

## Why a Commitment to Pluralism Should Limit How Humanity Is Re-Engineered

*Brett Frischmann and Evan Selinger\**

What does it mean to be human? What matters about being human? Can uses of technology dehumanize, and if so, how? What moral values should guide normative evaluations of the networked digital world that we're building for ourselves, our children and future generations? These are the highly-contested questions we tackle in our recent book, *Re-Engineering Humanity*.

Our contribution here is to provide three conceptual tools to help people grapple with these difficult questions. First, we discuss *humanity*, distinguishing between what it means to be human and what matters about being human. We argue that humanity is a shared resource consisting of intergenerational ideals and commitments. Second, we discuss Robert Nozick's famous thought experiment: *the Experience Machine*. The scenario raises fundamental questions about the good life and does so from an individualistic perspective. Finally, we discuss our thought experiment: *the Experience Machine n.o.* This scenario also raises fundamental questions about the good life, but it does so from an interconnected social perspective that emphasizes how world building engineers humanity by shaping the possible lives of others, including future generations.

We present this analysis to help readers make better sense of their own lives and the supposedly smart techno-social systems that exert ever-greater influence on how we think, act, and relate to one another and our environment. The main position we defend is that (i) different views of the good life and their underlying conceptions of what it means to be human are justifiable and (ii) a commitment to pluralism requires building worlds that allow people to pursue diverse paths towards their conceptions of flourishing. Endorsing (ii) requires being committed ethically and politically to two related ideals for freedom: freedom to be off (in the "not always on" sense) and freedom from engineered determinism. Each ideal entails positive and negative liberties<sup>1</sup> that are

\* This chapter is adapted from "To What End?," a chapter in our book, *Re-Engineering Humanity* (2018). For brevity, we include select references. For more, please consult the book.

<sup>1</sup> "Negative liberty is the absence of obstacles, barriers or constraints. One has negative liberty to the extent that actions are available to one in this negative sense. Positive liberty is the possibility of acting – or the fact of acting – in such a way as to take control of one's life and realize one's fundamental purposes." Carter

contingent on the dynamic relationships among individuals and between individuals and techno-social environments. We're asking for degrees of freedom engineered into our lived-in environments and thus our lives.

## INTRODUCTION

Life can be understood as a role-playing game with the earth being our experience machine. For better or worse, humans haven't optimized the planet to give us all of the experiences we desire. But this situation can change. Our roles, desires, and capabilities are engineered and re-engineered, more than we typically appreciate.

Let's unpack the gaming metaphor. Popular role-playing games allow players to create their own characters. In *Dungeons and Dragons*, players roll dice and assign points to different *attributes*, such as Strength, Intelligence, Wisdom, Dexterity, Constitution, and Charisma. They also select their *race*. If players want to stick with what they know, they can remain Human. Or, if they're feeling more adventurous, they can try being an Elf or a Dwarf. That's not all. Players also get to choose their *alignment*, such as Good, Neutral, or Evil. And they decide on a *class*. Fighters, Clerics, Mages, and Thieves are all popular options.

Many video games follow similar procedures. Sometimes, players are given the option of using a pre-generated character that allocates a fixed number of points to a variety of attributes. Sometimes, they can create their own with a character generation screen. In this case, players choose how to spread points among the attributes. In both *Dungeons and Dragons* and similar video games, players can improve their attributes, usually by performing tasks and gaining experience in the game. The built environment – the rules, story lines, maps, algorithms, and so much more – constrains and thus shapes players' lives.

These fictional narratives tell us something important about how we see and, more importantly, how we imagine ourselves as human beings. Humanity is represented as a collection of basic characteristics that each person starts with. You might believe a divine spark pre-generates your "character." Or you might believe that evolutionary biology plays this role. You could believe both of these things, or something else altogether, since you are largely a product of social conditioning. Your cultural background, your religious affiliation (or lack thereof), your political orientation, and all of the other features that make you who you are, influence which characteristics you value. You might believe that some characteristics are more

(2016); Berlin (1969); Berlin (1978). "[I]n the first case liberty seems to be a mere absence of something (i.e. of obstacles, barriers, constraints or interference from others), whereas in the second case it seems to require the presence of something (i.e. of control, self-mastery, self-determination or self-realization)." Carter (2016). See, generally, Berlin (1969), pp. 121–22. (Negative liberty is relevant when one answers the following question: "What is the area within which the subject—a person or group of persons—is or should be left to do or be what he is able to do or be, without interference by other persons?" Positive liberty is relevant when one answers the following question: "What, or who, is the source of control or interference that can determine someone to do, or be, this rather than that?")

essential than others. If given the opportunity, perhaps you'd allocate more "points" to autonomy than sociality, or maybe it would be the other way around.

Regardless of your perspective on how human "characters" are generated and the relative weight of various characteristics that they can be constituted by, the gaming analogy usefully emphasizes the importance of identifying basic characteristics that constrain, define, and inspire us to be more, and recognizing how the built environment shapes them. Crucially, we have opportunities to shape ourselves, and these opportunities emerge in and through our built worlds.

In *Re-Engineering Humanity*, we closely examine these relationships. How do we engineer ourselves through the techno-social world we build? How does techno-social engineering impact important aspects of our humanity? We do not attempt to identify everything important about being human. Instead, we provide a framework for identifying and evaluating techno-social engineering of core capabilities that distinguish humans from an idealized construct that we call "simple machines." Other baselines are possible, for example, other species. We leave consideration of alternatives for future work. For now, keep in mind that simple machines are programmed and fully predictable. Humans generally are not. If humans consistently behave like simple machines, however, we might need to evaluate the techno-social environment.

We examine some core human capabilities in depth, including commonsensical thinking, sociality, and free will. We touch lightly on others, such as how our sensory capabilities mediate how we look at and relate to others as well as the physical environment. In an extended discussion of transformative tools, we highlight the power of imagination, language, and the collective construction of shared ideals, institutions, and reality. These core capabilities and ideals – and thus humanity – may be at risk as supposedly smart techno-social systems spread.

A world in which engineered determinism governs is a world in which fully predictable and programmable people *perform* rather than live their lives. Such a world would be tragic. People living in it could be described as human and still would qualify as *homo sapiens*. Nonetheless, they would have a thin normative status as human beings because much of what matters about being human would be lost.

