

Digital Humanism, Digital Nudging and the Scientific Basis of Free Will

Oliver Braganza and Johannes Schultz, University of Bonn

1 Introduction

Digital humanism, and in particular digital nudging, is reigniting old debates about free will. The reason is that digital technologies are continuously uncovering ways to influence our behavior and beliefs without our full conscious perception and control. The increasing potential for digital nudging has furnished what could be cast as two opposing visions for digital humanism. One side argues nudging is inevitable and we should deemphasize free will to promote humanist goals. The other side insists that humanism requires emphasizing free will and constraining nudging. We suggest that the normative arguments of both sides have merit and the apparent contradiction can be resolved by a modern scientific understanding of free will. Such an understanding casts free will as an emergent phenomenon, arising from the continuous interaction of indeterministic and deterministic processes in a temporally recursive loop. The resulting model, which we label the *temporally recursive two-stage model* suggests free will is a graded, temporally unfolding, and partially social phenomenon. It acknowledges that both nature and nurture shape our behavior, character and will, but reasserts freedom as the ability of our emergent selves to become aware of, and cumulatively affect, this shaping. The upshot is that it is no longer sufficient to assume we are free unless we are being intentionally manipulated. Instead, asserting free will in a digital era must increasingly mean to *judge the nudge*.

Modern digital technologies and their social consequences have furnished a call to reinvigorate humanist ideals within what has been termed *digital humanism*.¹ Such technologies are reigniting ancient debates on “what it means to be human”²

1 Vienna Manifesto on Digital Humanism, DIGHUM, May 2019, <https://caiml.org/dighum/dighum-manifesto/>.

2 DIGHUM, Vienna Manifesto.

and, for instance, on whether we have free will or not.^{3,4} One reason is the increasing power to influence, or ‘digitally nudge,’ individual decisions and beliefs without us even noticing. As we will see, this technological development raises questions concerning common philosophical positions on free will, such as *compatibilism*. Compatibilism represents an attempt at reconciling the belief that humans possess free will with *determinism*. Determinism represents the view “that everything that occurs is the inevitable result of the laws of nature and the state of the world in the distant past.”⁵ According to what one could term ‘naïve compatibilism,’ even choices that are entirely determined by non-conscious factors, but are subjectively experienced as free, can be understood as evidencing free will.⁶ Compatibilists may for instance suggest that perceived conscious *will* simply operates at a different scale of organization than lower-scale physical determinism (or digital nudging), and that evidence of some subconscious determining factor thus does not undermine freedom.

On the one hand, compatibilism is an appealing position, because it provides a *definitive protection* of the concept of free will from a continuous onslaught of reductionist attacks.⁷ It allows those who believe free will to be essential, to rest assured—no scientific finding about how digital nudging (or neurobiology) covertly shapes our choices can become a threat. Even if the universe were fully deterministic, we could still base society on the premise of free choice.⁸ Unfortunately, the compatibilist focus on the *subjective experience* of free will also entails a great danger, which is crystallized by digital nudging—it makes a principled objection to manipulation much more difficult.

The issue has perhaps been examined in most detail in what is known as the *manipulation argument against compatibilism*.⁹ In many ways, digital nudging is practically playing out such a ‘manipulation argument’ in real time, at the societal level.

3 Alex Pentland, *Social Physics: How Social Networks Can Make Us Smarter* (New York: Penguin Books, 2015), 235.

4 Some also feel that advances in “artificial intelligence” (AI) raise the question if AI-systems need to be treated as something more than mere machines. This issue is not discussed in the present article.

5 Derk Pereboom and Michael McKenna, “Manipulation Arguments Against Compatibilism,” in *The Oxford Handbook of Moral Responsibility*, 1st ed., ed. Dana Kay Nelkin and Derk Pereboom (New York: Oxford University Press, 2022).

6 Daniel C. Dennett and Gregg D. Caruso, “On Free Will: Daniel Dennett and Gregg Caruso Go Head to Head | Aeon Essays,” *Aeon*, October 4, 2018, <https://aeon.co/essays/on-free-will-daniel-dennett-and-gregg-caruso-go-head-to-head>.

7 See Robert M. Sapolsky, *Determined: Life Without Free Will* (London: Bodley Head, 2023).

8 Daniel C. Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting* (Cambridge, MA, London: MIT Press, 1984), <https://doi.org/10.7551/mitpress/10470.001.0001>.

9 Pereboom and McKenna, “Manipulation Arguments Against Compatibilism.”

Briefly, the argument harnesses the intuition that manipulation (and nudging)¹⁰ can erode free will. Yet, for a compatibilist, any manipulation would arguably just be a switch from one subconscious determining factor to another. If only the subjective feeling of free choice matters, then what exactly is the problem with manipulation or nudging?

A compatibilist might argue that manipulation can be objectionable for other reasons than that it subverts conscious control over one's own choices. Subconscious influence may for instance be deemed freedom constraining *only* if it followed from someone's ill intentions or is otherwise undesirable (say if it is on a "nasty list").¹¹ In this view, the problem with manipulation (or nudging) is no longer that it acts beyond full conscious control. Instead, we need to base our assessment of freedom on a value judgement or speculate about others' intentions. Unfortunately, value judgements are subjective, and intent is often exceedingly difficult to establish. What is more, digital nudging is increasingly emerging spontaneously, without requiring any intentional design (see section 2.2). So, we are again left with the question whether digital nudging constrains free will and individual autonomy, or whether we first have to make a value judgement on the specific nudge.

These and similar questions have furnished two diametrically opposing visions of what 'digital humanism' could come to mean. One side is personified by Alexander Pentland, an intellectual figurehead of Silicon Valley, who promotes the study of what he calls "Social Physics." Pentland argues we should use digital technology to i) study determinants of behavior, and ii) use the generated data to digitally nudge individuals towards the greater good. Note that his agenda is explicitly humanist in the sense that his stated purpose is to improve human lives. He aims to "build a world without war or financial crashes, in which infectious disease is quickly detected and stopped, in which energy, water, and other resources are no longer wasted."¹² The other side has found a voice in Shoshanna Zuboff. She believes that Pentland's program undermines the "freedom of the will" which is the "existential bone structure" of a civilization that "favors the dignity of the individual."¹³ In her view, any "per-

10 Mario J. Rizzo and Glen Whitman, *Escaping Paternalism: Rationality, Behavioral Economics, and Public Policy* (Cambridge, New York: Cambridge University Press, 2019).

