The theory of bias and the ethics of technology design

This chapter focuses on Feenberg's development of a theory of bias in technology designs. The idea of bias is central to his overall project of developing a critical theory of technology, since it explains the entanglement of technology in issues of social power and domination. Feenberg argues that technology in modern societies is 'formally biased' and uses this idea to identify technology design as a field that is thoroughly political yet rarely recognised or theorised as such. The notion of formal bias establishes a space for critical and ethical concerns within technology and technology studies.

Like other critical theorists, Feenberg presents a philosophical argument whose viability rests crucially on philosophy's inclusion in its own discourse of insights from other scholarly disciplines, especially the study of politics and society. At the same time, the argument is intended to open a route for philosophy to 'speak' to other disciplines as well – in particular, to cast a critical light on contingent features of social reality that those disciplines study 'up close', so to speak.¹ In the case of critical theory of technology, if it is to be anything more than a scholastic reflection on technology's relation to society, Feenberg's philosophy must connect with disciplines that study and participate in technology design.

Feenberg presents the idea of bias as a way to clarify issues of social fairness at stake in technology design because it enables critical theory to engage with and mediate ideas from technical and other disciplines while decisively liberating the theory from its essentialist heritage. Drawing on social theory rather than philosophy, Feenberg aligns the 'neutral' appearance of technology with Max Weber's thesis that modern societies are characterised by 'rationalisation'. Technology is viewed both as an agent in that process and as being itself shaped by it, which is why it appears to be both rational (neutral) and a factor in alienation and the loss of meaning associated with cultural modernity.

In this chapter, I argue that the concept of bias is more useful when detached from the theme of societal rationalisation and deployed instead as the conceptual route towards a technology design ethics. Bias in technology design is better understood without invoking different modes of rationality, which are too broad to afford a secure grip on contemporary social phenomena. Instead, philosophical concerns can gain a foothold in technical politics by identifying normative principles immanent to design practice, understood as a special kind of communicative process.

Section 1 describes Feenberg's preferred account of modernity and its relationship to rationalisation. This section shows how, rejecting the option of an essentialist critique of technology, Feenberg clears the way for a strictly contextual understanding of bias in technology design. It presents his argument that in capitalist societies technology can both be formally neutral – that is, not obviously designed to be unfair – and yet function to produce and reinforce social injustices. Feenberg explains this, the formal bias of capitalist technology, with reference to the uniquely formal character of rationality in modern societies, which he says skews contemporary technology design. This section identifies a tension or discrepancy between Feenberg's emphasis on the problematic character of abstract 'formal reason' and his focus on concrete instances of bias in technology, of the kind that concerned Marx.

The next two sections present a detailed analysis of the concept of formal bias, concentrating particularly on Feenberg's presentation of the concept in his (2010) *Between Reason and Experience*. There he suggests that there may be several versions of formal bias and focuses on two of them, which he calls 'constitutive' and 'implementation' bias. Subjecting this argument to critical scrutiny, Section 2 argues that the distinction requires further clarification. In particular, there appears to be some overlap between constitutive formal bias and the kind of substantive design bias Feenberg says he rejects.

The discussion then moves to the test implied by Feenberg's discussion for the presence of formal bias in technology design. This turns on two features of the situation, namely (1) the nature of the intention through which technology comes to bear the impress of social determinations, and (2) the sociologically understood consequences of the design in practice. The formal bias of capitalist technology involves a distinctively 'neutral', or value-free, consciousness on the part of technologists that is focused on efficiency gains. Paradoxically, this is what generates biased designs that systematically favour one social group. Technology is formally biased only in context, and the outcomes generated by a technology design are contingent features that emerge when the technology is operational in a specific social setting. Of particular importance here are outcomes that favour the reproduction of what Feenberg calls the 'operational autonomy' of managers and owners in the economy.

Formal bias is present when neutral intentions at the scene of design are conjoined with the socially regressive enlargement of operational autonomy for favoured groups. Section 3 highlights the existence of cases that are problematic for this theory and clarifies the ambiguity concerning constitutive formal bias. The purpose of this argument is to rehabilitate a version of substantive bias but without opening a door to Feenberg's Heideggerian and conservative critics, whose substantivism consists in maintaining there is something *essentially malign* about modern technology. Instead, it becomes clear that substantive technical bias is a meaningful category because bias is a property of all technology, even though the precise nature of this bias can only ever be specified in social context. The point is that technology designs are biased – that is, they are substantively inclined to favour some interests over others – even when they do not fit either of Feenberg's definitions (constitutive or implementation) of formal bias. These technologies are sociologically significant in their effects, and their problematic nature arises from their technological character (they present issues that rarely arise in connection with other classes of object).

Feenberg's critics have misread his comments on substantivist critique of technology as reflecting an indecision on the question of the trans-historical nature of technology as something opposed to the human. His alignment of formal rationality with technical reason as a definitive property of modern technology reinforces this impression, but, nonetheless, the critics are wrong – Feenberg's true definition is thoroughly historical and relational. His point is that formal rationality biases technology design as a matter of contingent historical fact. The confusion, I submit, partially stems from the fact that this contingency has itself shifted, or ceased to obtain, so that technical bias is no longer well understood with reference to competing forms of rationality, especially not the opposition of technical or instrumental reason to communicative or other kinds. Feenberg's real point, which is that technology is always biased in context and not in virtue of some inhuman or anti-human essence, stands.

Here it is useful to supplement Feenberg's argument with ideas from post-phenomenological philosophies of technology, which emphasise the role of technical artefacts as agents or quasi-subjects rather than more or less inert objects. Feenberg's attempts to comprehend this are impeded by his framing of technology in terms of its association with a distinctive kind of societal rationality. Detached from a rationalisation-based historical perspective, technical politics becomes more messy and requires more diverse tactics than a battle between two opposed forms of reasoning. To illustrate this, I present examples of substantively biased technologies that ought to be subject to a strategy of containment – that is, considered unacceptable by critical theorists on terms that are immanent to design as a social practice. Section 4 considers Feenberg's

objections to containment of technology development and argues that they do not apply to these cases.