#### BEING HUMAN AND WHAT MATTERS ABOUT BEING HUMAN

Many criticize the negative impact of technology for being dehumanizing, especially in recent decades with the widespread adoption of computers, the Internet, and, more recently, smartphones. It is difficult, however, to know when a line has been crossed, when the techno-social engineering has gone too far, when something meaningful has been lost. Do we know when technology replaces or diminishes our humanity? Can we detect when this happens? To begin to answer these questions, we would have to know what constitutes our humanity, what makes us human, and what matters about being human. We need to understand and appreciate our

humanity if we are to preserve, protect, and sustain it for ourselves, our children and future generations.

Many have said that humanity can be taken away through slavery as well as its authoritarian political analogues, like totalitarianism. Psychologists label certain practices, such as confining people to pointless tasks and subjecting them to deindividuation, dehumanizing. On this view, humanity can be lost, fully or in increments, partial deprivations, or deteriorations.

Others, however, challenge the notion that one's humanity can ever be lost or taken away. They argue that one's humanity persists even when it is not acknowledged or respected. The slave is and always will be human, and thus, her humanity cannot be taken away. What those who support slavery do is fail to acknowledge and respect her humanity. While this is a reasonable perspective, we don't adopt it; the perspective doesn't adequately distinguish (i) being human from (ii) having access to, possessing, and sharing in humanity. We agree that the slave is and always will be human, but we don't think that means her humanity cannot be taken away. To elaborate, we'll say more about different conceptions of being human (descriptive) and what matters about being human (normative).

But first let us be clear: We are committed to what, over time, has come to be known as the Kantian rule on human dignity: "All human beings are worthy of respect and deserve to never be treated exclusively as a means to an end." Some might argue that our suggestion that slavery deprives the slave of her humanity means that the slave is no longer a human being and consequently no longer worthy of respect. Similarly, they might argue that since a newborn baby lacks various capabilities that we (elsewhere) identify as potentially essential components of humanity, the baby would not be a full-fledged human being worthy of respect. These arguments fundamentally misconstrue our approach. The slave and the newborn are and always will be human beings worthy of respect. The Kantian rule applies universally to all human beings, regardless of whether they have access to, possess, and/or share fully in the blessings of humanity.<sup>2</sup>

It's easy to believe that the meaning of "humanity" is simple, intuitive, and unproblematic. You know it when you see it. At least, it seems that way when we look in the mirror or talk to our children. We are human, and so humanity must be what we are and those who are like us are. Unfortunately, this commonsense view lulls us into believing that what matters about being human – what's special and

<sup>2</sup> The eighteenth-century philosopher Immanuel Kant insisted that this rule is a universal truth, a categorical imperative dictated by an inescapable moral logic. We maintain that the rule is the product of human imagination and collective recognition, just like Hammurabi's Code and the Declaration of Independence. The Universal Declaration of Human Rights enshrines the rule, recognizing "the inherent dignity and . . . the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world." United Nations – *Universal Declaration of Human Rights*. According to Article 1, "All human beings are born free and equal in dignity and rights." As many have noted, the Declaration reflects an incredible crosscultural and intergenerational commitment to human dignity.

important and – worth protecting and cherishing— is whatever we happen to be in the present context. In short: What is = what ought to be. Our humanity is taken as a given, stable and safe, *as if* it's a persistent and seemingly inevitable and natural state of affairs. Proponents of this view risk being profoundly *ignorant* about history. It obscures what the present generation has inherited from past generations. It underplays incredible cultural variations in the present generation. It also turns a blind eye to how our actions in the present affect future generations. And, frankly, it allows a *lack of imagination* about possible futures for humanity and the worlds we're building.

Some would say that what it means to be human can and should be described biologically in terms of what differentiates us from other species. On this view, we identify the distinguishable, evolved characteristics and capabilities of *homo sapiens*, the species of humans that outlasted the rest.<sup>3</sup> For example, in contrast with all other known species, only *homo sapiens* evolved the complex cognitive and social capabilities needed for widely shared language, fiction (e.g., myth, imagined realities), and social institutions (e.g., trust and rules) that scale beyond small close-knit groups (~n = 150).<sup>4</sup> These basic capabilities and social infrastructure enabled humans to conquer the earth and reconstruct the environment within which we evolve.

The descriptive, biological approach has the advantage of being scientific and revealing a continuum and set of functions that relate humans with other species as well as our natural environment. This approach has its limits, however. Biology doesn't explain everything that's important about humanity. For example, biology can't account for the complex and nuanced ways we relate to and evolve within our reconstructed, built environment. Biology also doesn't fully explain the tools we choose to build and use, much less whether, how, why, or when we should engineer ourselves through our techno-social tools. To more fully understand what matters about being human, how we engineer ourselves and future generations, and how to conceptualize humanity as a shared resource, we need to move beyond evolutionary biology.

After all, we may have evolved certain capabilities that enable survival and our rise to the top of the food chain.<sup>5</sup> However, we don't necessarily value all of those capabilities as central expressions of who we are. Precisely because we have evolved to the point where we can shape ourselves and our society, several philosophical questions have arisen. *How should we exercise such power? What about us should we sustain and cultivate? What should we let go? Who should we aspire to be? How should we engineer ourselves? What type of society should we build and sustain?*

Crucially, human beings can contemplate and act on such questions only because of the various capabilities we've gained through evolution *and* practiced, honed, developed, and sustained collectively. Evolution is necessary but not

<sup>3</sup> Harari (2014), pp. 3–25.

<sup>4</sup> Harari (2014).

<sup>5</sup> Harari (2014), p. 11.

sufficient. What is also necessary, and this is both controversial and incredibly important, is our built world, engineered with our techno-social tools to sustain the basic capabilities that allow us to flourish individually and collectively within and across generations.<sup>6</sup>

According to our approach, (1) what meaningfully distinguishes *homo sapiens* from other species is our capability to imagine, conceptualize, and engineer ourselves and our environment; and, (2) *what matters about being human* is how we exercise such power over generations to collectively produce, cultivate, and sustain shared normative conceptions of humanity.