11 For instance, Markus Gabriel, *Ich ist nicht Gehirn: Philosophie des Geistes für das 21. Jahrhundert* (Berlin: Ullstein Ebooks, 2015), 294f, seems to suggest freedom is only constrained if influence is undesirable. He argues that a 'nasty list' (German: "fiese Liste") of determining factors (or 'conditions') would have to outweigh a 'nice list' (German: "freundliche Liste") in order to undermine freedom. Interestingly, Gabriel's 'nasty list' emphasizes factors that are beyond conscious control, creating ambiguity as to whether it is the ' nastiness' or the absence of conscious control which undermines freedom.

12 Pentland, *Social Physics*, 18.

13 Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (London: Profile Books, 2019), 331.

son, idea, or practice that breaks these bones and tears this flesh robs us of a self-authored and we-authored future.”¹⁴

In this article we will argue that in order to resolve this tension and develop a robust conception of digital humanism, we need a scientific update of our understanding of free will. To make this argument, we will focus on how increasingly sophisticated forms of *digital nudging* are emerging without any intent, but with profound consequences. We are of course aware that conjoining three such complex topics (digital nudging, digital humanism, and free will) inevitably implies that their individual treatment must remain unsatisfactorily brief. However, the present argument depends on their conjunction.

The article will proceed as follows. First, it will outline how an onslaught of increasingly sophisticated digital nudging is challenging widespread views on humanism and free will (section 2). Second, it will address the normative question, i.e. whether (de)emphasizing free will undermines or furthers humanist goals (section 3). Third, it will turn to the positive question, examining how the millennia old debate on whether (or in which way) free will actually exists is informed by recent developments in science and philosophy (section 4). Fourth, it will outline an emerging scientifically updated model of free will as a *temporally recursive two-stage model*, including a simple formalism (section 5). Finally, we draw conclusions, highlighting that this model does not only help to resolve current controversies on digital humanism and free will, but also aligns well with traditional humanism (section 6).

2 Digital Nudging and Free Will

An ongoing surge of increasingly sophisticated *digital nudging* is creating problems for widespread views on free will. Section 2.1 introduces the concept of digital nudging and outlines why it merits our attention. Section 2.2 outlines why we must expect digital nudging to increasingly occur without explicit intention, forcing us to revisit naïve conceptions of free will.

2.1 Digital Nudging

We begin by defining the concept of *digital nudging*, as presently used. Note that our focus on digital nudging is for illustrative purposes. It is not meant to imply that the argument could not similarly apply to other techniques or factors which subcon-

14 Zuboff, *The Age of Surveillance Capitalism*, 331.

sciously affect our thoughts and choices, in both digital and non-digital realms.¹⁵ The term “nudge,” popularized by Thaler and Sunstein, refers to techniques to subtly influence our behavior in ways that do not draw on our conscious, rational capacities.¹⁶ While a nudge need not act (entirely) outside of conscious perception, the definition of nudging implies that a fully conscious, rational choice would not be altered by its presence or absence.

It is worth noting that the precise definition of the term nudge is contested.¹⁷ A frequent understanding is for instance that nudging should only refer to practices that are both intentional and beneficent.¹⁸ Here we follow Thaler and Sunstein, who require neither beneficence nor intention. They state explicitly that nudging can be used “for good or evil”¹⁹ and that “of course [...] some nudges are unintentional.”²⁰ The notion that nudging can be unintentional of course raises difficult terminological questions, which we will return to in section 2.2. But first, we want to motivate why a particular focus on *digital* nudging is warranted.

Digital nudging arguably represents a qualitative shift over previous non-digital forms of nudging due to a combination of i) big data, ii) machine learning and/or artificial intelligence, and iii) increasingly digital lives. This combination furnishes a continuous ratchet of iterative improvement, whereby increasingly powerful algorithms can harness increasingly precise knowledge about individuals and apply this knowledge in increasing portions of our lives to influence our decisions. Consider this illustrative example from Calo, “[a]n obese person trying to avoid snacking between meals could receive a text on his phone from the nearest donut shop exactly

15 See e.g. Chris van Tulleken, *Ultra-Processed People: The Science Behind Food That Isn't Food* (New York: W. W. Norton & Company, 2023), for how the chemical composition of foods reflects engineered nudging.

16 Richard H. Thaler and Cass R. Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (New York, London: Penguin Books, 2008), <https://psycnet.apa.org/record/2008-03730-000>.

17 See Stuart Mills, “Nudge/Sludge Symmetry: On the Relationship between Nudge and Sludge and the Resulting Ontological, Normative and Transparency Implications,” *Behavioural Public Policy* 7, no. 9 (2020): 309–32, <https://doi.org/10.1017/bpp.2020.61>; Viktor Ivanković and Bart Engelen, “Market Nudges and Autonomy,” *Economics & Philosophy* 40, no. 1 (2022): 138–65, <https://doi.org/10.1017/S0266267122000347>.

18 This is likely because Thaler and Sunstein also developed the concept of “libertarian paternalism” to describe a subset of nudges which conform to these two properties. See Cass R. Sunstein and Richard H. Thaler, “Libertarian Paternalism Is Not an Oxymoron,” *University of Chicago Law Review* 70, no. 4 (2003): 1159–202, <https://doi.org/10.2307/1600573>.

19 Richard H. Thaler and Cass R. Sunstein, *Nudge: The Final Edition* (London: Penguin Books, 2021), 64.

20 Thaler and Sunstein, *Nudge: The Final Edition*, 14.

when he was least likely to resist.”²¹ In other words, an algorithm can collect myriad data-points, to predict precisely when and how humans in general, but also each of us individually, can be maximally influenced.²² Zuboff describes another example, where digital nudging was used to literally “herd” unwitting humans towards “sponsored locations.”²³ Of course, the same technologies can also be used to promote healthy or beneficial behaviors with unprecedented efficiency.²⁴ The key point is that such algorithms will increasingly ‘know us better than we know ourselves,’ and are using this knowledge to ‘guide’ our behavior in ways we often do not fully control or understand.

