Governance by Technological Design, a Critique

João C. Magalhães[1]

This chapter develops a critique of the notion of *governance by design*, questioning whether approaches in this tradition provide a realistic perspective to make sense of and act upon our ever more complex technological systems. Given the seeming omnipresence of what is now often called simply 'tech' – from digital platforms to generative AI – even posing this problem might appear odd. Yet, whilst no-one can reasonably doubt the socio-political importance of materiality, approaches that see governance as mainly a function of how things are physically configured can lead to conclusions that are as alluring as they are simplifying, as I will argue. Even when these conclusions are firmly rooted in critical thinking and geared towards democratic values, they might still limit our understanding of how artifacts and power are associated, inadvertently directing attention away from the messy reality of their entanglements. If that is the case, the usefulness of ideas on governance by design might remain limited, regardless of their *de facto* impact among policymakers, researchers, and citizens.

My argument emerges from a critical reading of a few highly influential conceptualizations of the relationship between technology and social control. More specifically, I investigate these ideas as they moved from their post-structuralist foundations (Foucault, Winner, and Latour) to more recent work produced by legal scholars (Lessig and Yeung). This trajectory suggests a maturation of the thesis that we are politically affected by artifacts – but also the persistence of simplistic assumptions about technology, people, and their relations, assumptions that are often contradicted by actual evidence. This chapter at no point disputes the need to conceptualize and study how technology governs us, and is not, evidently, a sweeping criticism of this entire field – much less an analysis of state regulation. Rather, it should be taken as an invitation to keep questioning and denaturalizing those assumptions. Developing a full-blown alternative theorization of the role played by technological design in governance processes is beyond the scope of this critique. The conclusion, however, briefly suggests three problematic aspects that such new theories could address: the tendency to focus on isolated logics of control; a superficial view of human agency; and a lax (or insufficient) dialogue with empirical research.

The Power of Artifacts, in Five Acts

Governance by technological design is broadly defined here as encompassing the (i) processes whereby people are governed by things designed with the explicit goal of playing such regulatory role – and, then, (ii) how such processes are and should be themselves designed and governed. Put another way, governance by design encompasses the multiple dimensions of the entanglement of social control and the materiality of artifacts. As this section suggests, which "processes", "things", and "people" these are, and what, after all, it means "to govern" can vary considerably.[2]

This section discusses a small selection of seminal ideas that dealt with these problems. This paves the way for the next section, which formulates a more substantial critique of the thesis that we are governed by the things we create. Whilst this is not a systematic review, the works

[1] Centre for Media and Journalism Studies, University of Groningen, the Netherlands.

[2] The chapter employs "governance", "regulation" and related terms interchangeably to discuss how power operates.

I chose to discuss represent some of the most influential views on what governance by design is and ought to be. Authors like Foucault and Latour are now arguably part of the Western social thought canon; Winner and Lessig have played foundational roles in disciplines that are critical to works on power and technology – STS (Science and Technology Studies) and cyberlaw, respectively; and Yeung's writings, while more recent, have quickly become very popular, especially but not only among those who identify themselves as interested in "governance by design". The evolution of these ideas over time might not be exactly linear but it does suggest a distillation process, in which the role played by materiality becomes more salient and explicit, with its social consequences increasingly associated with individual behaviours, as I will argue.

Foucault

Current critical thinking on the power of and over technology have deep and largely forgotten theoretical roots in Marxist insights on the socially productive consequences of industrial machinery and on the Frankfurtian scepticism towards technocratic rationality (for a synthesis of these origins, see Feenberg, 2002). Yet the term "technology" in these works also orbited other concepts that were not described as primarily physical, such as ideology, class, and science.

It is in Foucault that we can arguably find the most influential antecedent of this approach. I mean, specifically, his idea of "panopticism". Given how well-known this concept is, it suffices to say that he coined this term to explain the "general principle" of a new form of social organization – what he names disciplinary (as opposed to sovereign) power, whereby control becomes self-realized as consequence of invisible surveillance (Foucault, 1977/1995, p. 208). Somewhat overlooked is Foucault's insight about the nature of the materiality-control complex. The most paradigmatic "architectural figure" of this new era of power, he argued, was the panopticon, a blueprint of a model prison invented by British philosopher and social reformer Jeremy Bentham in the late 18th century (Foucault, 1977/1995, p. 208). Foucault describes Bentham's project in some detail, noting how walls, windows, and a central tower, once arranged in a meticulously-thought spatial configuration, engender unequal visibility relations between inmates and guards, guaranteeing that the former would follow the orders established by the latter without the need of any direct intervention or physical punishment – a much more humane form of incarceration, in Bentham's view at least.