In concluding, I make the connection between desirable kinds of containment and Jürgen Habermas's idea of discourse ethics. Just as some discursive acts are ruled impermissible because they could not, in principle, be assented to by all participants in a given debate, so I argue that some technologies ought to be disbarred on similar grounds, and that this standpoint can be useful in developing the idea of design bias in the direction of an immanent ethics. The latter ought to form a supplement to Feenberg's project of democratic technical politics. A sociologically informed ethical critique at the scene of design is one of Feenberg's objectives, and clarifying the nature of technology's bias is an important step towards its fulfilment. The argument of this chapter suggests that his attachment to broad, historically grounded rationality concepts obscures relevant sociological details and inhibits Feenberg's successful pursuit of this goal, even though his introjection of social elements into the philosophical definition of technology makes it possible.

1 Varieties of bias

Feenberg views technology as a fundamental institution of modernity and as a fundamentally modern institution. Making the case for bringing modernity theory into discussions of technology, he asks rhetorically, 'how can one expect to understand modernity without an adequate account of the technological developments that make it possible, and how can one study specific technologies without a theory of the larger society in which they develop?' (in Misa *et al* 2003: 73). Technology is distinctive in modern societies because it bears the imprint of the capitalist and bureaucratic context associated with modernity. At the same time, technology is a catalyst or relay that amplifies some of the core tendencies of cultural modernity, so in this sense they work together. Technical systems work with causal propensities in the natural world by steering known physical regularities to achieve human ends, and, as such, they are inherently rule-bound. This imbrication with rules is particularly pronounced in modernity, and that is part cause, part effect, of technology's relationship to other institutions in modern societies.

Modern technology presents as neutral, objective and obvious, so that to reject it is to be perceived as foolish, anachronistic or worse. Similarly, the institutions of modernity present themselves as formally neutral in the sense that their behaviour is controlled by rules that have three defining properties. They are transparent, which means that anyone who accesses a given system can, in principle at least, find out what the rules are and use the system accordingly. They are universal in their application, so

that everyone is equally subject to the rules without exceptions based on social status or individual identity. Finally, the rules that constitute modern organisations are enforced in a manner that is separate from the issue of their purpose or that of the organisation of which they are part. This often creates the feeling that the rules get between people and whatever service they expect an organisation to provide. However, whether a modern institution works or not is a matter of its processes and the extent to which these are well ordered and efficient. The goals of the organisation are separate from the question of its performance.

The fact of being based on these kinds of rules is part of what differentiates modern from pre-modern institutions. In Britain before the reforms of the 1820s, for example, and for some time after them, law and its enforcement were inconsistent across the country. A local lord or bishop often had the authority to determine punishments and could vary them almost at will depending on their view of the plaintiff as well as their crime (Reiner 2000). The projects of standardising and publicising law are in large measure constitutive of the modernity of modern societies. Immanuel Kant placed particular importance on the publicity criterion as essential to legitimacy in law, writing that no 'actions affecting the rights of other human beings' could be valid if 'their maxim is not compatible with their being made public' (1992: 126). All modern political philosophy shares the conviction that systems are only fair if they apply equally to all and are objective in the sense entailed by the three-point definition of modern rule systems given above. Only such systems are untainted by direct, substantive sectional interests and, in this way, they achieve the appearance of neutrality. They are neutral and unbiased in the sense that they explicitly identify their correct functioning in terms of serving a *general* social interest.

Max Weber (1974) argued that the spread of such standardised, rule-governed systems constituted 'societal rationalization'. He exposed a kind of circularity at the heart of modern, rationalised social structures. Describing the 'spirit of modern capitalism', as, 'that attitude which seeks profit rationally and systematically', for example, Weber says that this 'attitude has on one hand found its most suitable expression in capitalistic enterprise, while on the other the enterprise has derived its most suitable motive force from the spirit of capitalism' (1974: 65). This relates to the third point above, about organisations' distance from the ends or values they are supposed to serve. As modern societies develop increasingly sophisticated methods for measuring performance independent of content, so their institutions tend to erode their own reasons for existing in the first place. Building on Weber, other thinkers in the critical theory tradition, most notably Adorno and Horkheimer (1997), argue that while ideology, especially religion, had served to legitimise domination in pre-modern societies, in modern, bureaucratically administered societies

this function increasingly passed to science and technology, where it was justified with reference to efficiency.

In *The Dialectic of Enlightenment* (1997), Adorno and Horkheimer suggest that in what they called the 'totally administered society', the contradictions of capitalism were cancelled by being managed through strategies of knowing manipulation that secure a culture of compliant consumerism. At the heart of this social system is identity-thinking – the principle that things can be grasped quantitatively and thereby brought under technical control. Modern science only sees the measurable world that it assumes is there in the first place, excluding from view nature's manifold qualitative aspects or considering them only as secondary. Modern culture extends this perspective to human beings and society, so that people are viewed as producers and/or consumers, their individual idiosyncrasies seen as problems to be handled by the system. This stark characterisation culminates in Marcuse's (1964) description of industrial modernity as a 'one-dimensional society'.

What Feenberg presents as 'modernity theory' is a synthesis of ideas from these and other sources concerned with the inner connection between increased differentiation of social capabilities achieved by modern institutions and the deleterious consequences of a socially diffuse emphasis on instrumental efficiency. In his reading of Weber, Feenberg contrasts the formal rationality characteristic of modern social systems with what he calls the substantive rationality of pre-modern societies:

Rationality is substantive to the extent that it realizes a specific value such as feeding a population, winning a war, or maintaining the social hierarchy. The 'formal' rationality of capitalism refers to those economic arrangements which optimize calculability and control. Formally rational systems lie under technical norms that have to do with the efficiency of means rather than the choice of ends. (1991: 68)

Feenberg views the question of how technology, which seems to be the embodiment of a neutral, scientific understanding of the world in tools that work for human ends, can be biased, through the prism of its relationship to this formal rationality. Viewed in this way, technology is entangled in the Weberian tendency of modern societies to erode cultural values in the pursuit of efficient procedures. Importantly, however, this is not inherent to technology but something it acquires under the unique cultural conditions of modernity. Feenberg writes:

The technical ideas combined in technology are neutral, but the study of any specific technology can trace in it the impress of a mesh of social determinations which preconstruct a whole domain of social activity aimed at definite goals. (1991: 81)