Humanity can thus be understood as a *set of ideals* about who we are and aspire to be. These ideals are valuable, intangible resources, particularly when shared, acted on, and reflected in our most sacred institutions as shared commitments. Ultimately, we might say that humanity as both a normative concept and as a collectively produced and shared resource (or set of resources) stems from the answers we give to the following fundamental questions:

- Who are we?
- What sort of people do we aspire to be?<sup>7</sup>
- What values and capabilities do we possess and commit ourselves to sustain?
- What sort of society do we want to build and sustain?
- What obligations do we owe to past, present, and future generations? And how should such obligations shape the technological and social institutions we build and sustain?

We – as societies, as communities, as generations, as families, as individuals – answer these constitutional questions directly and indirectly through our actions and the cultures, institutions, infrastructures, and environments we build and sustain.

Suppose we describe *states of human affairs* as the sum of who we currently are as a group, how we see ourselves, and who we want to be.<sup>8</sup> This description can include a set of characteristics and capabilities,<sup>9</sup> some of which we possess and some of which we aspire to possess. People with different normative conceptions (or value

<sup>6</sup> Several political questions accompany the philosophical ones. *Who decides? Who should exercise such power over humanity?*

<sup>7</sup> In this and the fourth bullet point, we use singular words (people, society) where we could use plural words (peoples, societies). We do so only for ease of reading. As we develop further in the text that follows, different groups of peoples and different societies can and do choose to possess and commit themselves to sustaining different values and capabilities. They can and do decide, in different ways, that they owe different obligations to past, present, and future generations.

<sup>8</sup> By focusing on the state of human affairs, we do not mean to overinflate the position of human beings. However, this chapter does have a particular focus: it is about the relationships between humans and the technologies we create and use. Of course, as Peter Singer has argued, the suffering of other species caused by humans and our technologies is important. Nevertheless, consideration of this issue would bring us beyond the scope of this chapter.

<sup>9</sup> The set could include more than characteristics and capabilities. One could focus on knowledge and moral virtues, for example.

systems) might disagree about what characteristics and capabilities ought to be in the set, how to prioritize them, and how we sustain them through social institutions.<sup>10</sup> Such disagreement and diversity produce and are products of different cultures.

Despite such disagreement and diversity, there are some widely shared core ideals, for example, as reflected in the Universal Declaration on Human Rights.<sup>11</sup> These multinational, macro-level normative commitments answer some of the most fundamental constitutional questions about who we are and aspire to be collectively. International human rights laws and institutions create a global community committed to crosscultural standards and moral floors.<sup>12</sup> These and other political processes enable but do not guarantee moral progress over time and across generations and cultures.<sup>13</sup>

Ideals don't become permanent, true, or worth preserving just because lots of people endorse them. They may change and be supplemented with other ideals that vary by culture, as the history of slavery taught us.<sup>14</sup> But across cultures and generations, human beings have exercised our capabilities to conceptualize and engineer ourselves and our environment, to build and sustain a collective heritage, which we characterize as nothing less than humanity itself.

<sup>10</sup> While debates have raged for millennia over what matters about being human and what constitutes a good human life, there are persuasive philosophical accounts that identify several basic human capabilities. See Sen (2005), pp. 151–66; Sen (1985); Sen (2001); Nussbaum and Sen (2004); Nussbaum (2011), pp. 33–34 (2011); Rachels (2014), pp. 15–32 (making the case for universal values that exist across all societies).

<sup>11</sup> Some might criticize our approach because it allows for too much variation and cultural contingency. This objection presupposes too much for the reasons stated in the text. Others might criticize us for not being as sensitive to diversity as we aspire to be. After all, appeals to culture risk focusing on shared values at the expense of recognizing differences in race, class, and gender, as well as commonalities found in subcultures (of which there are many) and norms that only make sense in specific contexts (e.g., what's acceptable at work might not be at home). Our straightforward response to this charge is that we're using culture broadly to refer to any group that's constituted, even if only temporarily, by shared commitments, values, experiences, or yearnings.

<sup>12</sup> For an interesting take on moral floors, see Nussbaum (2007), p. 126 ("any minimally just society will make available to all citizens a threshold level of ten central capabilities, as core political entitlements."); Nussbaum (2011), pp. 33–34.

<sup>13</sup> For the same reasons that we reject technological determinism, we reject corresponding notions of moral determinism. Moral progress and regress are possible.

<sup>14</sup> Throughout history, cultures have built diverse worlds that allowed different values to become preeminent or techno-socially engineered into existence. In *The Order of Things*, philosopher Michel Foucault contends that fundamental values like what constitutes "humanity" have been constantly redefined throughout history to suit a variety of agendas and powerful actors (Foucault 1994). Indeed, it's hard to deny that the confluence of power and prejudice – racism, sexism, classicism, and ableism, among other pernicious "isms" – has had an oversized influence in determining who gets to count as being sufficiently similar to ingroups to qualify as human. Moreover, formulations of humanity and the imagined worlds that support these conceptions can lose their hold on us, just like perceptions of the gods do – a shift that's aptly illustrated by polytheism being displaced by monotheism in large parts of the world. We could create a laundry-list of the features that have differentiated worlds across human history.

Preserving the “fundamental blessings” of humanity is the most important constitutional commitment that unites cultures across generations.<sup>15</sup> In his Lyceum Address of 1838, Abraham Lincoln recognized that the “fundamental blessings” passed on from generation to generation extend beyond the blessings of the Earth to include the blessings of society – the communal heritage of law, political institutions, and fundamental rights of liberty and equality.<sup>16</sup> Lincoln reminded his generation, as his words ought to remind us today, that the fundamental resources on which any society depends include the blessings bestowed on any present generation by sacrifices of its ancestors. Lincoln’s speech, like the Gettysburg Address,<sup>17</sup> offers a powerful vision of a transgenerational social contract firmly rooted in equity. Each generation inherits a wealth of natural and communal resources. In return for this boon, it’s obligated to transmit these resources “to the latest generation that fate shall permit the world to know.”<sup>18</sup> This duty to transmit a legacy to the future reverberates in many cultures. Lincoln’s speech implicitly harkens back to the Athenian Ephebic Oath by which men of ancient Athens swore to “transmit my fatherland not diminished [b]ut greater and better than before.”<sup>19</sup> The intergenerational moral obligation is rooted in a more traditional conception of equity, akin to the repudiation of unjust enrichment. The present generation is morally bound to perform its duty to transmit because its own welfare and humanity has been enriched by access to and use of the resources passed on to it. To accept the benefits without satisfying the attendant duty would constitute enrichment at the expense of future generations.<sup>20</sup>

Humanity, conceived of normatively as a shared set of ideals reflected in us and our built world of imagined realities, institutions, infrastructures, and environments, is at risk of deterioration by pervasive techno-social engineering. We focus on specific forms of techno-social engineering that affect the basic capabilities that enable us to ask and participate in answering fundamental questions about who we are and aspire to be, individually and collectively. Thus, we consider thinking

<sup>15</sup> This section draws from Frischmann’s article *Some Thoughts on Shortsightedness and Intergenerational Equity*. See generally Frischmann (2005).