This "machine", as Foucault repeatedly calls the panopticon, would assure the "*automatic* [emphasis added] functioning of power", which would become so "perfect" that "its actual exercise" would be "unnecessary" (Foucault, 1977/1995, p. 201). That is, once control is transfigured into artifacts, it becomes independent from controllers' individual presence, motivations, and skills (Foucault, 1977/1995, p. 201). One chief reason the panopticon was incredibly consequential, he pointed out, was due to how it undermined the inherently uncertain processes of human meaning-making and interpretation, gaining a quasi-deterministic capacity to do what its inventor wanted: the production of "docile and useful" subjects (Foucault, 1977/1995, p. 231). This "physical power" was not scientifically sophisticated as, say, the steam engine – erecting this sort of building required no major engineering feat (Foucault, 1977/1995, p. 225). The panopticon was, first and foremost, an intellectual innovation.

Few would disagree with Foucault's argument on the centrality of surveillance as a logic of power, in particular to digital forms of social control. His insight proved to be so fantastically accurate that, almost 50 years after the publication of *Discipline and Punish*, it has almost

become a truism of contemporary life. Yet, as I note in the next section, his account was also wrong on several counts.

At any rate, the historical novelty of panopticism is importantly premised on the explicit design of objects to automate, depersonalize, and massify social control. Curiously, however, Foucault's discussion about materiality is hardly specific or overt – he seems to take for granted that walls, windows, and towers direct the human gaze, and in so doing necessarily revolutionize subjectivities.

Winner

This sort of meta discussion can be found in a classic piece of early STS scholarship, Langdon Winner's *Do Artifacts Have Politics?* (1987). Originally published in 1980, Winner's essay does not refer to Foucault. Readers can easily see a continuation between them, though, as if they were both responding to similar social transformations.

Winner was not interested in discussing whether "technical systems" were "deeply interwoven in the conditions of modern politics" – they clearly were. He discusses whether technologies can be said to have, "in themselves", "political properties" (Winner, 1987, p. 20).

He proposes two ways whereby this seems to happen. Firstly, features of artifacts are designed to impose the "dominion" of some groups "over others" (Winner, 1987, p. 24), establishing "patterns of power" (Winner, p. 38). A paradigmatic example was how an alleged racist urban planner (Robert Moses) supposedly designed some bridges in New York to be so low that they physically prevented buses with non-white and poor people from accessing a public park.

Secondly, some systems can have almost fixed politics: their adoption engenders relations with a "distinctive political cast" – centralized/decentralized, repressive/liberating etc (Winner, 1987, p. 29; see Mumford, 1964, for a precedent of this idea). The technical complexity of industrial production or nuclear power plants for instance, necessarily creates a strong hierarchical organization, he says. In the first type, localized changes in the design of artefacts (e.g., building higher bridges) can change their politics. In the second type, this is irrelevant (e.g., all and every cotton factory would demand centralization).

Winner surfaces and sheds light on aspects that are mostly implicit in Foucault. Namely, intentionality and determinism. The first form of inscribing politics into artifacts (e.g., New York Bridges) is all about designers' goals; the second (e.g., industrial factories) concerns mainly the inevitable consequences of materiality, which might or not be intended by designers.

If Foucault was mostly concerned with what happens to the individuals who are subjected to the power of things, Winner seems much more interested in the things themselves. Therefore, what, exactly, is regulated by technology, how such regulation unfolds in practice, and with what consequences, remains largely unclear and, as the next section shows, even factually contestable in his account.

Some of the general difficulties raised by Winner's argument on intentions and determinism were discussed in depth during the 1980s and 1990s, as STS developed into a scholarly field in its own right. Yet, STS scholars were seldom overtly interested in power per se – or, when they were, this preoccupation could be mostly descriptive.

Latour

An important example of this approach is Bruno Latour's *Where are the Missing Masses: The Sociology of a Few Mundane Artifacts* (1992). In it, he proposes that designers embed in the artifacts they create (doors, seat belts, speed bumps, meat rosters) bundles of often contradictory norms that the users of these things are expected to follow (Latour, 1992, p. 247). In this sense, technology materializes "prescriptions", a form of "programming language" (Latour, 1992, p. 232) which cannot, however, "prevent the inscribed user or reader from behaving differently from what was expected" (Latour, 1992, p. 237). And "prescriptions" are indissociable from "pre-inscriptions", or how designers imagine users will engage with their invention. Indeed, "no scene is prepared without a preconceived idea of what sort of actors will come to occupy the prescribed position" (Latour, 1992, 237).