Feenberg here distinguishes between technology, the social context that shapes it and the consequences of that shaping in generating merely

technical definitions of social situations and problems. Technology is shaped by capitalism, and this is what makes it biased in favour of the class interests of the bourgeoisie. However, this bias is not one that can be grasped in terms of deliberate, intentionally discriminatory warping of the substantive design of technology. That would produce machines materially adapted to serve some users' interests while making life harder and less pleasant for others, and created because they had those effects. Rather, the bias of technology design rests upon its increasingly formal character. It is in its appearance as neutral, as the self-evident solution to a problem, that technology is in fact biased to serve some social interests rather than others.²

This might make it seem as if radical or democratic critique ought to reclaim a stake in what Feenberg calls substantive rationality and allow this to fashion a new technology or, indeed, to resist technology as the source of an alienating way of approaching the world. That is the preferred solution of thinkers like Albert Borgmann (1984), who denounces what he calls the modern 'device paradigm' as sustaining a false way of being and opposes to it other 'focal' cultural practices that enable people to resuscitate meaning and restore values to the centre of social existence. Borgmann writes, for example, of a 'historic decline in meaning' associated with the development of modern computing (1999: 15). A 'focal' practice is one that involves a relationship to some part of the world that is sufficiently rich and subtle that it definitively evades capture in any quantitative rubric (Borgmann 1984: 81). Borgmann gives the example of running, a practice that for him has meditative, even spiritual benefits.

However, as seen in the previous chapter, part of Feenberg's project is to distance critical theory from substantivist conceptions which maintain technology has a negative essence that might cause it to be biased in the sense that it always impacts negatively upon human culture, and, as a result, steers modern societies towards various kinds of catastrophe. He argues that substantivist theory tends to confuse the lack of values in modern technology design with the essence of technology as such (Feenberg 1991: 66). Feenberg wants to focus critique instead on the social factors that make technology biased, which also involves distancing his theory from the critique of science implicit in Adorno and Horkheimer's denunciation of identity-thinking.

What Feenberg rejects in the idea that technology might be substantively biased is the notion that the technological as such is held to be inimical to certain interests, cultural practices or social groups. Instead, he insists that all of technology's bias, all the ways in which it prefers some groups over others or implements contentious values, only arise when the impress of social determinations at the scene of design is conjoined to specific social contexts of implementation. Hence, he writes that when Jacques Ellul (1964) and Martin Heidegger (1987) present all technology

as incorrigibly instrumental, they confound ‘the essence of technology with the hegemonic code that shapes its contemporary forms’ (Feenberg 1991: 181).

Feenberg’s attitude to substantivist arguments like theirs, which focus on technical reason as itself inherently problematic, is circumspect and historical. At times he appears open to the idea that their analyses are persuasive representations of the negative impact of technology on modern society (e.g. Feenberg 2002: 19; 2010: 193; 2017: 69), writing that the ‘basic claims’ of Heidegger’s apocalyptic opposition of technology to culture ‘are all too believable’ (1991: 7). His clearly stated position, however, is that these arguments mistakenly impute something perennial to technology that is actually contingent on its design under capitalism (Feenberg 2002: 21; 2010: 194–196),³ and he repudiates the substantivist fallacy of ‘opposing spiritual values to technology’ (1991: 10). If such an opposition were integral there would be no possibility of technical politics, still less of the ‘common ground between critical theory and the scientific and technical professions’ (1991: 19) which, following Marcuse, Feenberg considers vital to it.

Feenberg’s theory of the formal bias of technology is developed with these debates in the background. He seeks a theory that can account for the real and perceived implication of technology in the most problematic dimensions of modernity while at the same time holding open the possibility of new designs that will combine technical elements to solve human problems in ways that are not possible under capitalism. Feenberg wants to combine a sense of the gravity of humanity’s current situation and of technology’s implication in it, which is normally associated with cultural conservatism and political pessimism, with the optimism implicit in his thesis that rationalisation can be democratised.⁴ In addition, the theory Feenberg elaborates has to have a degree of sociological precision when it comes to identifying how and why technology becomes problematic. This will free critical theory from its romantic, anti-technology past and enable it to acquire renewed political purchase, or as Feenberg puts it, critical theory will identify technology as, ‘not a destiny but a scene of struggle’ (Feenberg 1991: 14). Viewed in this way, the theory of formal bias is a main pillar of his project.

2 The theory of formal bias

Feenberg’s rejection of essentialism must be distinguished from his discussion of substantivism and substantive bias. In philosophy of technology these two terms (essentialism and substantivism) are closely associated, because it would seem that if technology is always and everywhere in some sense pernicious, as essentialism claims, then this must correspond to substantive properties it possesses – that is to say, properties

shared by all of its instances. Essentialists are committed to some kind of substantivism because they propound a non-relative property, or set of properties, as definitive of technology. However, substantivism need not entail essentialism. It is possible to maintain that technology always has some substantive impact and is therefore always ethically consequential (biased), without asserting that the character of this bias is always negative or regrettable.⁵

Feenberg associates substantivism with a kind of bias that might be found in social institutions or arrangements but which cannot exist in modern technological form. The pre-modern institutions referred to above might single out women or members of particular castes, for example, as unfit to own property or to have rights that other designated groups enjoy. Such social systems are designed by a biased intent. In *Between Reason and Experience* (2010), Feenberg argues that substantively biased reasoning like this simply could not enter the scene of technology design because were it to do so it would cause breakdowns and a loss of coherence in the resulting technology. He explains that 'substantively biased decisions in the technological realm, where cool rationality ought to prevail, lead to avoidable inefficiencies' (2010: 69).

Irrational hostility towards certain groups is grounded in feelings that would contaminate the kind of pure reflection on laws of motion of matter that is required to formulate a coherent technical design (2010: 163). Technical reason is irreducibly different from reasoning that includes a layer of feeling or prejudice. When Feenberg argues that technology is biased by contingent social factors specific to capitalism, he is not talking about prejudices or ideologies that issue from a ruling-class perspective on the world. Rather, the fact that technology requires a purely technical perspective leaves it peculiarly vulnerable to shaping by prevailing systems of thought and practice that are also unencumbered by values or meanings.

In place of substantive bias emanating from either irrational intentions at the scene of design or the purported negative essence of technology, Feenberg argues that modern technology is only ever formally biased. Formal bias 'prevails wherever the structure or context of rationalised systems or institutions favors a particular social group' (2010: 163). In other words, it occurs when a seemingly neutral system of rules is placed in a social context where it contributes to the systematic reproduction of unequal or unfair outcomes.

