may or may not accurately reflect those of their users. It is here where

a recognition of the cultural dimension of the conflict playing out can be particularly useful, as the long tradition of reflection and practice regarding the easing of intercultural tensions and misunderstandings can be applied to perhaps provide new foundations for developing personal and regulatory 'green lights and red lines' for overseeing the development of software platforms and products that collect and reuse our data.

Intercultural competence has cognitive, affective and behavioural dimensions. Its toolkit includes *knowledge* about other cultures, which enables a refined understanding of the nature of culture and a wariness of oversimplification. In addition, it fosters the development of personality traits such as empathy, openness and flexibility which are enablers of intercultural dialogue. Finally, it fosters an understanding of the subtlety of communication and a rigour towards the potential for miscommunication and misunderstanding, which one must assume is present in such encounters with inherent difference. It ensures that our hermeneutic powers, the cornerstone of humanistic investigation, do not get lulled into trusting false heuristics, mindful of Friedrich Schleiermacher's advice that 'misunderstanding occurs as a matter of course, and understanding must therefore be willed and sought at every point'.³⁷

Applying strategies developed to navigate intercultural interactions can provide a number of concrete supports to individual and collective responses to new and developing software platforms and tools, in particular ones that are facilitated by the gathering and use of big data. For example, the discussion of how unconscious biases have come to shape the products of big data is now well known and discussed in its cultural dimensions throughout this book. We can speak of these biases in a number of ways, such as the effects they have, the human rights they may infringe upon and the contributing factors that may cause them to be expressed in software, but none of these approaches has yet developed into a basis for strategies to both understand and counteract their incursion. It is instructive therefore to look at this problem in the light of some of the work on impediments to intercultural interaction, such as the framework developed by Brislin in the 1980s,³⁸ which included three parameters: prejudice, stereotypes and ethnocentrism. Although a more comprehensive mapping of these factors to the interaction between software creators and users will require more extensive study, evidence of all of these factors has already been seen above: prejudice (defined as a 'dislike based on a wrong and inflexible generalization') seems to well encompass the tensions between the 'coder' and the user, and ethnocentrism (defined as the 'tendency to view one's own culture as the only appropriate way of life') is also hugely

resonant with the tensions reflected in many of the accounts of conflicts between software developers and their management.

Perhaps most interesting of Brislin's three parameters, however, is that of stereotypes, which he defines as 'an exaggerated assumption about a certain category of people'. One can see this aspect realized perhaps most forcefully in the popular technology design tool known as a 'persona'. These snapshot biographies of potential users may have come into fashion as a way to 'help you step out of yourself'³⁹ as a technology designer, but the narrowing of user profiles down to the details of one particular (fictional) person's likes and dislikes, to the extent of including quotes in their imagined 'voice', largely promotes a reference to simplified stereotypes rather than the complexity of human experience. As an extension of the thinking of software culture, personas act as a way of promoting the cultural imperialism that forces the users to adapt to tools, rather than the other way around.

Intercultural competencies don't only give us a capacity to perceive and understand software development, they also give us strategies to respond to it differently. For example, Oberg's taxonomy of stages in adjusting to intercultural interaction⁴⁰ provides a very useful lens for observing some of the shifts in the relationship between users, their data and technology platforms. The first stage is known as the **Honeymoon Stage**: characterized by fascination and enthusiasm, and friendly but superficial relations between cultures. The many benefits of new technology, innovative uses for it and the habits that will predicate its later expansion are established in this stage. Before we had the Cambridge Analytica scandal, we saw Egypt's 'Facebook Revolution'; and before we saw publicly shared DNA records⁴¹ become a possible bounty for the surveillance assemblages of insurance companies and law enforcement, we saw a groundswell of public interest in what their genetic data could tell them about their health and their family histories. Regardless of whether the potential for abuse is always inherent in them, or emerges only as technologies become widespread and accepted enough for abuses of privacy and attention to emerge, it is clear that at this point in many cases we have segued in our relationship to the second phase of intercultural **Crisis**, in which culture shock arises as a result of differences in languages, values, etc., such as the many documented in this book, which lead to feelings of loss, rejection, frustration, anxiety or anger.