This brief description enables us to see the continuities and discontinuities between his vision of governance by design and that of Foucault and Winner. Latour seems to be essentially unpacking, in his characteristic kaleidoscopic stream of details and examples, Foucault's original idea that things automate power, and that they do so, as Winner more explicitly problematized, because designers want to accomplish certain goals. But there are important differences. A subtle one is how the Foucauldian view of technology as able to produce entire subjectivities is replaced by a focus on individual, discrete, and ordinary *behaviours* – defined as patterned bodily movements such as driving, entering a room, eating; another is his insistence that these behaviours are always influenced but never determined by material prescriptions.

The perhaps most important novelty introduced by Latour's thinking (of which this essay is only one example) is his contention that "programs of action" which make up the social world are inherently hybrid: some of their "sections . . . are endowed to parts of humans, while other sections are entrusted to parts of nonhumans" (Latour, 1992, p. 254). Whereas Foucault is mainly concerned with the effects of particularly designed things onto people, and Winner was mostly interested in understanding the ways in which these things can be political in themselves, Latour fully embraces a 'flat ontology' in which there is no a priori hierarchy between stuff and individuals. And yet, despite his overt calls for complexifying social theory through the acknowledgement of the role of all the "mundane artifacts" in everyday life, critics have long understood that this complexification is misleading, a topic to which I return in the next section.

Foucault, Winner, and Latour had different views of the relations between materiality and social control. But their works are premised on a similar starting point: that power is utterly transformed when it ceases being purely symbolic (language) and enters (or is finally seen as necessarily belonging in) the realm of physics (artifacts). Control becomes easier to exert but potentially harder to resist. This is a foundational premise of contemporary works on governance by design, which are strongly nested in legal and regulation studies and much more interested in how technology's power ought to be tamed than the post-structuralist antecedents I have just commented on. Normative, and specifically democratic concerns, become more prominent.

Lessig

Along with Joel Reidenberg's paper on "lex informatica" (Reidenberg, 1998), Lawrence Lessig's *Code* (2000) is one of the indisputable pioneers in this body of work. This

book brought some of the same concerns that had been percolating through post-structuralism for decades to a very different audience – legal and regulation scholars and practitioners, particularly Anglo-American and Western European ones. To be clear, Lessig does not explicitly build on his predecessors. He appears to stride into this discussion as if it was an almost new intellectual territory. Foucault and Winner are quickly mentioned in footnotes, no attention whatsoever given to the details of their ideas; Latour and other key STS scholars are simply ignored. Yet, his work is far from being a mere repetition or rehashing of those authors' previous insights. Some of those ideas became central (the insistence on controlling discrete behaviours); others seemed to be taken as negligible truisms – that architecture cannot determine "any particular result" (Lessig, 2000, p. 359).

Arguably, Lessig's main contribution to theorizations of governance by technological design regards how he helped redirect concerns on this topic to digital communication systems and contextualized the role of materiality vis-à-vis other institutions – especially but not only, the law. Indeed, he argued, "regulation" of human behaviour depends on at least four "modalities", in his terms: law, market, social norms, and architecture (Lessig, 2000, p. 123). Law refers to traditional forms of regulation such as state statutes and policies. Market regulation refers to the actions of corporations and the use of economic incentives. Social norms regulation refers to cultural and societal expectations. The fourth modality of regulation, architecture, is created through the design of technology. Lessig argues that architecture has regulatory effects because it can limit or enable certain actions. While he discusses several "offline" examples, his focus is of course on the Internet – or the "cyberspace". And "life in cyberspace", he points out, "is regulated primarily through the code of cyberspace" – the programming languages that function as a form of digital architecture (Lessig, 2000, p. 83). For example, the design of a website determined by code can limit the types of content that users can upload, or a piece of code (say, a "cookie") can be used to gather data on user behaviour, which can be then used to regulate their actions.

Importantly, code is seen as a particular type of artifact because it is particularly malleable and, if controlled by private companies, also rather opaque (Lessig, 2000, pp. 127). In this way, the popularization of the Internet expanded the possibilities for both ever more stringent forms of control over users' behaviours and ever more dynamic forms of control over code itself. And a key "modality" to regulate code is precisely the law – either directly or indirectly, through the regulation of markets and norms. In his view, governments and companies could collude to exploit code and undermine not only Internet's potential but democracy broadly understood. Protecting users' autonomy, which he understood as a function of individuals ability to make informed choices, demanded designing transparent and open code. He argues that while corporations might lead the development of such technology, state oversight would be needed – what was far from obvious in the late 1990s, when neoliberal beliefs faced relatively little pushback.