One possible point of confusion here concerns the role of the intentions that shape the technology. The point about intentions that produce formally biased design is that they are detached from their outcomes in the manner characteristic of modern, rationalised institutions. Technologists and others empowered to feed ideas into technology design do not often seek specific *social* benefits or prejudicial outcomes. Instead, their

intentions are narrowly focused on efficiency gains that will result from the new technology. As empowered agents in a capitalist economy, they understand the potential in new techniques in terms of productivity and performance, which they dispassionately associate with enhanced control for management at the point of production (Feenberg 2010: 70). The association of the goal of enhanced control over a process or practice with management interests is mediated by scientific and technical discourses on management and workplace design. Feenberg argues that there is a kind of innate convergence of such discourses with the similarly value- and meaning-free orientation of technical reason itself (2010: 185).⁶ These are the factors that shape technology design in modern capitalist economies, and the result is technology that enhances the operational autonomy of specific, privileged social groups (2010: 71).

Formally biased design is present where design shaped by this 'neutral' intention takes on a specific social function, namely that of unfairly enhancing what Feenberg calls the 'operational autonomy' (Feenberg 2002: 75–76)⁷ of privileged groups in the production process. Feenberg emphasises the systematic, sociological character of formal bias, arguing that 'to show discrimination in the case of a technological choice ... it is necessary to demonstrate that the discriminatory outcome is no accident but reproduces a relationship of domination' (1991: 181). It occurs when a system designed to maximise efficiency, placed in a social context, serves to generate, amplify or reinforce patterns of inequality and unfairness. Formal bias is only present in technology design when both aspects of this definition are found together. Technology is biased even though (in fact, because) the intentions that shape it are neutral and, despite being mere physical matter configured in a certain way,⁸ it operates to reproduce unfairness in a given social context (Feenberg 2002: 81–82).

Feenberg argues that formal bias has two forms, which he calls constitutive and implementation bias (Feenberg 2010: 164). In each case there is the determining intention just described and a social function defined in terms of the reproduction of social power relations. Constitutive formal bias occurs when 'values [are] embodied in the nature or design of a theoretical system or artifact' (Feenberg 2010: 163). Here the relevant test concerns whether it is conceivable that an artefact, placed in some other social context, might perform its function without systematically favouring the interest of a particular group over others. One of Feenberg's examples of such a system is the design of a production line which deskills workers and makes their lives unpleasant, while enhancing the operational autonomy of managers. Such a technology is constitutively biased because it is impossible to envisage a social situation where it might be used without having this effect. It is not substantively biased because it was not anyone's intention to design a system that would be unfair, or, rather, the design of the system was not motivated by a feeling

or sentiment of hatred towards or fear of workers. What distinguishes substantively biased technology from constitutively formally biased technology is the presence or absence of such an intention (Feenberg 2010: 163). In this case, the motivation for the design is the 'neutral' one of making production processes more efficient.

It is not clear, however, that constitutive bias really is distinct from some kind of substantive bias. As discussed in the previous chapter, Marx (1990: 562) wrote of technology designed by capitalists with the aim of disempowering workers, breaking up their associations and reducing their capacity to subvert employer domination. The prime motivating element there seems to be fear, though the practical implications include increased profits and a measure of efficiency, narrowly construed. This combination of motives and affects in an intention would efface the distinction from substantive bias that Feenberg seeks to maintain. What has changed since Marx's time, perhaps, is that this mix of motivations and beliefs about economic practices and the merits of machinery is now routinely parsed through social science disciplines, especially management science and business studies. Feenberg treats the latter as discourses in the Foucauldian sense, as sites of condensation of knowledge and power.

Implementation bias differs from constitutive formal bias in that alternative implementations of the same artefact or design are conceivable without the unfair outcomes. Here, Feenberg writes that 'values [are] realized through contextualizations' (2010: 165). In other words, it is only when it is placed in a *specific* social context that the technology takes on its formally biased character. This is the kind of unfairness that results from designing a system that works well in one, imaginary setting but generates unanticipated negative consequences when ported to another. In my view, implementation bias is a clearer specification of what Feenberg originally intended by formal bias. In his first formulation of the theory he wrote that '[t]he essence of formal bias is the prejudicial choice of the *time, place and manner of the introduction of a relatively neutral system*' (1991: 180; emphasis in original text). In contrast, constitutive formal bias is inscribed in the design of a system and travels with it across contexts, indicating that it is actually somewhat ambiguous in relation to the rejected category of substantive bias.

Feenberg clarifies the difference between substantive and formal bias with reference to the intention embodied in a design. What makes technical systems formally biased is the absence of human values in them, rather than the impress of malicious or prejudicial purposes. Formal bias occurs despite the absence of such intentions. Sociologically, it is what happens when the capitalist motivation to secure profits through improved efficiency and enhanced control over production processes converges with the purported emptiness, or freedom from prejudice,

of technical reason. As seen above, Feenberg suggests that this confluence is a key dimension of societal rationalisation in the Weberian sense. At the same time, however, his test for the presence of formal bias is a thought experiment that involves asking of any given technical system whether it might feasibly generate different outcomes under varied social and cultural conditions, which would demonstrate that it is the location of the system in a specific context that accounts for its role in producing prejudicial social consequences (Feenberg 2002: 81).

The outcomes Feenberg is most concerned with here relate to the enhanced operational autonomy of managers. Technology is biased in context when we can see that the interests of a particular group are systematically favoured over those of others involved with the technology. Most commonly, this involves managers gaining more control over production processes as a consequence of the application of scientific management and organisation principles. The latter are generic formulations that enhance efficiency, in a narrow sense, in proportion as they abstract the manager from production processes where they might have and feel obligations to the people they work with. In this way, formal bias mirrors societal rationalisation. As Feenberg puts it, 'formal bias prevails wherever the structure or context of rationalized systems or institutions favors a particular social group' (2010: 163).

3 Operational heteronomy

There are, then, two aspects to formal bias, and both are necessary, while neither is sufficient on its own, to produce it. First, there is the distinctive kind of intention that shapes a technology design, namely one that is free of concern with values and meanings and therefore dovetails with the kind of thinking that needs to happen if there is to be technical design as such. Second, formal bias only arises when, once set to work in a social context, the resulting artefacts systematically enlarge the operational autonomy of a social group.