Digital nudges are clearly already affecting the everyday thoughts, beliefs, and behaviors of billions of citizens around the world.²⁵ Consider for instance, the *recommender systems* of Google, YouTube or any social media site.²⁶ Not only does the sequence and way in which results are presented shape what we choose, but the very set of options that enters our conscious deliberation is preselected. To borrow the words of former Google employee Tristan Harris: “a handful of technology companies through their choices will steer what a billion people are thinking today.”²⁷ To name just one widely discussed consequence: the “rise of extremist opinions and

-
- 21 Ryan Calo, “Digital Market Manipulation,” *George Washington Law Review* 82, no. 4 (2013): 995–1051.
 - 22 The dystopian future to which such technology could take us was viscerally illustrated in the short film “The Best Option”, directed by Serge Gorioli. In brief, a personal AI-assistant continuously and subtly nudges its user into a paranoid state in which it ultimately commits what seems to be a murder. Crucially, the AI at every step appears to help and empower the user to freely choose the “best option” for themselves.
 - 23 Zuboff, *The Age of Surveillance Capitalism*, 318.
 - 24 Pentland, *Social Physics*, 18.
 - 25 Fabrizio Esposito and Thaís Maciel Cathoud Ferreira, “Addictive Design as an Unfair Commercial Practice: The Case of Hyper-Engaging Dark Patterns,” *European Journal of Risk Regulation* 15, no. 4 (2024): 999–1016, <https://doi.org/10.1017/err.2024.8>. See also, Christian Montag et al., “Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of Psychological and Economic Theories,” *International Journal of Environmental Research and Public Health* 16, no. 14 (2019): 14, <https://doi.org/10.3390/ijerph16142612>. See also, Shota Ichihashi and Byung-Cheol Kim, “Addictive Platforms,” *Management Science* 69, no. 2 (2023): 1127–45, <https://doi.org/10.1287/mnsc.2022.4392>. See also, Jan Trzaskowski, “Manipulation by Design,” *Electronic Markets* 34, no. 1 (2024): 14, <https://doi.org/10.1007/s12525-024-00699-y>.
 - 26 Mathias Jesse and Dietmar Jannach. “Digital Nudging with Recommender Systems: Survey and Future Directions,” *Computers in Human Behavior Reports* 3 (2021), <https://doi.org/10.1016/j.chbr.2020.100052>.
 - 27 Tristan Harris, “How a Handful of Tech Companies Control Billions of Minds Every Day,” TED Talk, June 2016, https://Opgaver.Roedkilde-Cym.Dk/Engelsk/Engelsk%20B%20stx.%2031.%20maj%202021/Files/193760_non_fiction_4.Pdf.

behavior orchestrated by social media”²⁸ appears to be an increasingly urgent threat to modern democratic societies. Conversely, a well-functioning recommender system can nudge us towards content that is true as well as socially and individually advantageous.²⁹ The point here is not to paint digital nudging as inherently bad (or good). Rather, it is to highlight how such technology is practically playing out the manipulation argument against compatibilism. If we are compatibilists, then manipulation and nudging would appear to be just as compatible with free will as any other subconscious influence factor.

As mentioned above, a compatibilist might respond that the problem is not subconscious influence per se, but *intentional* manipulation (or nudging). The next section elaborates the problems with this response. It outlines how the process of optimizing proxy measures, like ‘click-rate’ or ‘watch-time,’ is leading to the emergence of increasingly pervasive and sophisticated digital nudges, *without requiring any underlying intention*, but with highly problematic consequences.

2.2 Proxy optimization and unintentional nudging

Already Thaler and Sunstein note that “of course [...] some nudges are unintentional.”³⁰ For us, it makes sense not to require intention in the definition of nudging for many reasons, for instance because intention can be difficult to ascertain.³¹ But of course defining the term ‘nudging’ without requiring intention also creates difficulties, as such a broad definition seems to allow to label *almost every aspect of our environment* as a nudge. The reader may spot the resonance of this difficulty with the ‘manipulation argument against compatibilism,’ outlined above.³² In practice, it may often be impossible to draw a line between what should be termed an influence factor and what a nudge. This is one reason why it is *insufficient* to exclusively focus on intentional nudging.

But there is a more fundamental reason why it is important to abandon an exclusive focus on intentional nudges than mere uncertainty. Specifically, we should expect digital nudging to *increasingly* arise without intention, as a side effect of the algorithmic optimization of *proxy measures* such as click-rates, time-on-platform or more generally profit.³³ A remarkable surge in the power of such proxy-optimizing

28 DIGHUM, *Vienna Manifesto*.

29 Pentland, *Social Physics*.

30 Thaler and Sunstein, *Nudge: The Final Edition*, 14.

31 Oliver Braganza, “Nudging Society to Consume—the Aggregate Consequences of Consumption Nudges,” *Review of Evolutionary Political Economy* (2025), <https://doi.org/10.1007/s43253-024-00144-y>.

32 Pereboom and McKenna, “Manipulation Arguments Against Compatibilism.”

33 Yohan J. John, Leigh Caldwell, Dakota E. McCoy, and Oliver Braganza, “Dead Rats, Dopamine, Performance Metrics, and Peacock Tails: Proxy Failure Is an Inherent Risk in Goal-Oriented

algorithms, often discussed under the heading of artificial intelligence, is likely to accelerate this development.

A proxy measure is some quantitative measure used to further an abstract goal, where the goal can itself not be quantitatively captured.³⁴ Programmers may refer to this quantitative approximation of a goal as an ‘objective function.’ In private enterprise, the proxy measure is usually some derivative of the generic economic goal to “maximize the satisfaction of *revealed preferences*.”³⁵ A large digital platform will construe a ‘click’ or a ‘minute watched’ as precisely that: a revealed preference. More generally, some measure of revealed preference is taken to approximate value to consumers (and thus ultimately profit). *Prima facie*, it is completely plausible to assume that the proxy of ‘click-rate’ will reasonably approximate consumer interest, or that the proxy of ‘watch-time’ approximates consumer value. After all, are consumers not ‘freely choosing’ to click or to continue watching?