Lessig's book immense influence strongly nuanced the then common tropes of the "cyberspace" as non-regulable environment (Barlow, 1996) and played a rather important role in advancing the idea that the Internet, particularly through its "code", should be regulated by democratic states. Some of his ideas went on to become almost common sense: the threat posed by large corporations to digital freedom, the opaque nature of code, the urgent need of transparency. Curiously, early conservative critics attacked Lessig for what, today, seems to be exactly what he predicted correctly – e.g., the catastrophic political dangers associated with Internet's profit-driven development (Cato Institute, 2009; Post, 2000). A much more

fundamental criticism concerns Lessig's subtle technological determinism, as I explain in the next section.

Yeung

As a rich literature on regulation by design developed in the wake of Lessig and other "cyberlawyers" (see e.g. Brownsword & Yeung, 2008), digital technologies changed drastically. Notably, few private organizations that generate huge profits from the constant surveillance and attempted manipulation of their users became the *de facto* controllers of much of the Internet. This chapter cannot do justice to all the works that tried to make sense of this transformation, from a governance perspective. The idea of *hypernudge*, as developed by Karen Yeung (2017), offers, however, an important and representative example of how earlier concerns around "code" mutated into concerns about "algorithms" and "data".

Yeung's central argument is that the regulatory power of code to "shape individual decision-making" has been drastically enhanced by the use of big data-driven AI systems controlled by "commercial Big Data barons" (Yeung, 2017, p. 119). Despite their complexity, "these applications ultimately rely on a deceptively simple design-based mechanism of influence: 'nudge' (Yeung, 2017, p. 119). The concept of "nudge" refers to the idea that, by altering the way choices are presented to individuals, often by changing the default option or simplifying complex information, designers/regulators can exploit cognitive shortcuts/biases, and subtly influence people's behaviour, nudging them in the direction they actually wish (Thaler & Sunstein, 2008). In this original theorization, "nudges" are – not so differently from the early websites Lessig discusses – fixed, based on the assumption that all individuals would react to stimuli in similar ways. Big data-driven technologies nudge differently, and much more powerfully, Yeung contends. At the heart of this *hyper*nudge, according to her, is a "recursive feedback loop", which silently "allows dynamic adjustment of both the standard-setting and behaviour modification phases of the regulatory cycle, enabling an individual's choice architecture to be continuously reconfigured in real time" (Yeung, 2017, p. 122). In other words, the rules embedded in these systems (e.g., those defining *which* exact online ad one should see) will constantly change in order to increase the likelihood that their users will indeed comply with broader rules that undergird the system's very existence (e.g., that which defines that users *should* see online ads). The hypernudge is much more efficient because it is radically – and automatically – personalised in a way that is both unobtrusive and aligned with designers' – not necessarily users' – goals. Yeung argues that this form of control is a much more potent version of Foucault's disciplinary power as it modulates users' actions in a way that appears to match people's actual intentions. Often inescapable, due to the market dominance of their developers, and deceptively designed to fit users' perception of free will, these systems challenge liberal remedies, such as notice and formal consent, she points out.

The idea that data-driven personalization is used by companies to manufacture profitable behaviours (from clicking on ads to ceding personal data) was not new (see e.g., Zuboff, 2015). Yeung work is particularly relevant because it offers a critical continuation of Lessig's original concerns with individual autonomy and democratic control vis-à-vis design decisions. As he, she sees in the particular materiality of advanced computer systems developed by for-profit organizations a unique threat to freedom. But she manages to persuasively demonstrate in which ways Lessig's normative expectation of informed choice ended up undermined by innovations that can only be explained by private economic goals and incentives. She thus avoids Lessig's overly liberal view of human agency and, in consequence, of the political

subject. But, as detailed below, this argument is also premised on similar simplifications of both technology and users.

Troubling Simplifications

I have so far described how the idea that power can be rendered physical through the design of artifacts emerged, changed and developed, travelling from thinkers as Foucault to scholars like Yeung. This section contends that, regardless of the acuity and originality of those conceptualizations, they also depend on some narrow assumptions about the nature of technology and our ability to design, understand, and negotiate it. This narrowness can be intuited by reading those texts, of course, but have also been evidenced by empirical evidence, which often contradicts those authors' suppositions. As it will hopefully become clear, what unites these conceptualizations is a certain disregard for the indeterminacy of social life, and eventually an urge to identify clear-cut and "solvable" design problems.