The critical theory of technology is a normative theory, and its critical focus is on the biased nature of technology in capitalist society. It is worth emphasising, therefore, that one of the – perhaps surprising – consequences of the theory of formal bias is that it narrows the range of cases with which the critical theory of technology is concerned. Even technology designed with explicit nasty intent, for example, will be deemed irrelevant if it does not have the requisite social consequences. Technology with negative social consequences that are not the effect of design informed by the valueless evaluations characteristic of modernity will not feature in technical politics. Perhaps most importantly, technology that impacts negatively on specific groups in ways that are not

also advantageous to others by enhancing their operational autonomy will not be within the range of the theory.

It is also significant that the theory allows for a potentially quite large number of technologies that do not raise any normative issues. In principle at least, by insisting on the presence of a particular kind of intention in design and a specific kind of social outcome, Feenberg allows technology that might be considered 'neutral'. This seems to be unintentional, given his insistence that technology is socially shaped all the way down and his fundamentally agonistic conception of the social (considered further in the next chapter). Such technology might be comfortably accommodated in a pragmatic, Habermasian framework, within the normal development of a 'systems' sphere.⁹

Moreover, Feenberg rejects the possibility of designs that are motivated by or involve prejudicial intent, and insists that only what he calls 'cool rationality' (2010: 69) rules at the scene of design. I have indicated above that Marx's account of capitalist industrial machinery as designed to oppress workers would seem to be in tension with this requirement. Another example would be the printing technology described by Cynthia Cockburn (1983), which was designed with the malicious purpose of keeping women out of employment in nineteenth-century print factories. There, some machinery was kept deliberately and unnecessarily heavy as a result of agreements reached between unions and management. Since women were prohibited from working with machinery over a certain weight, better-paid work was thereby effectively reserved for male workers. In Feenberg's terms, the effect was to enlarge the operational autonomy of male workers, as well as managers and owners. In these examples, shaping intentions other than the one of cold neutrality appear to contribute to the regressive social outcome.

Feenberg addresses such cases through his category of constitutive formal bias, in which some groups of people are disadvantaged and excluded because designers fail to think about the variety of human beings who may be affected by or reliant upon the technology. In these cases bias is the result of neglect rather than malign intent, and Feenberg maintains that this kind of bias is most common in modern capitalism because of the focus on efficiency that dominates most thinking about technology and the narrow framing of the latter concept promulgated in disciplines like organisation theory and management science. He introduces the example of pavements that are too narrow to accommodate wheelchairs to elaborate on this (2010: 164). Another example might be the case of the Windows operating system, which disadvantaged blind and visually impaired computer users, putting many of them out of work (Goggin and Newell 2003).

Feenberg cites as an example of constitutive formal bias that is designed into the technology a case that is neatly analogous to Cockburn's

example – that of factory machines designed to be used by children (2010: 163). The factory machines could be used by small adults or even, uncomfortably, by adults of average height, so that in principle they are only biased in context. The real biasing factor is not the shaping of the technology but its role in a context that includes a web of restrictive rules and conventions. In Cockburn's study she points out that male workers and trades unions sought to preserve their wages, rather than to exclude women, which tends to support such an interpretation.

Interestingly, however, in these cases, while the requirement of a neutral shaping intention is conserved, the clearest social consequence of the technology is not to enhance the power and privileges of a group but rather to adversely impact on a particular group, perhaps a minority. This would be better described as *operational heteronomy* than enhanced operational autonomy. It is quite conceivable, then, that technology could present no discernible benefit to ruling groups yet still have exclusionary or oppressive consequences for others.¹⁰ In terms of Feenberg's theory, even when some technology negatively impacts certain groups more than others, perhaps obliging them to perform actions or tasks that others do not have to perform, it remains formally unbiased.

Feenberg's rejection of the category of substantive bias is, as was seen above, partly a result of his denial of any role for biased intentions in design, but it also follows from his insistence that bias is a contextual and not an essential property of technology. However, there are technology designs that are substantively biased in terms of his definition of the phrase and, in my opinion, they are too important for critical theory to overlook. An example of substantively biased technology in this specific sense is the mosquito device, which has been used in the UK to disperse groups of young people who are perceived as a nuisance by property owners.¹¹ The device emits a sound at a pitch that can only be heard by people under the age of 18. Changes in the human ear after that age mean that it becomes inaudible. The sound is, apparently, uncomfortable to those who can hear it and the device discourages young people from standing outside shops or on street corners where the owners do not desire their presence.

This technology was shaped by an intention that was at least in part motivated by negative feelings about a target group in the population. It worked with knowledge of something specific to them and identified that thing as a target. Even if this attitude was subordinate to the one of seeking a kind of control over social spaces, which might be cast as neutral, the intention to create a device that could only ever be experienced as uncomfortable by a specific group is surely prejudiced. The mosquito could be seen as enhancing the control of some groups of people, namely middle-class property owners, but a more exact description would be that it *limits and reduces* the autonomy of the targeted group. Moreover,

the mosquito is substantively biased in the sense that variations in context do not change the fact that it only has adverse effects on a selected group in society and is therefore inherently prejudicial. While it is debatable whether it actually empowers anyone except in the most tendentious sense (the shopkeeper has the power to remove some people from the environment around his or her premises, but appreciating this as a 'benefit' seems to involve participating in paranoid and prejudiced fantasies), it is certain that it impacts negatively on the freedom of a targeted group to exercise the basic human right of association in a given public space.

Other examples of substantively biased technologies come from military designs. For example, the flechette is a particularly unpleasant kind of tank shell that is designed to disperse 4cm metal darts, which enter and can remain in targeted bodies. Once there, the shards of metal may take days or longer to kill. Although the rationale for the design of the flechette is that it enables offensive action against enemies concealed in dense foliage, it is difficult to see how the weapon constitutes any kind of an advance on regular shells, which can be so powerful as to assure complete destruction of their targets without taking the additional step of torturing them as well. The flechette would seem to be a clear case of substantive bias in technology design. It is inherently inhuman in its conception and design and so seems to be biased 'all the way down'; its cruelty is not something that it acquires only when placed in context.

It would be anomalous if the critical theory of technology pushed these and similar cases beyond the scope of its concerns. They are problematic for the theory because the intentions in their design are not neutral in the sense required for formal bias and because their social effects are unfair even though they do not obviously enhance anyone's operational autonomy in the sense Feenberg gives to that phrase.