The open question is of course how the ecosystem of interactions between individuals and their data, and the companies and systems that use them, can make the next two shifts, first to **Recovery**, in which the crisis is gradually resolved, as the expatriate starts to learn how to act within the norms of the new

language and culture and finally **Adjustment**, in which the expatriate accepts cultural differences and even comes to appreciate them, although there may still be occasional instances of anxiety. We can see progress against these phases in the shift over time in how the conversation around ethics for big data and AI are developing, however. These phases (as documented by Kind)⁴² have moved discussions of AI ethics from initially being led by ethicists and philosophers, producing strong and informed responses, but expressed in a language and cultural code very foreign to those building software, to proposals consisting of technical fixes, emerging from the software development community in response to criticism of their output. What is emerging now in the ‘third wave’ of ethical frameworks to govern large technological systems is an approach with a much more hybrid set of actors, including regulators and the courts, citizen activists, NGOs and independent think tanks.

The work that underpins this book was designed in conversation with this third wave, but also aimed to address some of the gaps between the first and second ones, drawing from a deep understanding of culture to suggest ways in which technology builders and users might productively converge on a set of principles to guide a new generation of software design. To do so effectively, we must, however, be very clear, with ourselves and with our readers, about the biases and values we ourselves bring to this work. Each of us is steeped in both the appreciation of and the professional study of culture and its place in our lives. When we observe how the knowledge technologies we write of are deployed in society, we see them not only through the filter of what they are and do, but also of the creative memory- and identity-forming processes they are silent participants in. We are each personally and professionally committed to the translation between these values and their possible actuation in software.

While it is not clear if software culture has come close to realizing the vision of technologies being developed ‘by the people for the people’⁴³ viewing software development as culture allows us to read the conflicts between culture moving at a human’s pace and the much faster machine and corporate drives to datafication, as intercultural relations: after all, software systems and the people who build them do display distinct languages, norms, shared understandings, biases and ethics. This holistic approach will hopefully not only complement strategies to guide responses to developments in the digital society growing out of the social sciences, but also enable a paradigm allowing humanities methods and knowledge to take a leading role in the study of the digital in human cultures and lives.

Notes

- 1 Zuboff, Shoshana, *The Age of Surveillance Capitalism* (London: Profile Books, 2019).
- 2 Yakowitz Bambauer, Jane R., 'Tragedy of the Data Commons', *Harvard Journal of Law and Technology*, 25/1 (2011), 1–67.
- 3 Hertzmann, Aaron. 'Computers Do Not Make Art, People Do', *Communications of the ACM*, 63/5 (May 2020), 45–8.
- 4 Duportail, Judith, 'I Asked Tinder for My Data. It Sent Me 800 Pages of My Deepest, Darkest Secrets', *The Guardian* (2017), <http://www.theguardian.com/technology/2017/sep/26/tinder-personal-data-dating-app-messages-hacked-sold>.
- 5 European Commission, 2018.
- 6 Intelligence and Security Committee of Parliament, 'Intelligence and Security Committee of Parliament Annual Report 2016–2017' (2017), 123.
- 7 Subramanian, Samanth, 'Inside the Macedonian Fake-news Complex', *WIRED* (2017), <https://www.wired.com/2017/02/veles-macedonia-fake-news/>
- 8 Fortune 500 2000, <https://fortune.com/fortune500/2000/>.
- 9 The 100 largest companies in the world by market capitalization in 2020, <https://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-capitalization/>.
- 10 McLean, Paul, *Culture in Networks* (Cambridge: Wiley, 2017), 1.
- 11 Hofstede, Geert, 'Dimensionalizing Cultures: The Hofstede Model in Context', *Online Readings in Psychology and Culture*, 2/1 (1 December 2011).
- 12 Hardey, Mariann, *The Culture of Women in Tech: An Unsuitable Job for a Woman* (Emerald Publishing, 2019).
- 13 Ensmenger, Nathan L., *The Computer Boys Take Over Computers, Programmers, and the Politics of Technical Expertise* (Cambridge, MA: MIT Press), 15.
- 14 Cannon, William M. and Dallis K. Perry, 'A Vocational Interest Scale for Computer Programmers', in *Proceedings of the Fourth SIGCPR Conference on Computer Personnel Research*, SIGCPR '66 (New York: Association for Computing Machinery, 1966), 61–82.
- 15 Thompson, Clive, *Coders: Who They Are, What They Think and How They Are Changing Our World* (London: Picador, 2019), 77.
- 16 Frehill, Lisa, 'Satisfaction: Why Do People Give up on Engineering? Surveys of Men and Women Engineers Tell an Unexpected Story', *Mechanical Engineering*, 132/01 (2010), 38–41.
- 17 Bort, Julie, 'Some Tech Workers over 50 Are Literally Working Themselves to Death – and Other Things We Discovered about Their Careers', *Business Insider*, <https://www.businessinsider.com/stressful-lives-of-older-tech-workers-2015-11>.