Foucault was never oblivious to the multiple dimensions of human agency but, at least in the first part of his career, seemed convinced that, as such agency became the very target of governance rationalities and techniques, it could indeed be subjugated. It is not that the inmates of the panopticon were unable to make critical sense of their condition but that the visibility regime created by the building's architectural configuration would, *by* and *in itself*, overrule and discipline these capabilities, leading them to act, believe, and feel as the designers of the panopticon wanted.

On paper, this might appear convincing. But its limitations become notable when one considers what happened when panopticons were actually built. Bentham never managed to convince British authorities to build a prison based on his plans, and the few panopticons that were constructed after his death hardly functioned according to his ideals, or in line with Foucault's interpretation of them. A relatively well-researched example is the Stateville Penitentiary, inaugurated in 1925 in the state of Illinois (US). Some issues were technical. Its explicit attempt to create a panopticon "did not function so ideally" as "the design which allowed the tower guard in each cell house to see into the cells also permitted the inmates to see when the guard's back was turned" (Jacobs, 1978, p. 428). Others problems related to how prisoners reacted to the architectural design of the building. For they could "cover their walls with blankets or cardboard in order to create a private space, free from the gaze of other prisoners and guard" (Alford, 2000, p. 131). Alford suggests that, in reality, guards did not really "care" to monitor prisoners: "Why should they? Their power depends not on supervising prisoners, but on controlling the entrances and exits" (Alford, 2000, p. 131). That is, at least in this case, surveillance was neither the only nor the most important governance logic at play. A broader view of the inmates as not deserving, in fact, so much attention from the prison administrators (as long as they remained inside the building) seemed more important. Furthermore, in his history of Stateville, Jacobs (1978) documents extensively how that "real" panopticon simply failed to produce the "docile and useful" subjects Foucault predicted. Riots, violence, and escapes were rather common. The architecture of the building was not irrelevant, however. For example, "as soon as the inmates" began to "shout to each other", the panopticon's walls acted "as an amplifier" and the place was "soon deafening", leading to insufferable levels of noise – a rather material consequence of a panopticon that Foucault did not predict (John Howard Association, 2010).

An argument can be made that buildings like Stateville are imperfect realizations of panopticons. But the messy reality of these prisons also raises the question of whether such

perfect "machine", in Foucault's words, could ever be implemented as Bentham imagined. Designers' intentions, realizations, and the ensuing consequences are seldom if ever aligned.

This can be seen in the subtle doubts that, confusingly, Winner introduces into his own argument. For instance, he concedes that the "patterns of power" created by machines might not be exactly what designers wanted. The example he uses is a kind of highly efficient and expensive tomato harvester which, he claims, increased economic concentration among agribusinesses in California in the 1970s. At the same time, there is no evidence that this was an intended outcome of the developers of the harvester. Similarly, not all systems will necessarily require particular forms of social organization. For instance, solar energy, with its disaggregated materiality, is only "compatible" with democratic governance – but might also be administered by a central manager, he suggests. To what extent do designers' intentions or the objective characteristics of things determine political consequences, then?

This question is made further opaque by the factual inconsistencies of Winners' account. Consider the example of the New York bridges. Against the story presented by Winner, there is little evidence that, even if Robert Moses ordered the construction of lower bridges for racist reasons (itself a point of contention), racial minorities and poorer people were indeed prevented from accessing said public park (on this, see Woolgar & Cooper, 1999). Public parks can be usually reached via multiple ways. Perhaps the hypothetical racist bridges could make access harder – but certainly not impossible.

The fact that such influential piece of writing was based on contested facts suggests another issue that pervades much of the work on regulation by design: a tendency to rely on anecdotes, often using secondary literature as its basis (in Winner's case, Moses' biography by Robert Caro), or the use of dramatic extrapolations (as Foucault's reading of Bentham's mere plans for a prison as paradigm of a new, all-encompassing form of power).

The simplifications that afflict Latour's view of material "prescriptions" are different. He was of course quite aware of the potential frictions between designers' goals and the reality of technology usage. Yet, as others have noted before (see e.g., Bowker, 2014, p. 1796), his theory of regulation is one in which categories that cannot be noted and described by simply "following the actors" are automatically discarded. His approach might more realistically portray the multiple elements that make people/things act as they do but it hardly illuminates the perennial (and perhaps exclusively human) moral reasons why we seek, resist, and study power in the first place: greed, oppression, suffering, accountability, reform. In that essay, "morality" becomes a matter of agnostically mapping and describing rules, while the opacity of human subjectivity is flattened in favour of observable behaviours taking place within a "network" whose topology is a direct consequence of researchers' own subjective choices.