What this discussion highlights is a discrepancy between the philosophical cast of Feenberg's theory and its reliance on sociological factors. The distinctions between varieties of formal bias and the rejected notion of substantive bias begin to break down on contact with the kinds of cases studied by social historians. Cockburn's study of print technology designed to exclude women highlights an aspect of design that seemingly eludes the theory of formal bias because of the theory's exclusive focus on enhancements to operational autonomy. The mosquito demonstrates the possibility of a technology that is biased by a malign intent and socially regressive without enhancing anyone's operational autonomy. Feenberg's reluctance to treat such negative and exclusionary impacts in themselves, independent of the question of how they are thematised in discourses of power, reflects the political orientation of his theory, especially its concern to identify potential sources of political agency.¹²

However, as presented, the theory of formal bias threatens to abstract technical politics from the sociological analysis of the dynamics of exclusion and disempowerment.

4 Dilemmas of containment

The examples of substantive bias just presented indicate the need to consider the possibility that technology will entail restrictions on the development of certain kinds of technology, perhaps similar to moves taken internationally to abolish land mines in the 1990s. One of the things that a critical theory of technology ought to be able to say is that designs like the mosquito and the flechette are not legitimate. Feenberg rejects containment mainly because he views it as a conservative strategy, alien to a critical theory that sides with progressive rationalisation. If these cases are important, however, then this is an important oversight. The politics of technology design must include space for such a proscriptive element.

Feenberg gives four reasons for rejecting containment as a strategy, and it is worth devoting some attention to them because of what they disclose about the broader orientation of his theory. First, he says that technical politics should not be conceived in terms that cast culture as resistant to technological innovations and change but should instead be concerned with a design politics whose goal is to identify and cultivate technology's currently neglected potentials. Here Feenberg aligns critical theory of technology with the Enlightenment belief in a positive directionality to the historical process based on the development of reason and deepening human understanding of the world. This casts critique and the advance of knowledge as breaking down 'traditional' barriers and overcoming 'substantively' instituted regimes of social power based on irrational prejudices and privileges. Such a view favours technology development rather than its inhibition and comports with Feenberg's positive focus on politicising technology design rather than simply opposing modern technology.

Second, Feenberg maintains that a containment stance perpetuates the opposition of cultural values to technology, while a progressive position envisages their reconciliation. For critical theory of technology, it should be conceivable that social values and goals would be entirely consistent with those of technology development, which is why it is important to break with essentialism. Third, he argues that it is impossible to specify at this or any point in history which domains of social and cultural life should be 'protected' from technology, since change is pervasive and what seems permanent and important today can turn out to have been transient and weightless tomorrow, and vice versa. Finally, containment presents the paradox of an instrumentally conceived preservation of the

non-instrumental, which, Feenberg maintains, would have the perverse consequence of transforming the latter into a goal of technical action.

Feenberg rejects the idea of containment of technology by culture because he wants to distance the critical theory of technology from substantivist and conservative positions. However, he does acknowledge that there are occasions when containment of technology, in the form of an active proscription of some kinds of design, may be necessary. He writes that:

Objects introduced to technical networks bear the mark of the functionalization to which they have been submitted. Not everything of value can survive that transformation. Hence, we reject the idea that more or less technically efficient means can best accomplish things such as forming friendships and enjoying Christmas dinner. (2010: 76)

This way of construing matters is, it seems to me, back to front. It is not necessary to invoke special domains of cultural practice to see the necessity of limiting or prohibiting certain technologies. Doing so plays into the hands of those critics who suggest Feenberg prevaricates over essentialism.¹³ But the discussion in the previous section shows that inappropriate application of technical reasoning in human affairs is not the point at issue.

It is important to separate the idea that technology may be substantively biased from essentialist and conservative readings of that bias as located in technology's 'instrumental' character, which is held to contrast with other rationalities operative in special cultural domains, like the family or friendship relations. The point is not that such domains should be protected from technology in principle but rather that some technology is bad in itself and should be proscribed for that reason. Feenberg should be able to claim there are technology designs that are in conflict with the goals or rules of technology design, properly understood, in principle.

As Feenberg points out, technology only becomes threatening when its design reflects a particular disposition or cuts a certain kind of path through social relations. The way that it does this, when it becomes problematic, is specific to technology (other kinds of natural object and human artefact do not pose the same menace), but it is not caused by the technical reasoning involved in any technological project. It is not because of the reasoning behind technology that it can be disposed in this way but because of the specific entwinement of social purposes with physical properties in what Latour (2005) calls 'hybrid' combinations, which give rise to a distinctive kind of agency.¹⁴ The mosquito illustrates that it can also involve a distinctive kind of social outcome, which is discriminatory in a negative sense, rather than primarily serving a positive interest.

This kind of bias extends beyond formal bias as Feenberg has presented the idea, since it involves recognising that, as post-phenomenological philosopher of technology Peter-Paul Verbeek points out (2006), technical objects *act* in ways that help some people and disadvantage others. They nearly always do this, and as such are inherently or substantively biased. At the same time, this is not substantive bias in the traditional sense: the cases discussed in this chapter reveal nothing about the ‘essence of technology’, except in the sense that they pertain to a particular modality of the intrusion of physical nature or brute causality into human affairs. Such intrusion can be associated with regret, especially when technology becomes a systematic agent of operational heteronomy. But equally, even ‘good’ technology would not be useful if it did not make a substantive difference with social consequences.

Verbeek argues that the complex imbrication of human intention and the agency of objects at the scene of design necessitates an ethical approach immanent to technology design practice. He presents his view as radically at odds with Feenberg’s,¹⁵ but detaching the definition of formal bias in terms of intentions and outcomes from the wider question of societal rationality makes it usable as the basis for just such an immanent ethical foundation. Feenberg’s approach illuminates the range of intentions that might be present and suggests ways in which they might serve as conduits for unacknowledged social interests. Opening up the way in which an apparently narrow focus on ‘technical objectives’ might introduce foreseeable design bias is critical theory’s distinctive contribution. It is not necessary to relate technical reasoning to wider historical developments to achieve this. Moreover, if it were true that prejudice could play no part in technology design because it is grounded in irrational feelings, this would also be problematic for Feenberg’s project of a democratic transformation of technology, which surely presupposes a different mix of emotional, social and rational elements at the scene of design.