The alternative view is that proxy-optimization *per se* has a tendency to result in “proxy failure,”³⁶ which in the present example entails the algorithmic emergence of digital nudging. In this view, proxy-optimization spontaneously gives rise to “addictive” features which “hack” our decision systems,³⁷ effectively undermining free autonomous choice. For instance, social-media sites systematically hijack our biological attention and reward system, e.g. through the presentation of bright colors,³⁸ extreme content,³⁹ or curated social feedback.⁴⁰ Crucially, these features will result simply from the algorithmic optimization of what people ‘reveal’ to prefer, without requiring human intention. To again use the words of former Google employee Tristan Harris: “Technology [...] becomes this race to the bottom of the brain stem.”⁴¹

Systems,” *Behavioral and Brain Sciences* 47, no. e67 (2024): 1–56, <https://doi.org/10.1017/S0140525X23002753>; Oliver Braganza, “Proxyeconomics, a Theory and Model of Proxy-Based Competition and Cultural Evolution,” *Royal Society Open Science* 9, no. 2 (2022): 1–41, <https://doi.org/10.1098/RSOS.211030>.

34 John et al., “Dead Rats, Dopamine, Performance Metrics, and Peacock Tails.”

35 Braganza, “Nudging Society to Consume.”

36 John et al., “Dead Rats, Dopamine, Performance Metrics, and Peacock Tails.”

37 Pier Luigi Sacco, “It’s the Biology, Stupid! Proxy Failures in Economic Decision Making,” *Behavioral and Brain Sciences* 47 (2024), <https://doi.org/10.1017/S0140525X23002947>.

38 Claire E. Robertson et al., “Changing the Incentive Structure of Social Media May Reduce Online Proxy Failure and Proliferation of Negativity,” *Behavioral and Brain Sciences* 47 (2024) <https://doi.org/10.1017/S0140525X23002935>.

39 Marc Faddoul, Guillaume Chaslot, and Hany Farid, “A Longitudinal Analysis of YouTube’s Promotion of Conspiracy Videos,” *arXiv*, March 6, 2020, <http://arxiv.org/abs/2003.03318>.

40 Ana da Silva Pinho, Violeta Céspedes Izquierdo, Björn Lindström, and Wouter van den Bos, “Youths’ Sensitivity to Social Media Feedback: A Computational Account,” *Science Advances* 10, no. 43 (2024), <https://doi.org/10.1126/sciadv.adp8775>.

41 Harris, “How a Handful of Tech Companies Control Billions of Minds Every Day.”

How exactly can a non-intentional mechanism lead to sophisticated nudges? As developed in Braganza (2025), the algorithmic optimization of, say, click-rates is sufficient in itself to lead to a cultural evolutionary⁴² discovery process, identifying which ‘design features’ are most efficient at eliciting clicks. A design feature may simply be the sequence in which content is presented, or it may be the context or visual presentation of content. The term ‘cultural evolution’ is a technical term, describing how the iterative application of selection and variation of some set of practices (or digital interface features) can lead to the emergence of apparent design without a designer, or *seeming intention* without actual intention. Dawkins famously used the metaphor of a “blind watchmaker,” to convey this remarkable phenomenon in biology.⁴³ Others have outlined why similar mechanisms can occur even faster in a cultural sphere, where variation and selection can take place on much shorter timescales and be infused with actual human creativity.⁴⁴ Since selection, in the present case, is based on proxy measures like ‘click rate,’ this process will give rise to ever more sophisticated digital nudging whenever this increases the relevant proxies.

This is not to say that digital nudges are never intentionally designed or that the question “who does the nudging” is not ethically relevant.⁴⁵ We of course still need to be wary of bad-actors who intentionally manipulate us. We should also hold actors responsible when they fail to mitigate entirely foreseeable adverse consequences (conspiracy spirals come to mind). But the key point is that exclusively focusing on intentional nudging can be a dangerous distraction in a digital context. The effect of digital technologies on our autonomy, behavior, and welfare does not hinge on whether they came about by evil intention or some other mechanism. Yet gauging the effect of digital technologies on our autonomy, behavior, and welfare is precisely what digital humanism must accomplish.

42 Cultural evolution can be thought of as a form of distributed computation. See Paul E. Smal-dino, and Peter J. Richerson, “Human Cumulative Cultural Evolution as a Form of Distributed Computation,” in *Handbook of Human Computation* (New York: Springer, 2013), https://doi.org/10.1007/978-1-4614-8806-4_76. Note that recognizing an explanatory potential of evolution-like processes in culture in no way implies that all of culture can be explained by such processes. See Gabriel, *Ich ist nicht Gehirn*, 215.

43 Richard Dawkins, *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design*, reissue Edition (New York: W. W. Norton & Company, 2015).

44 Joseph Patrick Henrich, *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter* (Princeton, New Jersey: Princeton University Press, 2016); Alex Mesoudi, and Alex Thornton, “What Is Cumulative Cultural Evolution?” *Proceedings of the Royal Society B: Biological Sciences* 285, no. 1880 (2018), <https://doi.org/10.1098/rspb.2018.0712>.

45 When presenting these ideas, one participant of the *Humanism and AI* conference (Center for Science and Thought, University of Bonn, May 23–24, 2024) (RPH) suggested that the most relevant question is “who does the nudging.”

In summary, digital nudging is of steadily increasing relevance in modern societies. It infuses age-old questions about how free our decisions are or should be with renewed practical relevance. Digital technologies, in particular, are highlighting that traditional stances which simply condemn *intentional* manipulation are insufficient. It is because of this that age-old philosophical questions about free will, such as whether we should (de)emphasize it (section 3) and whether it actually exists (sections 4,5) require renewed attention.