Furthermore, it could be said that, for all his claims about the centrality of things for understanding the social world, the artifacts he uses to make his case are too rudimentary to be considered general proxies of "technology". E.g., it is wholly unclear whether we can study how very large computational networks govern their users in the same way we would study the morality of a door, to use Latour's example. In essence, his limited promise of complexification may actually shrink our ability to understand the ways in which technology governs.

While issues of justice reappear in Lessig's book, human agency and the intricacy of technology are hard to find in his strongly institutional perspective. I would argue that Lessig's provides a rather superficial view of what "the Internet" is, after all. Fundamentally, the Internet

has never been only about "code". It is an unfathomably complex network of software, hardware, corporate and governmental decisions, rules, and human labour – to name some of the most obvious ones. When he talks about the ways in which "code" regulates users' behaviour he seems to mean, in essence, the ways in which computational language creates certain end users' *interfaces*: the visual configurations of certain digital objects on screens and the functionalities that enact/are enacted by these interfaces, including intrusive software such as "cookies". In so doing, he ignores the vast array of moments in which "technology" shapes "behaviour" – of not only users but engineers, corporate leaders, policy makers, data workers etc. This narrow fascination with "code" (which is hardly exclusive to Lessig – see e.g. Lash, 2007) may preclude us from understanding the actual nature of the power relations created by such sprawling systems. Furthermore, noting how Lessig downplayed two decades of STS literature, Mayer-Schönberger argued that the book suffers from a sort of determinism, which can be summarised as a "view that society is shaped through a *linear* process of technological innovation in which society has no independent role; commercial actors innovate technology, which in turn impacts society" (Mayer-Schönberger, 2008, p. 739, emphasis added). Very little attention is given by Lessig to how code is actually developed and the ways in which users understand and engage with it.

Particularly telling is the idea that what code regulates is behaviour, and his disinterest in several other registers of human experience (affects, thoughts, beliefs) that might be the target of governance techniques. Given nudge theory's behaviouristic origins, this point is arguably even more important for Yeung's argument. She is of course critical of the ways in which Big Tech allegedly manipulates users. However, she does not challenge the underlying assumption of this concern – i.e., that such manipulation is indeed effective. In fact, her paper relies on the alleged power of big data-driven algorithms to justify the threat they pose. Yet critical data studies have repeatedly found a kind of human reflexivity towards these systems that is not consistent with the idea of an all-powerful and completely stealth manipulator. There is now considerable evidence that many users do know that *something* is actively trying to influence them. The popularization of terms such as "algorithm", "platform", and even "surveillance capitalism", even whe not accurate, is likely to further undermine the epistemological assumptions on which the hypernudge proposal hinges.[3]

Murky empirical evidence regarding the efficiency of nudging (broadly understood) reinforces these doubts. A meta-analysis of over 200 studies concluded that most (not all) "choice architecture interventions . . . successfully promote" behaviour change (Mertens et al., 2022, p. 1). However, when others corrected the same dataset for publication bias (i.e., researchers' tendency to report data that confirm their hypothesis and disregard those that do not), they discovered that those indicators of efficacy all but disappeared (Maier et al., 2022). Certainly, nudging might work – but this seems far from guaranteed.

Hypernudging remains a persuasive concept to describe what some tech companies intend to do when designing their products. Nevertheless, similar to Foucault's panopticism, it is wholly unclear whether this form of governance reliably produces the consequences it is designed to produce. Furthermore, even if hypernudging indeed led people to act against what appear to be their own interests, serious questions would remain regarding the meanings users associate

[3] This does not render manipulation attempts inconsequential or harmless to freedom – it is just that the exact nature of this harm might not be that which designers expected (see Magalhães and Yu, 2022).

with these control structures, and the potential for resistance, transgression, and reappropriation.

Conclusion

That relations of power are materialized into artifacts, and that these materialisations are designed by some to govern others, remains one of the most far-reaching insights in contemporary social theory. This chapter critically discussed the ways in which this idea was formulated in ground-breaking writings by Foucault, Winner, Latour, Lessig, and Yeung. In so doing, it underlined some well-known but unresolved issues – the gap between designers' intentions and the use of their products, what technology actually governs, what counts as technology, and the role of human agency in it. Simplifications abound, I argued.

In revisiting these problems, I tried to demonstrate that our theoretical vocabulary to consider the regulatory capabilities of technological design often fails to explain what exactly is going on. Real-life panopticons may not produce docile subjects, and might not even really be about surveillance; Moses' bridges might not have stopped black people from entering New York public parks (and perhaps were not designed to do so); the "morality" of technology certainly involves more than behavioural "prescriptions"; the Internet encompasses much more than "code", and maybe we are not so perfectly manipulated by "algorithms", however ruthlessly greedy and authoritarian their corporate controllers are.