<sup>16</sup> Lincoln (1838).

<sup>17</sup> Jaffa (1959), p. 228 (citing Lincoln 1863). (“The ‘people’ is no longer conceived in the Gettysburg Address, as it is in the Declaration of Independence, as a contractual union of individuals in the present; it is as well a union with ancestors and with posterity; it is organic and sacramental.”)

<sup>18</sup> Lincoln (1838). See also Akwesasne Notes (1977) (often referred to as “The Iroquois’ Law of Seven Generations”); Morris (1995) (discussing the “centuries-old Haudenosaunee philosophy that all major decisions of a nation must be based on how those decisions will affect at least the next seven generations.”)

<sup>19</sup> Swift (1947), p. 4 (describing the Athenian Ephebic Oath translation by Clarence A. Forbes).

<sup>20</sup> The analogy to unjust enrichment is imperfect. Unlike unjust enrichment in which the beneficiary to whom the benefit is conferred compensates the person who conferred the benefit and so involves only two parties looks to the past, the dynamic we describe involves three parties (past, present, and future generations) and looks to the future. We thank John Breen for pointing this out.



capacities, the ability to socialize and relate to each other, free will, autonomy, and agency.

Some may disagree with our choices of capabilities to examine. They may choose to examine others. The bottom line is that across cultures and generations, humans have engineered themselves and their built environments to sustain these and other core capabilities. In our view, they are part of our shared heritage, our humanity. And again, they are at risk of being whittled away through rampant techno-social engineering driven by many different forces and logics. Taylorism extended and fetishized computational power, and the allure of ever more powerful intelligent control systems promise tremendous gains in efficiency and productivity along with the convenience and happiness of optimized lives. But at what cost?

#### OPTIMIZED LIFE ON THE EXPERIENCE MACHINE

Over forty years ago, Robert Nozick wondered whether he or anyone else would choose to plug into a hypothetical “experience machine” that could convincingly simulate any desired experience. In the blink of an eye, the experience machine would let you take on the role of a renowned novelist, a caring father, an ascetic saint, or any one of myriad other possibilities, like rock star, brilliant scientist, or world-class athlete. Nozick’s scenario offers a choice to plug into a virtual reality machine that guaranteed a “lifetime of bliss.”<sup>21</sup>

If you were presented with the opportunity, would you choose to plug into the experience machine?

We don’t claim to know the right answer for you. People have different intuitions. This classic thought experiment prompts imagination and deliberation about one’s basic conception of a good life. Is there more to life – or better yet, the good life – than a person’s subjective experience of happiness? Or, as Nozick put it, than how one’s life feels from the inside?

For hardcore hedonists, the decision is straightforward: Plug in. It guarantees optimal happiness, the highest aggregation of moment-by-moment positive feelings. The movie *The Matrix* toyed with a more reflective version of the thought experiment when a character named Ciphher announces that he understands he’s living in a simulation but still prefers that world to alternatives where he’d be less happy. “I know this steak doesn’t exist. I know when I put it in my mouth the Matrix is telling my brain that it is juicy and delicious. After nine years you know what I realize? Ignorance is bliss.”<sup>22</sup>

<sup>21</sup> To make his objections vivid, Nozick introduced the experience machine thought experiment (1974 and revisited in 1989), a hypothetical scenario that bears a striking resemblance to Ray Bradbury’s earlier short story, “The Happiness Machine” (1957). Weijers (2011).

<sup>22</sup> For an extended discussion of this issue, see Grau (2005).

Most people would not choose to plug in.<sup>23</sup> And that's because most people are not hardcore hedonists.<sup>24</sup> Of course, happiness matters for most people and is an important component of a life lived well. But many other things matter too, and they are not reducible to or commensurable with happiness. Pursuing pleasure exclusively, as the ultimate end, would lead to a rather shallow life.<sup>25</sup>

Nozick contended that it would be a mistake to plug into the experience machine.<sup>26</sup> His reservations revolved around a core conviction that many people share today: “[R]eality has intrinsic prudential value.”<sup>27</sup> No matter how realistic a simulated world feels, it lacks the features of an independent reality conducive to human flourishing. It seems the epitome of a dehumanizing environment. It's a fully engineered reality wherein simulated people lack free will, simulated things bend to our will and desires, history has no weight, interdependence doesn't exist, and even the laws of physics can be broken. In such a programmed world, our actions wouldn't be meaningful. Our accomplishments, including caring for others, would be hollow. Our relationships would be fake. And at least some of our basic human capabilities would atrophy. Fortunately, we wouldn't realize any of this while living on the machine because we'd be programmed to be oblivious to such concerns. And we'd feel really good.

There's a related concern that plugging in would lead to an inauthentic life determined by others. A life lived on the experience machine would not be one's own. Too much of the life lived and happiness experienced would be determined by the machine, or, to be more precise, the engineers who created the machine. This concern has weight for those who insist that the means matter when evaluating the quality of one's life. On this view, a well-lived life requires some human agency and exercise of free will. By contrast, a fully programmed and determined life on the machine is not a well-lived life, regardless of how much pleasure the machine supplies.

It's important to appreciate that these considerations are perfectly legitimate products of the thought experiment. By forcing someone to choose upfront whether

<sup>23</sup> Weijers (2014); Kolber (1994). There is a debate among philosophers and experimentalists about whether thought experiments like Nozick's are amenable to empirical study. Smith (2011); De Brigard (2010).

<sup>24</sup> Bramble (2016). This may be an empirical claim worth exploring. We are unaware of a definitive study and are not certain that empirical testing would work well.

<sup>25</sup> Susan Wolf offers compelling examples that illustrate why thinkers like Nozick don't reduce meaningful experiences to instances where we strive for pleasure or even necessarily experience it. Wolf (2015), p. 51.