3 Digital Humanism and Free will

Next, we collect *normative* arguments suggesting that we either should (or should not) emphasize free will in order to further humanist goals. Before we do so, we should at least briefly outline what we are referring to when speaking of “digital humanism.” Digital Humanism has been summarized as a call to shape digital technologies to *serve* humans in accordance with traditional humanist and enlightenment values.⁴⁶ These values include, in particular, respect for the “dignity of every human being” and an affirmation “that human beings have the right and responsibility to give meaning and shape to their own lives.”⁴⁷ The call seems to, above all, be a response to the observation that current digital technology is undermining such values. The *Vienna Manifesto on Digital Humanism* opens with an unambiguous quote by the web’s founder, Tim Berners Lee: “The system is failing.” For example, the “formation of filter bubbles and echo chambers” and “the rise of extremist opinions and behavior,”⁴⁸ are diagnosed to threaten the very foundations of democratic society. As noted above, these kinds of diagnoses have been met with two starkly different responses, which align in their ‘humanist’ goal to improve the human lot, but differ starkly on whether this goal requires to emphasize or deemphasize free will. In the following two sections we review these *normative* arguments. Section 3.1 collects claims made in the literature that emphasizing free will harms humanist goals, while section 3.2 reviews the case that emphasizing free will is crucial to humanist goals.

Importantly, the normative arguments presented here are *not* meant to address the positive question whether (or in which way) free will actually exists. We treat the positive question separately in sections 4 and 5. We highlight this because, within

46 Hannes Werthner, Erich Prem, Carlo Ghezzi, et al., eds., *Introduction to Digital Humanism* (Cham: Springer, 2024), <https://doi.org/10.1007/978-3-030-86144-5>.

47 Julian Nida-Rümelin, and Dorothea Winter, “Humanism and Enlightenment,” in *Introduction to Digital Humanism*, ed. Hannes Werthner et. al. (Cham: Springer 2024), 6, <https://doi.org/10.1007/978-3-031-45304-5>.

48 DIGHUM, *Vienna Manifesto*.

the debate, it seems almost obligatory to accuse one's opponents of doing precisely that—concluding on the actual existence of free will based on its desirability. Skeptics, for instance, argue that belief in free will is often driven by “emotional reasons that reflect that there being no free will is pretty damn unsettling.”⁴⁹ Conversely, free will defenders might argue that skeptics are driven by a desire to relinquish responsibility for their own actions and lives.⁵⁰ So it is worth reemphasizing that the following two sections are in no way meant to address the positive question.

3.1 The Humanist Case Against Free Will

Several authors have argued or implied that we *should* partially⁵¹ or fully⁵² reject free will in order to further humanist goals.⁵³ In this view, we should focus on i) quantitatively studying the determinants of behavior, and ii) harnessing such knowledge to foster compassion and improve human lives. The normative argument is supported in at least four ways.

First, understanding subconscious determinants of behavior (i.e. limitations of free will), can open up unused potential for social organization and reduce unnecessary harm and suffering. E.g., Pentland believes that “social physics” will allow to “build a world without war or financial crashes, in which infectious disease is quickly detected and stopped, in which energy, water, and other resources are no longer wasted.”⁵⁴ He offers empirical evidence of, e.g., how digital nudging can be used to improve individual financial choices by “tuning” social network interactions, with the positive side effect of decreasing systemic financial risk.⁵⁵ In this example, “tuning” entailed nudging some individuals towards more diverse social contacts, in order to “break up the echo chamber.”⁵⁶ Other examples include a famous Face-

49 Sapolsky, *Determined: Life Without Free Will*, 7.

50 See Gabriel, *Ich ist nicht Gehirn*, 113.

51 Pentland, *Social Physics*.

52 Sapolsky, *Determined: Life Without Free Will*.

53 Note that free will skeptics will often suggest that their main interest is positive. See Sapolsky, *Determined: Life Without Free Will*, 9. They seemingly follow a broader Zeitgeist of “reductionism,” “physicalism,” and “neurocentrism,” which generally delegates normative questions and the “humanities” to the sidelines. See Gabriel, *Ich ist nicht Gehirn*.

54 Pentland, *Social Physics*, 18.

55 Ibid. 38.

56 Ibid.

book study by Bond and colleagues,⁵⁷ in which subtle digital nudging was used to significantly increase voter turnout.⁵⁸

A second prominent normative argument against overemphasizing free will is that it undermines compassion and justifies inhumane punishment in criminal justice.⁵⁹ In this context, new scientific knowledge about determinants of behavior can directly translate to a more humane treatment of offenders (and often already does). In fact, Sapolsky's impassioned argument against free will was, in part, motivated by his having "worked with public defenders on various murder trials, teaching juries about the circumstances that produce brains that make horrible decisions."⁶⁰ Recognizing limits to free will in the context of digital nudging can directly foster empathy and compassion in criminal justice.⁶¹ More importantly, it shifts the focus from blame and punishment to understanding and prevention.

A third key argument against overemphasizing free will applies the same reasoning to social justice. For instance, Dennett and Caruso argue that emphasis on free choice is routinely used to "to excuse and perpetuate social and economic inequalities."⁶² Those on top of society like to believe that their own free actions and their intrinsic merit led to their position and that those who are less fortunate are also less deserving. However, it is well established that countless social ills, including teenage pregnancy, obesity, criminal offence rates, addiction, etc. are more likely for those born into poverty.⁶³ To name just one salient example, there is clear neuroscientific evidence that the brain areas enabling us to control problematic impulses are less developed in adults which suffered from childhood deprivation.⁶⁴ It is un-

57 Robert M. Bond, Christopher J. Fariss, Jason J. Jones, et al., "A 61-Million-Person Experiment in Social Influence and Political Mobilization," *Nature* 489, no. 7415 (2012): 295–98, <https://doi.org/10.1038/nature11421>.

58 It is of course difficult to overstate the potential for abuse of such technologies, as rightly pointed out by critics such as Zuboff, *The Age of Surveillance Capitalism*. But this does not in itself contradict that there is also massive potential to increase welfare and reduce suffering.

59 Gregg D. Caruso, "Rejecting Retributivism: Free Will, Punishment, and Criminal Justice, A Précis," *Journal of Practical Ethics* 10, no 2 (2023): 2, <https://doi.org/10.3998/jpe.3544>.

60 Sapolsky, *Determined: Life Without Free Will*, 383.

61 See for example *ibid.*, 381.

62 Dennett and Caruso, "On Free Will."

63 Richard G. Wilkinson, and Kate Pickett, "The Spirit Level: Why More Equal Societies Almost Always Do Better," *Allen Lane* 330 (2009).