We, critical researchers of technology, are constantly seeking mechanisms of control that are specific enough to be located, critiqued, and reformed. Yet we are repeatedly faced with the serpentine realities of how these mechanisms exist in the world, realities that are arguably becoming more complex as the technical systems we study become puzzlingly difficult to even describe accurately (e.g., Large Language Models). Perhaps that is inherent to the act of theorizing – that is, perhaps all theories are, at best, vague approximations of their objects. The crux, as I see it, is whether the approximations I identified above are the best we could do.

I would argue that they are not, and would posit that advancing the theorization on this topic could involve addressing at least three aspects that might limit our ability to understand how technological design governs us. These aspects might apply to a variety of sub-fields, from works on choice architecture to research on how to embed moral values in technology.

The first is the very expectation that this power follows a coherent, single regulatory logic – "discipline", "prescription", "hypernudge" etc. This might make sense when "governance" is reduced to isolated processes that are directly mediated by seemingly uncomplicated artifacts – how walls enable guards to see inmates, whether buses can pass under bridges, using a seat belt, choosing what to click on an interface. Yet, as soon as the analyst zooms out from these somewhat simple processes, that logic might prove to be just a small piece of much larger structures – the state, racism, public health, corporate datafication. And these structures are mediated directly or not by many more individuals, things, and sociomaterial relations than researchers can examine in detail. Using that relatively simple process as a paradigmatic instantiation of such larger systems is seductive and, often, a powerful rhetorical device. For instance, Foucault's grandiloquent claim that a core characteristic of "society" was encrypted in a text about how to build a penitentiary remains fascinating. Indeed, a synecdoche is attractive *precisely* because of its neat, illuminating narrative. But these narratives can only take us so far, as the messy reality of the panopticons attests. An alternative approach should start from the assumption that specific control logics are not, necessarily, the most important,

much less the only ones that matter. In this sense, one of our chief tasks is to map out these multiple logics, their associations and hierarchies, which might be themselves contextual and unstable. In so doing, we might arrive at more realistic accounts of how broader regimes of power operate through multiple, unequal, and perhaps conflicting instruments of governance, of which "design" is most likely to be just one among many others – such as discourse, explicit coercion, and cultural norms.

Secondly, those who study technology (even someone like Latour) are often drawn to this topic due to an irremediable fascination with artifacts – not with people (see Wyatt, 2008). Asserting to reject any form of determinism or pledging to attend to human agency (as Lessig does) is much harder than accounting for the sprawling consequences of actually doing so. It is extremely hard, for instance, to both care about such agency and focus solely on how artifacts supposedly create bodily actions that can be observed and quantified – "behaviours". Investigating exclusively what designers do and want to do will most likely yield conclusions that are, more than merely incomplete, distorted and potentially misleading. Taking seriously the idea that power is *inherently* situated in the crossroads of agency and structure involves at least two things. Firstly, being explicitly cautious about the limited insights one can gather from studying only the design of technology. Secondly, and more fundamentally, this approach entails shifting the very locus of concern: from things to the relations between things and people. This would necessarily demand listening to what individuals have to say about what they do, think, and feel about the artifacts built to govern them.

Thirdly, one needs not blindly embrace empiricism to comprehend the centrality of rigorous empirical research for advancing this field. The literature I analysed above is overflowing with questionable hypotheses and convenient case studies, which, given the right conditions, might become influential before their sweeping propositions can be thoroughly assessed. In addition to the examples I discussed above, consider e.g., how resilient has become the Lessigian idea that social media platforms create informational "filter bubbles", despite a growing number of studies demonstrating the precarity of that hypothesis (for an extended comment on this, see Bruns, 2019). In its first conception, over ten years ago, the idea of "bubble" served to channel a collective but then formless dissatisfaction with digital platforms. It was quickly taken up as a self-evident reality, directing our scarce attention to a problem that, in fact, does not appear to be nearly as common or as important as many seemed to believe. This is not to say that research must be exclusively empirical but that scholarship should be driven by a strong and productive scepticism towards ideas that are either empirically controversial or have not been empirically evaluated at all. Theoretical choices and methodological procedures should be highlighted and accounted for as entailing real limitations, not as forgettable formalities. This might sound obvious but it can also be easily neglected.