5 Immanent ethics of design

The emphasis, in the theory of formal bias, on a shaping intention makes it possible to articulate a formal principle that might realistically be operative within, or immanent to, the design process itself. Such an ethical principle might consist in something akin to the founding principle of Habermas’s discourse ethics, according to which ‘only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity as participants in a practical discourse’ (Habermas 1990: 93). To suggest that something like this might be operative in technical design contexts is not to commit the ‘fallacy of false concreteness,’¹⁶ but rather to claim that there are quasi-transcendental features of the

technology design situation that ought to constrain what may reasonably be proposed there. Violation of this pragmatic foundation would involve the violator in a performative contradiction, in much the same way that some forms of political speech conflict with the rational foundations of democratic politics.

Just as Habermas (1990) distinguishes the rules of his ideal speech situation from the substantive principles discussed by participants in moral discourse, so Feenberg's theory of bias puts him in a position to advance similar rules as immanent to technical practice. There is a benign, transformative orientation towards the future that underscores all technical thinking and activity, which Feenberg draws out most effectively when he discusses medicine (2010: 81), writing that 'the value of healing practices presides over biological knowledge of the human body in medicine' (2010: 81). As such, technical activity contains implicit norms that can only be violated at the price of contradiction with the technological purpose itself, however this is manifest in a given society. This focus on immanent ethical foundations aligns Feenberg's theory with Habermas's pragmatic notion of 'quasi-transcendental' principles.

As we have seen, Feenberg includes a social dimension to the philosophical definition of technology, bringing it into dialogue with other disciplines, especially sociology. As David Stump points out, Feenberg's inclusion of a social dimension in the philosophy of technology is the single, 'most powerful and innovative aspect of his philosophy' (in Veak 2006: 5). To capitalise fully on it, however, he needs to go further in enriching his philosophy with sociology. Instead of deepening his engagement in this way, Feenberg chooses to embed the theory of bias in a historical conflict of rationalities. Drawing on Weber, he argues that modernity is defined by a capacity for conceptual distancing (2010: 173) that distinguishes modern societies from traditional ones, where the rationality involved in technical projects has not been 'purified' in 'technical disciplines' (2010: 177). Feenberg aligns this development to societal rationalisation, arguing that 'modern societies are unique in the exorbitant role they assign social rationality' (2010: 179).

However, a wealth of historical scholarship casts doubt on the distinction of modern from traditional societies in terms of the form of their technical knowledge. Martin Bernal (1987) famously demonstrated, for instance, that ancient cultures possessed advanced mathematics, which they used to make astronomical forecasts. Great engineering projects of the ancient world indicate high levels of technical understanding. Similarly, Zaheer Baber (1996) describes inoculation programmes in pre-colonial India, which presuppose medical knowledge, and Feenberg himself refers to proto-industrial activities in the pre-capitalist world, which could not have occurred without detailed knowledge of the production process. Numerous other examples can be produced of a kind of

reasoning informed by abstract thinking in the centuries prior to modernity and in societies and cultures not normally included in it (e.g. Adas 1989; Baber 1996).

Feenberg emphasises (2010: 173) that he is not concerned to make a case for the cognitive superiority of modern societies; his focus is on the socio-cultural mediation of knowledge, and it is here that he considers they are distinctive in placing emphasis on the abstraction of function from meaning and codifying this into arcane disciplines that enforce social power. Detaching this argument from the notion of a wider societal rationalisation does not have to entail that modernity theory is abandoned altogether.¹⁷ Marx's characterisation (Marx and Engels 1967) of modernity as a period of unprecedented rapid change, in which commodities become increasingly salient, is an alternative reading that does not invoke the rather obscure idea of society-wide transformations underpinned by a novel form of 'rationality' but instead concentrates on class antagonisms and social contradictions as the drivers of technical change. Feenberg's critical focus on the convergence of (neutral) technical reason with broad historical developments tends itself to abstract from the deep entanglements of technology with practices of domination.¹⁸

At the same time, Feenberg's introjection of a specifically sociological dimension into the philosophical definition of technology makes it possible to advance the claims of critical theory in more detailed, situated kinds of enquiry. However, one of the hazards of incorporating a social dimension into the definition of technology is that society changes, forcing amendments to the theory. Feenberg's concern with the operational autonomy of managers is a case in point because the interface of management and workplace technology has significantly altered in the last three to four decades, associated with digitisation or informationalisation of technology and related new management practices. This is reflected in changes to the discourse on management and organisations, which, as seen above, was pivotal to Feenberg's grasp of the biasing of technology in design.

Critical analysis of management science texts was one of the foundations for Luc Boltanski and Eve Chiapello's pathbreaking study, *The New Spirit of Capitalism* (2005). Feenberg has been fairly dismissive of this work,¹⁹ but the challenge it presents to the category of operational autonomy is real. Increasingly, self-management, including a large degree of worker or user autonomy, is a presupposition of technology design.²⁰ Workers are now obliged to actively internalise system imperatives, converting them into personal norms of conduct. Meanwhile, contemporary work itself has to be seductive, appealing and exciting, leading some to characterise it as resembling an adventure, while others even write of a pervasive 'gamification' of the labour process.²¹ This restructuring of work is part of the way that capitalism recuperated itself after

the challenge it faced to its hegemony in the late 1960s and early 1970s. It entails new, 'streamlined' workers who use their autonomy to benefit the system because they have internalised its values (competitiveness, esteem for material success, etc.), incorporating them into their sense of themselves in a way that represents a new level of penetration of societal demands into individual psychology and interpersonal relationships.²² All this has implications for the theory of bias in technology design.

Feenberg's vision is partly inspired by these developments, in the sense that he prefers to think through the transformation of technology in terms of a widening of the cultural and normative inputs to the scene of design rather than advocating a defensive, conservative strategy aimed at limiting the development of new technologies. However, there is a tension between his declaration of the impossibility of substantive bias and his vision of a future technology shaped by wider values.

Democratising technology opens the way to designs (Feenberg sometimes calls them 'concretisations') that incorporate ethical and aesthetic values excluded from modern technology. This, he suggests, will create 'a new direction of technological progress' (2002: 150), which he calls 'progressive rationalisation'.