<sup>26</sup> Ben Bramble observes: “Hedonism has few contemporary advocates. This is mainly due to a single, highly influential objection to it, widely considered to be decisive: Robert Nozick's experience machine. Discussions of well-being—whether in scholarly journals, academic conferences, or university lecture halls—often begin with a quick dismissal of hedonism by reference to Nozick's objection before turning to ‘more interesting matters’ (usually the question of which desire-based or hybrid theory of well-being is true.” Bramble (2016), p. 136.

<sup>27</sup> Weijers (2011).

to plug in, the thought experiment prompts deliberation. Some argue that critical resistance to plugging in is rooted in the framing effects of the narrative. For example, status quo bias may lead people to prefer what they currently experience over the unknown life on a machine.<sup>28</sup> Or people can question whether the experience machine will actually work: *What happens when the power goes out? Will the unplugged be able to take advantage of those who are plugged in?* These arguments are valid, but too easily taken as the final word. Framing effects, cognitive biases, and other related concerns can and should be dealt with by engaging the thought experiment thoughtfully, rather than simply dismissing it. Behavioral psychology offers excellent grounds for criticizing and improving thought experiments but does not provide answers to the underlying normative questions.

Others, such as economist Richard Layard and philosopher Joseph Mendola, criticize the thought experiment for being “wildly unrealistic” and pumping unfair intuitions with “unfamiliar gadgetry which invokes our fear of the unfamiliar.”<sup>29</sup> These criticisms might have had purchase a few decades ago. But they seem wildly exaggerated today. Modern technology and culture make the experience machine scenario reasonably familiar and sufficiently realistic for people to consider. Regardless, these objections to the thought experiment hardly provide a better explanation for people’s preferences than the claim that reality has intrinsic prudential value.

In our view, a significant reason why many people would choose not to plug into the experience machine is that they believe in free will and value having it. To plug into the experience machine would be to accept engineered determinism and know, even if only at the moment of deliberation about whether to plug in, that any subsequent experience of free will would be illusory. Many people highly value free will. Arguably, it is one of the fundamental blessings of humanity, collectively cultivated and sustained across generations through the cultures, institutions, infrastructures, and environments we build and sustain.

But what if the promise of optimal happiness on the machine makes the thought experiment too outlandish? The utopian allure might exacerbate the framing effects, or it might trigger doubts, fears, or even guilt.

Suppose we adjust the thought experiment so that life on or off the machine would be the same in terms of aggregate happiness.<sup>30</sup> Suppose a person’s life on the experience machine is assumed to be identical to the life they’d live off the machine.<sup>31</sup> If the

<sup>28</sup> Weijers (2014), p. 3.

<sup>29</sup> Mendola (2006), pp. 441–77; Layard (2005). See also Bronsteen, Buccafusco, and Masur (2014), pp. 172–75 (suggesting that the experience machine thought experiment pumps “inadmissible intuitions”).

<sup>30</sup> Nozick tilted the scales in favor of hedonism by guaranteeing optimal happiness. He probably did this to show why hedonism, even in its most tempting form, is repugnant once you think carefully about it. By making the lives on and off the machine equivalent, we adjust the extreme framing of the tradeoffs. This might eliminate speculation about optimal happiness and elevate status quo bias as a factor.

<sup>31</sup> Crisp (2006), pp. 635–36. “According to hedonism, [the lives on and off the machine] have exactly the same level of well-being. And that is surely a claim from which most of us will recoil.” Crisp goes on to argue that our beliefs about the value of accomplishment (and by extension free will) might be “an

life Jane experiences on and off the machine is identical from her subjective perspective, hedonists argue Jane would be equally well on and off the machine. Her well-being – the goodness of her life – is the same. And so, they argue, she should be indifferent to choosing one life over the other. She should be willing to flip a coin.

*What about you? Would you be willing to flip a coin?* Heads would mean you plug in. And tails would mean you do not plug in. Keep in mind that life on and off the machine is guaranteed to be identical.

If you are truly indifferent, you should be willing to flip a coin and accept the outcome either way. Perhaps some people would do so. We suspect most people would not actually be willing to flip a coin. Further, it's hard to imagine that anyone would choose to plug into the experience machine if it promised an identical life. Most people would choose not to plug in. Some might argue that unwillingness to flip a coin should be construed as an indication of irrational biases, whether the status quo bias or some irrational fear that fights the hypothetical. But we think a much better explanation is that they would not be indifferent to the two lives. They would choose to maintain the status quo and for good reason. They value free will, or, at least, the illusion of free will. In other words, they would take a pragmatic approach and wager in favor of free will rather than plug into the experience machine, which would eliminate free will, and, at best, offer an engineered illusion of it.<sup>32</sup> Thus, what people fear about the experience machine is precisely what Nozick highlighted: *The loss of authenticity in a world of engineered determinism*.

Let's flip the scenario around. Suppose you are informed that you're currently plugged into an experience machine and that you've been selected for a one-time opportunity to unplug and experience an authentic life. Unlike Ciper in *The Matrix*, your life unplugged is guaranteed to be identical to the plugged-in life you're currently living.

What would you choose if you were given the opportunity to (1) unplug, (2) stay plugged in, or (3) flip a coin such that heads would mean you unplug and tails would mean you stay plugged in?

It's a little more difficult to predict what people would do, although for somewhat surprising reasons. We don't think many people would choose to flip a coin. Most people are not hardcore hedonists; they wouldn't be indifferent to the two lives (worlds). Initially, one might think people would unplug and choose an authentic life. But most people would not do so.<sup>33</sup> Some have suggested that this demonstrates a cognitive bias infecting the thought experiment. A better explanation is that in the reality we know and experience, most people would take a pragmatic approach and wager in favor of free will, meaning their existing belief in free will. In a sense, this is a reaction that fights the hypothetical. But, importantly, it only fights the hypothetical

example of a kind of collective bad faith, with its roots in the spontaneous and largely unreflective social practices of our distant ancestors." Perhaps. But perhaps he has it backwards.

<sup>32</sup> See *Re-Engineering Humanity* for our Free Will Wager.

<sup>33</sup> De Brigard (2010), pp. 43–57.

claim about our reality being fully engineered and determined. Absent a high level of proof that our current reality is fully determined (naturally or by engineering), wagering in favor of free will remains the best strategy.