This critique by no means suggests that studying governance by technological design is unnecessary or an esoteric endeavour, the exclusive task of any particular discipline. Consider for instance Natasha Dow Schüll's (2012) multi-year, interdisciplinary study of how discourses, places, machines, and subjectivities co-construct – in highly uneven ways – gambling addiction, a story of blatant exploitation that avoids the moralizing temptation of resorting to schematic conclusions about manipulation. Her work is an example of how, by empirically acknowledging the multiple forms of "design" and the multiple fora of "governance", researchers can construct truly realistic accounts of how power works.

References

Alford, C. F. (2000). What Would it Matter if Everything Foucault Said About Prison Were Wrong? Discipline and Punish after Twenty Years. *Theory and Society*, 29*(1)*, 125-146.

Barlow, J.P (1996). *A Declaration of Independence of Cyberspace*. https://www.eff.org/nl/cyberspace-independence

Bowker, G. (2014). The Theory/Data Thing. *International Journal of Communication* 8 (2043), 1795–1799. https://ijoc.org/index.php/ijoc/article/view/2190/1156

Brownsword, R., & Yeung, K. (Eds.). (2008). Regulating technologies: legal futures, regulatory frames and technological fixes. Bloomsbury Publishing.

Bruns, A. (2019). *Are Filter Bubbles Real?* Cambridge: Polity Press.

Cato Institute (2009). *Ten Years of Code: A Reassessment of Lawrence Lessig's Code and Other Laws of Cyberspace*. https://www.cato-unbound.org/issues/may-2009/ten-years-code-reasessment-lawrence-lessigs-code-other-laws-cyberspace/

Feenberg, A. (2002). *Transforming Technology: A Critical Theory Revisited*. Oxford: Oxford University Press.

Foucault, M. (1995). *Discipline and Punish: The Birth of the Prison*. New York: Vintage Books. (Original work published in English in 1977)

Jacobs, J.B. (1978). *Stateville: The Penitentiary in Mass Society*. Chicago, IL: University of Chicago Press.

John Howard Association (2010). *Monitoring Tour of Stateville Correctional Center*. https://static1.squarespace.com/static/5beab48285ede1f7e8102102/t/5d03e3901e07180001db0bac/1560535952383/Stateville+Report+2010.pdf

Latour, B. (1992). Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts. In Bijker, W.E., & Law, J., *In Shaping Technology/Building Society,* 225–258. Cambridge: MIT Press.

Lash, S. (2007). Power after Hegemony: Cultural Studies in Mutation. *Theory, Culture & Society*, 24*(3)*, 55-78.

Schüll, N. D. (2012). *Addition by Design: Machine Gambling in Las Vegas*. Princeton: Princeton University Press.

Lessig, L. (2000). *Code and other Laws of Cyberspace*. New York: Vintage.

Magalhães, J.C., & Yu, J. (2022). Social media, social unfreedom. *Communications: The European Journal of Communication Research*, *47*(4). https://www.degruyter.com/document/doi/10.1515/commun-2022-0040/html

Maier, M., Bartoš, F., Stanley, T. D., Shanks, D. R., Harris, A. J., & Wagenmakers, E. J. (2022). No Evidence for Nudging after Adjusting for Publication Bias. *Proceedings of*

the National Academy of Sciences, 119(31), e2200300119. https://doi.org/10.1073/pnas.2200300119

Mayer-Schönberger, V. (2008). Demystifying Lessig. *Wisconsin Law Review*, 4, 713-746.

Mertens, S., Herberz, M., Hahnel, U. J., & Brosch, T. (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences, 119*(1), e2107346118.

Mumford, L. (1964). Authoritarian and Democratic Technics. *Technology and Culture,* 5, 1-8.

Post, D. (2000). What Larry Doesn't Get: Code, Law, and Liberty in Cyberspace. *Stanford Law Review,* 52*(5)*, 1439-1459.

Reidenberg, J. R. (1998). Lex Informatica: The Formulation of Information Policy Rules through Technology. *Texas Law Review*, 76(3), 553-593. https://ir.lawnet.fordham.edu/faculty_scholarship/42

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

Winner, L. (1986). *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. Chicago: University of Chicago Press.

Woolgar, S, and Cooper, G. (1999). Do Artefacts Have Ambivalence? Moses' Bridges, Winner's Bridges and Other Urban Legends in S&TS. *Social Studies of Science,* 29*(3)*, 433–49.

Wyatt, S. (2008). Technological Determinism is Dead; Long Live Technological Determinism. In Hackett, E., Amsterdamska, O., Lynch, & M, Wajcman, J., *The Handbook of Science and Technology Studies* (3rd edition), 165-180. Cambridge, MA: MIT Press.

Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, *20*(1), 118-136.

Zuboff, S. (2015). Big other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, 30(1), 75–89. https://doi.org/10.1057/jit.2015.5