64 Gregg D. Caruso, *Public Health and Safety: The Social Determinants of Health and Criminal Behavior* (UK: ResearchersLinks Books, 2017). A key region adversely affected by childhood deprivation is the prefrontal cortex, one of the evolutionarily most novel brain structures. It is now well established that the prefrontal cortex is the region responsible for impulse control, meaning affected individuals – through no fault of their own – will find it much more difficult to guide their own emotions and impulses. Such knowledge undermines inhumane retributivism and politically motivates measures to combat childhood poverty.

controversial that digital nudging can compound social problems by, e.g., fostering addictive online gambling.⁶⁵ Emphasizing limits to free will can thus move us away from attributing (potentially unjustified) blame and praise, and towards empathy and social justice.

A fourth, closely related argument highlights that, an emphasis on free will is routinely used to justify manipulative corporate practices.⁶⁶ There are, for instance, intense ongoing debates between public-health and free-market proponents, about whether increasingly unhealthy food choices are freely made or the result of powerful nudging.⁶⁷ Public health officials may present quantitative evidence of how some type of food or marketing surreptitiously nudges individuals towards obesity and addiction⁶⁸ and free market proponents respond by insisting that consumers *freely chose*, and that any government interference with this free choice is to be resisted.⁶⁹ Emphasizing free consumer choice is also the central argument in defenses of advertising more generally,⁷⁰ including for tobacco,⁷¹ and can be done to cast even addictive chain-smoking as an expression of free will.^{72,73}

65 Javier Mora-Salgueiro, Aitana García-Estela, Bridget Hogg, et al., "The Prevalence and Clinical and Sociodemographic Factors of Problem Online Gambling: A Systematic Review," *Journal of Gambling Studies* 37, no 3 (2021): 899–926, <https://doi.org/10.1007/s10899-021-09999-w>.

66 See for example Jerry Kirkpatrick, *In Defense of Advertising: Arguments from Reason, Ethical Egoism, and Laissez-Faire Capitalism* (Claremont, CA: Kirkpatrick Books, 1994), <https://philpapers.org/rec/KIRIDO>.

67 Van Tulleken, *Ultra-Processed People*; Braganza, "Nudging Society to Consume."

68 Deborah A. Cohen, "Neurophysiological Pathways to Obesity: Below Awareness and beyond Individual Control," *Diabetes* 57, no 7 (2008): 1768–73, <https://doi.org/10.2337/db08-0163>; Matthias Blüher, "Obesity: Global Epidemiology and Pathogenesis," *Nature Reviews Endocrinology* 15, no. 5 (2019): 288–98, <https://doi.org/10.1038/s41574-019-0176-8>; Kevin D. Hall, "Did the Food Environment Cause the Obesity Epidemic?" *Obesity* 26, no. 1 (2018): 11–13, <https://doi.org/10.1002/oby.22073>.

69 Kirkpatrick, *In Defense of Advertising*; Rizzo and Whitman, *Escaping Paternalism*. For instance, Kirkpatrick, *In Defense of Advertising*, 23, argues that the entire reasoning of critics "denies [...] free will" and should thus be discounted.

70 Kirkpatrick, *In Defense of Advertising*.

71 David Stuckler et al., "Manufacturing Epidemics: The Role of Global Producers in Increased Consumption of Unhealthy Commodities Including Processed Foods, Alcohol, and Tobacco," *PLoS Medicine* 9, no. 6 (2012): e1001235, <https://doi.org/10.1371/journal.pmed.1001235>.

72 Gabriel, *Ich ist nicht Gehirn*, 289.

73 The issue is of course complex, because we can always assume that an addict would also choose addiction if he/she were not influenced by free-will-undermining addictive processes. Not coincidentally, this is precisely what tobacco companies (or any other company charged with consumer manipulation) argues. See Braganza, "Nudging Society to Consume."

Viewing the last argument in the context of environmentally problematic *consumption nudges*⁷⁴ implies that overemphasizing free will (and underestimating the role of nudging) is contributing to one of the most fundamental threats humans have ever faced, namely climate change.⁷⁵ Pervasive consumption nudging is arguably compounding the unfolding climate crisis by surreptitiously influencing consumers under an illusory pretense of autonomous choice.⁷⁶ Simultaneously, powerful digital misinformation campaigns are nudging consumers (and voters) to disbelieve or disregard scientific evidence.⁷⁷ Not coincidentally, insisting that overconsumption and climate skepticism is the result of *free choice* has become the central talking point of those seeking to undermine climate action.⁷⁸ This suggests that a better appreciation of the limits of free will might be essential to safeguard against manipulation and build a society that supports genuine liberty in the long term.⁷⁹

In summary, recognizing the limits of free will can lead us to “prioritize prevention and social justice.”⁸⁰ Indeed, recognizing subconscious influence factors seems to be a prerequisite for fact-based social discourse concerning issues such as obesity or climate change (or election integrity⁸¹). Paradoxically, deemphasizing free will (i.e. acknowledging influence) would appear to be a prerequisite to genuinely free and fact-based choice.

74 Consumption nudging, in its various shapes, is of course increasingly digital. See Vivian Frick et al., “Does online advertising stimulate overconsumption?” *Ökologisches Wirtschaften – Fachzeitschrift* 37, no. 1 (2022), <https://doi.org/10.14512/OEW370146>; Kathie M. d’I. Treen et al., “Online Misinformation about Climate Change,” *WIREs Climate Change* 11, no. 5 (2020): e665, <https://doi.org/10.1002/wcc.665>.

75 Andrew Simms, and Leo Murray, *Badvertising: Polluting Our Minds and Fuelling Climate Chaos* (London: Pluto Press, 2023).

76 Oliver Braganza, and Jakob Kapeller, “Reappraising Consumption Nudging – on Liberty in the Age of Climate Catastrophe,” *Humanities and Social Sciences Communications* (2025).

77 Erik M. Conway, and Naomi Oreskes, *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming* (New York: Bloomsbury Press, 2011); Li Quian Tay, and K. H. Ecker, “Continued Influence of Disinformation and the Information Disorder,” in *The Remaking of Memory in the Age of the Internet and Social Media*, ed. Qi Wang and Andrew Hoskins (New York: Oxford University Press, 2024).