The experience machine thought experiment usefully highlights competing value systems and prompts deliberation about what matters to you about your life. Hardcore hedonists would choose to plug in; most other folks would not. There are many reasons people might choose not to plug in. But an important reason is that a programmed world governed by engineered determinism is not conducive to human flourishing. At the most basic level, life on the experience machine doesn't jibe with how most people answer any of the basic constitutional questions. We don't see ourselves, our lives, or our reality as fully programmed or determined, even though we understand that much of what we do and who we are is the product of various factors outside of our control. Neither do we aspire to such a fate.

*Do we aspire to build such a world for our children and future generations?* Notably, this is a moral question outside the scope of Nozick's thought experiment. To see why, consider the importance of the cord and plug. The decision about whether or not to plug into the experience machine served a few purposes:

- It kept the thought experiment focused on an individual decision about an individual life.
- It prompted deliberation by an individual about that individual's conception of a good life for that individual, which is often referred to as the individual's well-being or welfare.
- It eliminated concerns about paternalism, social engineering, or social welfare.
- It implied the existence of multiple worlds – at least, worlds on and off the machine.

The myopia of the thought experiment limits its relevance, however, when we turn our attention to social arrangements and world building. Yet that is where we must go if we are to answer the fundamental questions about the world we're building for posterity.

#### EXPERIENCE MACHINE N.O – LIFE IN THE MACHINE/WORLD

Let's eliminate the plug. Consider a different version of the experience machine, which we refer to as the Experience Machine n.o. Imagine a ubiquitous smart techno-social environment that spans the Earth and optimizes the planet to provide human inhabitants optimal happiness. Rather than ask you whether you'd plug yourself in, we'd like for you to consider whether such a machine/world should be built.

If you were given the ultimate decision-making authority, would you build such a machine/world?

Moving from a decision about whether (i) to plug yourself into the machine to (ii) building a machine/world for all humans complicates the analysis and shifts the focus from individual well-being to broader moral and social concerns. To make the situation less abstract, let's be clear about what we mean by optimal happiness. Assume that the Experience Machine n.o supplies all human beings on Earth with *maximum happiness*, measured moment-by-moment for each individual and aggregated over a lifetime, at *minimal social cost*. (Our arguments hold if we replace happiness with pleasure, positive affect, or other positive mental states.)

We must admit that this sounds pretty darn good. Optimal happiness appears to be an end worth pursuing. But *what about the means? Do they matter?*

What could the Experience Machine n.o look like and how might we build such a machine/world? Extrapolating from the present to the near future, we envision that the Experience Machine n.o would be comprised of interconnected sensor networks and data-driven automation of socio-technical systems around, about, on and in human beings. Imagine that within the next few decades, the following occurs. Large multinational companies<sup>34</sup> gradually build and connect smart techno-social environments that actually deliver on their promises. The scope of deployment expands to the point where there is seamless interconnection and integration across all environments within which humans live. The normative agenda executed throughout all this construction and deployment is optimal efficiency, productivity, and happiness.

Guided by the optimization criterion of maximum happiness at minimal social cost, the Experience Machine n.o necessarily would engineer human beings. After all, human beings are inefficient and costly to sustain. Optimization would entail minimization of various costs associated with humans being human. For example, our bodily and mental engagement with the physical world entails logistical, navigational, and various other transaction costs. Outsourcing to intelligent control systems would minimize these costs. Making decisions, experimenting, and learning (among other mental processes) are also costly endeavors. Again, the optimizing logic would press toward minimizing and potentially eliminating these costs. Interacting with one another, coordinating behaviors, developing relationships, and many other aspects of interdependent human relations entail transaction costs to be minimized.

Finally, there is a subtler type of techno-social engineering prevalent in the Experience Machine n.o. Since what makes us happy is in large part contingent on environmental conditions and experiences, and since those factors are fully determined within the machine/world, optimization also would entail engineered tastes, beliefs, preferences, and other factors that feed into our affective feelings of pleasure.

<sup>34</sup> We are not committed to identifying multinational companies as the architects. Governments, public-private partnerships, and others would also presumably be involved. As we explain elsewhere, the point is not to allocate blame to any master planner.

Bluntly, the cheapest way to make billions of human beings perfectly happy – particularly when using the sorts of technological means we’re imagining – is to set the bar very low. In this case, the techno-social system need only meet or even barely surpass expectations. As hedonists know and often are prone to emphasize, people adapt their beliefs, preferences, and expectations to their conditions and, subsequently, their corresponding happiness levels typically adjust.<sup>35</sup> So, the goal might very well be to shape beliefs, preferences, and expectations in a manner that makes supplying happiness as cheap and easy as possible. At the end of the day, optimal happiness would boil down to satiation of engineered will.

There are many possibilities, however. Perhaps machine learning would settle on different equilibria. We can imagine extremely dystopian science fiction scenarios, such as *Wall-E*, where humans are dumb satiated fools. But we also can imagine scenarios where the machine/world manufactures higher happiness aggregates through different types of programs. John Stuart Mill famously argued that it is “better to be Socrates dissatisfied than a fool satisfied.”<sup>36</sup> Perhaps the Experience Machine n.o would produce a world filled with happy sages. Cheap engineered bliss need not exclude higher pleasures of the sort Mill defended. Higher pleasures often are cultivated and learned, and cultivation and learning entail costs.<sup>37</sup> On one hand, minimizing these costs might lead to outsourcing or pushing toward engineering wills satiated with lower pleasures. On the other hand, cultivating certain sets of tastes that correspond to higher pleasures might be worth the cost if they produce even more net pleasure. *Who knows?* Frankly, it’s impossible to know exactly how the optimization would work out. But that’s really beside the point. One way or another, techno-social systems would determine whether and what to cultivate, who and what we are. What makes humans happy would be predetermined rather than produced through the exercise of free will, experimentation, and self-authorship.

Regardless of the scenario, our thought experiment raises a fundamental normative question: Would maximizing human happiness at minimal social cost through the Experience Machine n.o justify forcing everyone to live a fully determined life?