- 18 Lyons, Dan, *Disrupted: Ludicrous Misadventures in the Tech Start-up Bubble* (London: Atlantic Books, 2017).
- 19 Thompson, *Coders*, 44.
- 20 Stevens, Barry, 'Probing the DP Psyche', *Computerworld* (21 July 1980), 26.
- 21 Quoted in Thompson, *Coders: The Making of a New Tribe and the Remaking of the World*, 68–9.
- 22 Kouw, Matthijs, Andrea Scharnhorst and Charles Heuvel, 'Exploring Uncertainty. Classifications, Simulations and Models of the World' (2013), 89–125, https://www.researchgate.net/publication/260438168_Exploring_Uncertainty_Classifications_Simulations_and_Models_of_the_World.
- 23 Thompson, *Coders*, 45.2.
- 24 Crawford, Matthew B., *The World beyond Your Head: On Becoming an Individual in an Age of Distraction*. Reprint edition. (New York: Farrar, Straus and Giroux, 2016).
- 25 Horsley, N., 'Intellectual Autonomy after Artificial Intelligence: The Future of Memory Institutions and Historical Research', in *Big Data – A New Medium?* (Abingdon: Routledge, 2020), 130–44.
- 26 Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is proprietary software that uses algorithms that are trade secrets to assess the risk of recidivism in criminal defendants and is used by courts in the USA to aid judges in sentencing. It has been shown to discriminate against black people.
- 27 Cited in Awati, Kailash and Simon Buckingham Shum, 'Big Data Metaphors We Live By' (14 May 2015), <https://towardsdatascience.com/big-data-metaphors-we-live-by-98d3fa44ebf8>.
- 28 Eyal, Nir, *Hooked Book – Product Design to Boost Customer Engagement* (Bingley: Portfolio Penguin, 2013).
- 29 Leslie, Ian, 'The Scientists Who Make Apps Addictive', *The Economist* (20 October 2016), <https://www.economist.com/1843/2016/10/20/the-scientists-who-make-apps-addictive>.
- 30 Stecklow, Steve, 'Why Facebook Is Losing the War on Hate Speech in Myanmar', *Reuters* (15 August 2018), <https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/>.
- 31 Google, Eric Schmidt, *Princeton Colloquium on Public & Int'l Affairs* (2009), https://www.youtube.com/watch?v=9nXmDxf7D_g.
- 32 Mark Zuckerberg, (10 May 2017), Facebook update.
- 33 Thompson, *Coders*, 24.
- 34 Simmons, Anjuan, 'Technology Colonialism', *Model View Culture* (2015), <https://modelviewculture.com/pieces/technology-colonialism>.
- 35 Coldewey, Devin, 'This Clever AI Hid Data from Its Creators to Cheat at Its Appointed Task', *TechCrunch*, (31 December 2018), <https://social.techcrunch.com/2018/12/31/this-clever-ai-hid-data-from-its-creators-to-cheat-at-its-appointed-task/>.

- 36 Quach, Katyanna, 'MIT Apologizes, Permanently Pulls Offline Huge Dataset That Taught AI Systems to Use Racist, Misogynistic Slurs,' *The Register*, https://www.theregister.com/2020/07/01/mit_dataset_removed/.
- 37 Kimmerle, Heinz, *Hermeneutics: The Handwritten Manuscripts* (Atlanta: Oxford University Press, 1986).
- 38 Brislin, R. W., *Cross-cultural Encounters* (New York: Pergamon Press, 1981).
- 39 Dam, Rikke Friis and Teo Yu Siang, 'Personas – A Simple Introduction,' *The Interaction Design Foundation* (2020), <https://www.interaction-design.org/literature/article/personas-why-and-how-you-should-use-them>.
- 40 Oberg Kalervo, 'Cultural Shock: Adjustment to New Cultural Environments,' *Practical Anthropology*, 7/4 (1960), 177–82. doi:10.1177/009182966000700405.
- 41 A concern most recently aired when it was announced that leading private DNA testing company 23andMe was planning to go public: <https://www.theguardian.com/technology/2021/feb/09/23andme-dna-privacy-richard-branson-genetics>.
- 42 Kind, Carly, 'The Term "Ethical AI" Is Finally Starting to Mean Something' *VentureBeat* (23 August 2020), <https://venturebeat.com/2020/08/23/the-term-ethical-ai-is-finally-starting-to-mean-something/>.
- 43 Kalluri, Pratyusha, 'Don't Ask If Artificial Intelligence Is Good or Fair, Ask How It Shifts Power,' *Nature*, 583/7815 (7 July 2020), 169.