78 Michael E. Mann, *The New Climate War: The Fight to Take Back Our Planet* (London: Scribe Publications, 2022).

79 Braganza, and Kapeller, “Reappraising Consumption Nudging.”

80 Caruso, “Rejecting Retributivism.”

81 On January 7, 2025 (one day prior to the writing of these lines), Meta, one of the largest social media companies worldwide has announced that it will stop working with fact checkers in order to “empower the community to decide” what is true. Of course, ensuring factual accuracy limits the possibility for digital nudging (say via inaccurate posts on election fraud) and “empowering the community to decide” is an appeal to *free choice*. See Joel Kaplan, “More Speech and Fewer Mistakes,” *Meta Newsroom*, January 7, 2025, <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>.

3.2 The Humanist Case for Free Will

On the other side, many scholars emphasize that the concept of free will is, and must be, central to any viable form of humanism. Zuboff argues that “freedom of will is the existential bone structure” for a civilization that “favors the dignity of the individual.”⁸² In this view, any “person, idea, or practice that breaks these bones and tears this flesh robs us of a self-authored and we-authored future.”⁸³ Nida-Rümelin and Winter highlight that the focus on free choice is in fact what “unites the various humanist perspectives and movements,” specifically the “concept of human authorship and the related human capacity to give reasons, freedom, and responsibility.”⁸⁴ To reiterate: reviewing normative arguments (why emphasizing free will might serve humanist goals) is *not* meant to address the positive question (whether or in which way free will actually exists), which we treat separately in sections 4 and 5. In the following we present four reasons why humanists insist on an emphasis of free will.

First, free will is part of our lived experience and self-image. Our conscious experience is one of continuous meaningful choices. Rejecting free will would force us to experience a constant state of logical self-contradiction and deflated meaning.⁸⁵ We would have to continue to make choices while at the same time viewing these choices as completely illusory and utterly meaningless. Free will deniers argue that we would grow used to such a state of constant performative self-contradiction and deflated meaning simply because we would have to.⁸⁶ But such a state of being certainly seems to directly conflict with humanist goals, which emphasize “that human beings have the right and responsibility to give meaning and shape to their own lives.”⁸⁷

Second, rejecting free will (and, with it, moral responsibility) would undermine the foundation of almost all of our social life. In the words of Mitchell: “The stakes here could not be higher. The idea of moral responsibility is the foundation not only of our legal systems but also of all our social interactions.”⁸⁸ The worry is that re-

82 Zuboff, *The Age of Surveillance Capitalism*, 331.

83 Ibid.

84 Nida-Rümelin, and Winter, “Humanism and Enlightenment,” 8.

85 See, for example, Gabriel, *Ich ist nicht Gehirn*, 302. For instance, when Sapolsky argues we should reject free will, then he is appealing to our conscious self to choose to do so – he is providing arguments for our conscious deliberation to convince us that arguments cannot affect conscious deliberation in any meaningful way. See Sapolsky, *Determined: Life Without Free Will*.

86 Susan Wolf, “The Importance of Free Will,” *Mind* 90, no. 359 (1981): 386–405; Sapolsky, *Determined: Life Without Free Will*.

87 See for example Nida-Rümelin, and Winter, “Humanism and Enlightenment,” 6.

88 Kevin J. Mitchell, *Free Agents* (Princeton: Princeton University Press, 2023), 17, <https://press.princeton.edu/books/hardcover/9780691226231/free-agents>.

jecting free will might undermine the fabric of society, which could destabilize current institutions and cause widespread human suffering. And this would only be one manifestation of a more fundamental psychological shift. Susan Wolf argues that abandoning free will would “change drastically or, as most would say, reduce the quality of our involvement or participation in all our human relationships.”⁸⁹ Human interactions would be recast as “bits of positive and negative reinforcement meted out in the hopes of altering the character of others in ways best suited to our needs.”⁹⁰

Third, rejecting free will would imply that objecting to digital nudging or manipulation solely on the grounds that it subverts free choice⁹¹ is incoherent. The problem resembles the one in the manipulation argument against compatibilism, since an outright free-will denier, just like a compatibilist, sees all actions as nothing but the result of subconscious determining factors. If protecting individual autonomy and conscious self-determination is a humanist goal, then this is clearly undesirable. This is arguably a key reason why Zuboff places an uncompromising belief in free will as an essential prerequisite for a “self-authored we-authored future,”⁹² i.e. a future that is not defined by pervasive digital nudging.

Fourth, rejecting free will (and the broader underlying reductionist worldview) undermines core humanist concepts such as *human dignity* and *value*. Lee and George make the link between free will and human dignity explicit: “[W]hat makes human beings *persons* rather than *things* [...] are their] capacities for conceptual thought, deliberation, and free choice, that is, the natural capacity to shape their own lives.”⁹³ Frankfurt makes a similar point, arguing that what makes us *persons* rather than animals or machines is free will (or what he calls “volition”).⁹⁴ In fact, even Sapolsky tacitly acknowledges that, without free will, humans are nothing more than machines, devoid of intrinsic meaning or value.⁹⁵ It clearly conflicts with humanism to conceive of ourselves and others as lacking inherent dignity and value.

It seems that, notwithstanding the arguments presented in section 3.1, fully rejecting free will undermines the very foundations of humanism: First, the idea that

89 Wolf, “The Importance of Free Will.”

90 Ibid.

91 Rizzo and Whitman, *Escaping Paternalism*.

92 Zuboff, *The Age of Surveillance Capitalism*, 331.

93 Patrick Lee, and Robert P. George, “The Nature and Basis of Human Dignity,” *Ratio Juris* 21, no. 2 (2008): 173–93, <https://doi.org/10.1111/j.1467-9337.2008.00386.x>.

94 Harry C. Frankfurt, “Freedom of the Will and the Concept of a Person,” *The Journal of Philosophy* 68, no. 1 (1971): 5–20, <https://doi.org/10.2307/2024717>.