We strongly suspect few people would answer in the affirmative, at least on reflection.<sup>38</sup> Think about what it would mean. Building the Experience Machine n.o is functionally equivalent to forcing everyone to plug into Nozick’s Experience

<sup>35</sup> As Amartya Sen argued long ago in his critique of welfarism, an incredibly poor person with very little opportunity in life might be subjectively happy because she has adapted to her conditions in life, but that cannot mean that society should not be committed to reducing poverty or investing in building the capabilities of her daughters and sons or of future generations of similarly situated people. Sen (1985); Sen (2001); Nussbaum and Sen (2004).

<sup>36</sup> Mill (1962), p. 9.

<sup>37</sup> Historically, not everyone has had the means to experience Mill’s higher pleasures. Significant distributional and class-based concerns challenge appeals to higher pleasures. Further, strong commitment to individualism triggers concerns about paternalism – *who’s to say what is higher?* We don’t address these concerns, except to note that perhaps Experience Machine n.o could level the playing field. Of course, it is not clear how such leveling would affect humanity.

<sup>38</sup> Of course, if we’re wrong, then that would tell us something about people’s baseline normative values.

Machine. Even the most hardcore hedonist would hesitate before imposing a machine life on others. Such paternalism conflicts directly with longstanding ideals concerning free will, autonomy, and agency, which are shared by many cultures and widely regarded as fundamental blessings of humanity.

Yet hesitation does not mean rejection. We posed this thought experiment to Peter Singer, a renowned philosopher and incredibly thoughtful and generous person, and after some discussion and deliberation, he replied: "I'm a sufficiently hardcore hedonist to think that democracy is a means to an end rather than an end in itself. If we can really imagine achieving optimum happiness for all sentient beings, forever, that would be a greater value than democracy, which may be better than any other system of government, but so far hasn't got anywhere close to producing optimum happiness for all sentient beings." Optimal happiness for everyone on Earth is an end worth considering carefully, particularly in light of how much misery and suffering exists in our world.

The Experience Machine n.o poses a threat to the liberal democratic ideal that people should be generally free to choose their own ends with minimal interference. Commitments to forms of this ideal have proved to be the best way for diverse people to live together, have real opportunities to experiment with different ways of living, and determine, over time and with the accrual of experience, which ways to live well. A world that makes it very hard or impossible to opt out of Experience Machine n.o would violate this ideal by interfering too strongly with our capacity to freely set ends for ourselves. Such interference is morally wrong and should be politically resisted. Although different in some respects, the spread of hedonism through neo-Taylorism amped up by interconnected and ubiquitous smart devices functions much like an authoritarian government imposing a mandatory state religion.

But if longstanding ideals are socially constructed and thus contingent on techno-social engineering systems, would building the Experience Machine n.o be defensible so long as it was done gradually? Deployed and integrated incrementally over decades, people could be gradually prepared for and conditioned to accept this brave new world. Engineered beliefs could pave the slippery sloped path to the Experience Machine n.o. This is why we've emphasized the importance of asking these questions at a time when it's still possible to recognize and evaluate the path we're on.

Some would argue that any resistance people currently have toward building the machine/world is itself the product of techno-social engineering of beliefs, preferences, and values by past generations. This is true. The present generation has inherited the fundamental blessings of the past, including shared ideals about who we are and who we aspire to be. But this point doesn't undermine the inquiry or make it any easier to answer the difficult and pressing questions highlighted by our thought experiment and more thoroughly presented throughout our book.

Finally, some will argue that like Nozick's version, the Experience Machine n.o thought experiment is wildly unrealistic and nothing more than an unfair intuition



pump. They also will argue that the scenario is scary and framed in a manner that unfairly triggers psychological biases that distort rational analysis. These are fair points. But they are easily overstated and taken too far. Again, we use the thought experiment to prompt deliberation, analysis, and discussion. It is not a theoretical proof or empirical experiment.

While we truly hope our imagined machine/world is wildly unrealistic, there are various reasons to believe the slippery sloped path we're on is headed in that direction. Regardless of whether we reach the end-state, we need to think more deeply about the world we're building, both because of what it means for our children and future generations and because of how it affects us as we proceed down the path.

Twenty-first century techno-social engineering deeply affects how we think, feel, and interact with one another. Outsourcing so many of these functions to techno-social systems can't and shouldn't be assumed to be in our interest, neutral, or mere natural extensions of ourselves. We need to be aware of atrophying capabilities, mind control, and the gradual loss of human dignity as more aspects of our lives are determined by smart techno-social systems.<sup>39</sup>

*We are not fully predictable and programmable machines.* In all likelihood, we never will be. But that is no reason to become complacent. Much of what matters about being human can be lost in partial deprivations as we march down the slippery sloped path we're on.

#### PLURALISM AS A CONSTRAINT ON TECHNO-SOCIAL ENGINEERING OF HUMANITY

The transition from Experience Machine to Experience Machine n.o marked a significant shift in focus, from evaluating individual well-being to evaluating humanity in a social world. We hope that on reflection, most people, including hardcore hedonists, would decline to build Experience Machine n.o because they recognize that there is more that matters about being human than subjective feelings of happiness.

We can build much better worlds that are conducive to human flourishing and a plurality of values.<sup>40</sup> Free will and agency matter, above and beyond how those

<sup>39</sup> The techno-social engineering test framework developed in *Re-Engineering Humanity* (2018) provides a set of conceptual tools for identifying and contextually evaluating these risks.

<sup>40</sup> This is an important caveat. Suppose the Experience Machine n.o thought experiment framed the choice in binary terms as follows: (1) Build the Experience Machine n.o; or (2) Retain our current world as it currently exists. This choice presents a more difficult decision. The first option might sacrifice actual free will, but it would provide the illusion of free will and tremendous welfare gains in terms of longer and happier lives for billions of people. The second choice might sustain free will, but it would sacrifice the opportunity for massive social welfare gains. If these are the only choices, the tradeoff might be very difficult. It is critical, in our view, to recognize that these are not the only choices. We can build much better worlds.

capabilities cash out in terms of subjectively experienced well-being. But worst case, even if most people say they would choose to build the Experience Machine n.o., that empirical attestation of their values does not provide sufficient justification. This is where liberalism and paternalism clash.<sup>41</sup> A confounding difficulty with liberalism and relying on individuals' existing preferences and beliefs is that such preferences and beliefs are contingent on the techno-social environment within which they've lived their lives. It's hard to fully credit preferences and beliefs that significantly discount the constitutional value of free will and agency. After all, deferring to individuals' choices only makes sense if the individuals themselves have free will and practical agency, and if they do, then it seems morally wrong to deny the same capabilities to future generations, to deny them access to the fundamental blessings of humanity passed on to the present generation.