There is a kind of implicit, positive substantive bias in this vision of a society in which 'values would be installed in the technical disciplines themselves' (2010: 81). In the society of the future, technology that promotes equality, for example, will not be less subject to social distortion than contemporary technology. Rather, such technology would harness the agency of physical objects to benefit some, currently disadvantaged, social groups rather than others. Feenberg shies away from utopian speculation on these possibilities, preferring to identify interventions that challenge 'elite power structures inherited from the past in technically rational forms' (2010: 71). But what Boltanski and Chiapello's work demonstrates is that in many ways this technocratic form of social organisation is already largely superseded in informational capitalism, while the need for technical politics remains urgent.

In this chapter I have argued that in the theory of formal bias, Feenberg includes a sociological dimension in the philosophical definition of technology, while making a decisive move to free the critical-theoretic conception of technology from its essentialist, dystopian heritage. This opens up a perspective in which it is possible to clarify the combination of ruling intention and sociological outcome that characterises systematic and pervasive bias in technology design. To some extent these achievements are concealed by Feenberg's attachment to the transcendental thesis of a historically ambivalent societal rationalisation. As technology sheds its association with bureaucracy and top-down management processes in the real world, however, it becomes possible to

sharpen focus on an ethics immanent to technology design, an idea that is crucial to Feenberg's theory of technical politics.

Notes

- 1 Habermas writes that the task of philosophy today is to 'mediate interpretively between expert knowledge and an everyday practice in need of orientation' (1992: 17–18).
- 2 This is a very different line of attack to that of Marx, a difference Feenberg neglects to examine. As we saw in the previous chapter, Marx described a direct, intentional shaping of technology by capitalists, the substantive consequence of which was machines that were horrible to work with and limited workers' abilities to influence the production process or to build effective solidarity with one another.
- 3 It's important to note that this distinction is also present in Adorno, for whom technological rationality has a 'historical essence' (2000: 25). Defenders of Heidegger maintain that he also is not an essentialist, and Iain Thomson alleges that Feenberg 'simply us[es] essentialism as a descriptive term to characterize a fairly wide range of theories about technology with which he disagrees' (in Veak 2006: 65).
- 4 The sense of this claim will be explored in the next two chapters.
- 5 This conflicts with an intuitive sense that all bias is wrong because it involves systematic disadvantage to some individuals or groups, played out in ways that are not thematised and legitimated through public discourse. However, I would contend that all societies include bias in this sense in their fundamental structures and this is not always a matter of normative concern. A socialist society would institute measures to systematically prevent the materially better-off from gaining further advantage: indeed, if it failed to establish such bias it would not warrant the name 'socialist'. These arrangements would not be regrettable because they were not discussed on a regular basis.
- 6 It might seem as if technical reason is then an essential and indeed substantive property. Feenberg's proposal is that it is a *historical essence* in the sense discussed above – that is, it becomes an essential feature with the emergence of the modern social formation.
- 7 Feenberg takes this phrase from organisation theory (personal correspondence).
- 8 Latour argues that technology's behaviour in any situation is not susceptible, merely by dint of its material or substantive nature, to proper explanation by seemingly applicable scientific or causal laws. In his view, the reality in excess of science mournfully identified as essential yet unattainable by critical theorists is directly accessible now, in the sense that much of practical life already operates with a reality that ranges beyond what is sanctioned by the regime of epistemology, and always did (Latour 2013b: 70–71; 85). At one point he even observes that the 'material world' is simply not big enough to accommodate the multiple 'modes of existence' produced by actors (2013b: 103). Here as elsewhere, however, Latour neglects the practical entanglement of questions about the validity of knowledge with those concerning the operation of power.

- 9 Feenberg tends not to acknowledge this, perhaps because he has a polemical interest in charging other second- and third-generation critical theorists with neglecting technology. Once recognised, though, it becomes a matter of judgement how important the critique of technology is to the wider project of developing a critical theory of society. Notwithstanding this, detaching the critique of technology from global claims about the evils of societal rationalisation is a theoretical advance implicit in the theory of formal bias which, in my view, Feenberg does not fully capitalise on.
- 10 Such negative effects will remain out of view until they are successfully highlighted by the affected groups. Feenberg's neglect of this category turns out to be telling in an important respect, since it highlights the disjunct between the critical theory of technology as formulated and sociological analysis of the processes whereby injustice gets thematised by affected parties in the first place. Since the theory excludes technology that is not currently perceived as problematic, it seems to lack the resources to account for the struggle those subject to operational heteronomy face to articulate their concerns and gain recognition for them. I revisit this point in the next chapter, on technical politics.
- 11 The device has been banned in some European countries. See www.guardian.co.uk/society/2010/jun/20/teenager-repellent-mosquito-banned-europe. Accessed 19 June 2012.
- 12 This reproduces Marcuse's focus on the radical political potential of technical elites.
- 13 For example, David J. Stump writes that Feenberg 'keeps the general analytic framework that essentialism makes available while rejecting essentialism' (in Veak 2006: 8).
- 14 In other words, it is possible to say that there is something specifically technological about the way that some technologies are bad, without thereby claiming that there is badness in all technology.
- 15 In a review article Verbeek (2013) misrepresents Feenberg as the exponent of an unreconstructed modernism in which the sterile oppositions of traditional critical theory are all intact. I discuss the real differences between the two further in Chapter 5.
- 16 Habermas applies this phrase to arguments that purport to identify actual situations corresponding to his ideal speech situation.
- 17 As recommended by Latour (1993).
- 18 The identification of technology with the allegedly distinctive rationality of modernity is a bugbear of critical theory that limits its contemporary relevance – a point I return to in Chapter 4.
- 19 Curiously, Feenberg maintains (in Khatchatourov 2019) that studying management science literature, which is their principal methodology, is not a good way to understand contemporary labour processes and argues that the fundamental dynamics of workplace control have not shifted since the pioneering studies carried out in the 1950s and 1960s (Braverman 1974). It is worth noting that already in the 1960s Marcuse had observed that technology demanded that workers participate more in their own subjection as part of the modern industrial production process (Marcuse 1964: 30).

- 20 Boltanski and Chiapello overlook the issue of technology, neglecting to consider the ways in which the changes they describe to the labour process have been facilitated by new technology and have themselves shaped things like user interface design and the highly specific and limited use society has made of networked computing.
- 21 See Kirkpatrick (2015) for discussion of this and critical theory's response.
- 22 Sometimes grasped in terms of 'neo-liberal governmentality'; see e.g. Dardot and Laval (2014).