In the end, a commitment to pluralism demands freedoms engineered into our built environments. Accordingly, one of the most important constitutional questions of the twenty-first century is how to sustain the freedom to be off, to be free from techno-social engineering, to live and develop within underdetermined techno-social environments. A precursor to implementing any potential responses is recognizing the threat that engineered determinism poses to humanity. The sirens call of data-driven, supposedly smart techno-social systems promises cheap bliss – a world of optimal happiness, but it ignores the carnage of humanity lost.

#### REFERENCES

- Berlin, Isaiah (1969). Two Concepts of Liberty, in I. Berlin, *Four Essays on Liberty*, London: Oxford University Press. New ed. in Berlin 2002.
- Berlin, Isaiah (1978). From Hope and Fear Set Free, in I. Berlin, *Concepts and Categories. Philosophical Essays*, ed. H. Hardy, London: Hogarth Press; Oxford: Oxford University Press, 1980. Reprinted in Berlin 2002.
- Bramble, Ben (2016). The Experience Machine. *Philosophy Compass*, 11(3), 136–45.
- Bronsteen, John; Buccafusco, Christopher; and Masur, Jonathon S. (2014). *Happiness and the Law*. Chicago, IL: Chicago University Press.
- Carter, Ian (2016). Positive and Negative Liberty, *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta.
- Crisp, Roger (2006). Hedonism Reconsidered. *Philosophy and Phenomenological Research*, 73(3), 619–45.
- De Brigard, Felipe (2010). If You Like It, Does It Matter if It's Real. *Philosophical Psychology*, 23(1), 43–57.
- Foucault, Michel (1994). *The Order of Things: An Archaeology of the Human Sciences*. New York: Vintage Books.
- Frischmann, Brett M. (2005). Some Thoughts on Shortsightedness and Intergenerational Equity. *Loyola University Chicago Law Journal*, 36(2), 457–67.

<sup>41</sup> Paternalism kicks in where there is good reason to conclude that people don't know what's best for them; liberalism is skeptical of any such alleged state of affairs. The battle usually devolves into the core political question of "Who Decides?"

- Grau, Christopher (ed.). (2005). *Philosophers Explore the Matrix*. New York: Oxford University Press.
- Harari, Yuval Noah (2014). *Sapiens: A Brief History of Humankind*. New York: Vintage.
- Jaffa, Harry V. (1959). *Crisis of the House Divided: An Interpretation of the Issues in the Lincoln-Douglas Debates*. Chicago, IL: University of Chicago Press.
- Kolber, Adam J. (1994). Mental Statism and the Experience Machine. *Bard Journal of Social Sciences*, 3, 10–17.
- Layard, Richard (2005). *Happiness: Lessons from a New Science*. New York: Penguin Group.
- Lincoln, Abraham (1838, January 27). The Perpetuation of Our Political Institutions, Address Before the Young Men's Lyceum of Springfield, Illinois (January 27, 1838). Retrieved from <http://Federalistpatriot.us/histdocs/Lincolnlyceum.html>.
- Mendola, Joseph (2006). Intuitive Hedonism. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 128(2), 441–77.
- Mill, John S. (1962). *Utilitarianism: On Liberty*. London: Collins.
- Morris, Glenn. For the Next Seven Generations: Indigenous Americans and Communalism. Fellowship for Intentional Community. Retrieved from [www.ic.org/wiki/next-seven-generations-indigenous-americans-communalism/](http://www.ic.org/wiki/next-seven-generations-indigenous-americans-communalism/) (last visited January 6, 2005).
- Nozick, Robert (2013). *Anarchy, State, and Utopia*. New York: Basic Books.
- Nussbaum, Martha (2007). The Capabilities Approach and Ethical Cosmopolitanism: A Response to Noah Feldman. *Yale Law Journal Pocket Part*, 117, 123–29.
- Nussbaum, Martha (2011). *Creating Capabilities: The Human Development Approach*. Cambridge, MA: Harvard University Press.
- Nussbaum, Martha and Sen, Amartya K. (2004). *The Quality of Life*. Oxford: Oxford University Press.
- Rachels, James (2014). The Challenge of Cultural Relativism. In Stuart Rachels (ed.), *Elements of Moral Philosophy* (pp. 15–32). New York: McGraw-Hill Education.
- Sen, Amartya K. (1985). *Commodities and Capabilities*. Amsterdam: North-Holland.
- Sen, Amartya K. (2001). *Development as Freedom*. Oxford: Oxford University Press.
- Sen, Amartya K. (2004). *Rationality and Freedom*. Cambridge, MA: Harvard University Press.
- Sen, Amartya K. (2005). Human Rights and Capabilities. *Journal of Human Development*, 6(2), 151–66.
- Sen, Amartya, K. (2011). *Development as Freedom*. New York: Anchor Books.
- Smith, Basil (2011). Can We Test the Experience Machine?, *Ethical Perspectives*, 18 (1), 29–51.
- Swift, Fletcher H. (1947). *The Athenian Ephebic Oath of Allegiance in American Schools and Colleges*. Los Angeles, CA: University of California Press.
- United Nations. *Universal Declaration of Human Rights*. Retrieved from [www.un.org/en/universal-declaration-human-rights/](http://www.un.org/en/universal-declaration-human-rights/).
- Weijers, Dan (2011). A Review and Assessment of the Experience Machine Objection to Hedonism. Retrieved from [www.danweijers.com/pdf/A%20Review%20and%20Assessment%20of%20the%20Experience%20Machine%20Objection%20to%20Hedonism%20-%20Dan%20Weijers.pdf](http://www.danweijers.com/pdf/A%20Review%20and%20Assessment%20of%20the%20Experience%20Machine%20Objection%20to%20Hedonism%20-%20Dan%20Weijers.pdf).
- Weijers, Dan (2014). Nozick's Experience Machine Is Dead, Long Live the Experience Machine. *Philosophical Psychology*, 27(4), 513–35.
- Wolf, Susan R. (2015). *The Variety of Values: Essays on Morality, Meaning, and Love*. New York: Oxford University Press.