95 Sapolsky admits that it “is logically indefensible, ludicrous, meaningless to believe that something ‘good’ can happen to a machine”, but that to his “intellectual shame” he cannot overcome the “illogic” that it is “good if people feel less pain and more happiness.” See Sapolsky, *Determined: Life Without Free Will*, 392.

our conscious deliberations affect our choices and that these choices can and should shape our lives. Second, the idea that human lives have value, that every person has inherent dignity and that we should act accordingly. Sapolsky's admission in this respect is remarkable. Even while he passionately argues that relinquishing our belief in free will can lead to a more humane society, he accepts this goal is utterly meaningless, because humans are in essence nothing more than machines.⁹⁶

4 Towards a Scientific Understanding of Free Will

Above, we have outlined several *normative* arguments brought forth to either emphasize or deemphasize free will, i.e. assertions that either one or the other will promote humanist values. The attentive reader might also have noted that many of the arguments are not necessarily in conflict. One side mostly warns against attributing free will where it is illusory, while the other mostly warns against assuming that it is *nothing but* an illusion.

But, regardless of these normative considerations, the crucial question is arguably the positive one, namely whether or in which way free will actually exists. Unfortunately, the positive question seems at least as fraught as the normative one. Whether or not free will exists, or is even a coherent concept, is notoriously controversial. Despite over 2000 years of intense debate,⁹⁷ book-length treatments with diametrically opposed conclusions continue to be published until the present.⁹⁸

In the following two sections we will try to briefly outline what has likely impeded understanding in the past and lay the groundwork to introduce the temporally recursive two-stage model described in section 5. Section 4.1 suggests that much of the philosophical debate is plagued by a false dichotomy between determinism and indeterminism which systematically obscures phenomena that occur at the *interface* of deterministic and indeterministic processes. Section 4.2 suggests that, particularly within the natural sciences, unjustified metaphysical convictions in ontological and epistemological reductionism have obstructed understanding. Importantly, both sections are strictly naturalistic in the sense of requiring “no

96 Note that Sapolsky's assault on meaning is far more fundamental than that of the existentialists. For instance, Sartre accepts the notion of a meaningless universe but asserts meaning and humanism to derive precisely from the human freedom to make choices and shape one's own life. See Jean-Paul Sartre, *Existentialism Is a Humanism* (New Haven: Yale University Press, 1996).

97 Prominent historical debates about free will were for instance between Epicurus (pro) and Democritus (contra) in antiquity or between Erasmus of Rotterdam (pro) and Martin Luther (contra) at the beginning of modernity.

98 Mitchell, *Free Agents*; Sapolsky, *Determined: Life Without Free Will*; Gabriel, *Ich ist nicht Gehirn*.

miracles”⁹⁹: no uncaused causes or noumenal selves and nothing that defies any physical laws.

4.1 The False Dichotomy

The contemporary debate is shaped by a central dichotomy between determinism and indeterminism, as illustrated by what has become known as *the standard argument against free will*.¹⁰⁰ It is “the very simple and logical claim that *either* determinism or indeterminism is true. If determinism is ‘true,’ we are not free, if indeterminism is ‘true,’ we are not responsible.”¹⁰¹ These two problems have been called the *problem of determinism* and the *problem of luck* respectively.¹⁰² The problem of determinism states that if all our actions are predetermined, then how could they possibly be labeled ‘free.’ The problem of luck states that if our actions are in any way due to chance (i.e. indeterminism), then how could this possibly be labelled as ‘willed.’

Typical arguments proceed by *sequentially* considering these problems, and drawing conclusions based on whether either or both seem insurmountable.¹⁰³ For instance, some compatibilists seem to begin by intuiting that *the problem of luck* is insurmountable¹⁰⁴ and then proceed to deal with the *problem of determinism*. They might, for instance, argue that even if the lower level is fully deterministic, it still makes sense to speak of free will at the level of cognition and social interactions due to some kind of emergence.¹⁰⁵ Sapolsky instead follows the sequential script of the

99 A naturalistic explanation should require “no miracles”, i.e. it should violate no physical laws or require any supernatural phenomena. See David Papineau, “Naturalism,” *The Stanford Encyclopedia of Philosophy* (Fall 2023), ed. Edward N. Zalta and Uri Nodelman, <https://plato.stanford.edu/archives/fall2023/entries/naturalism/>. The present account is ontologically naturalistic in the sense that it rejects the supernatural. But it simultaneously furnishes epistemological anti-naturalism within the humanities – the reductive approach of the natural sciences has clear limits. See Markus Gabriel, *Fields of Sense: A New Realist Ontology* (Edinburgh: Edinburgh University Press, 2015), 14.

100 Galen Strawson, “Free Will,” in *Routledge Encyclopedia of Philosophy* (London, New York: Routledge, 1998), <https://doi.org/10.4324/9780415249126-VO14-1>.

101 Robert O. Doyle, “The Two-Stage Solution to the Problem of Free Will,” in *Is Science Compatible with Free Will?* ed. Antoine Suarez and Peter Adams (New York: Springer, 2013), 237.

102 Mark Wulff Carstensen, “Freedom and Chance: Towards a Graded Approach to Free Will,” (PhD diss., Ludwig-Maximilians-Universität München, 2021), https://edoc.ub.uni-muenchen.de/28639/1/Carstensen_Mark_Wulff.pdf.

103 Strawson, “Free Will.”

104 Gabriel, *Fields of Sense*, 281. And Gabriel argues for a “compatibilist” account which is based on an understanding of determinism where not everything is determined, just as the model fleshed out below suggests. See Gabriel, *Ich ist nicht Gehirn*.

105 Marij van Strien, “The Vienna Circle against Quantum Speculations,” *HOPOS: The Journal of the International Society for the History of Philosophy of Science* 12, no. 2 (2022): 359–94, <https://doi.org/10.1086/721137>; Dennett, *Elbow Room*.

standard argument to ultimately reject free will.¹⁰⁶ The sequential consideration of these arguments leads to, in Strawson's terms, a "great difficulty [...] to explain why the falsity of determinism is any better than the truth of determinism when it comes to establishing our free agency and moral responsibility."¹⁰⁷

What is remarkable about the debate is that physics, since the beginning of the 20th century, has suggested that the world is a complex mesh of *both deterministic and genuinely random processes*, which continually interact.¹⁰⁸ Briefly, physical systems at very small (specifically "quantum"¹⁰⁹) scales are considered fundamentally indeterministic—they can be described only by probability or density distributions. Furthermore, quantum indeterminacy must be expected to continually translate into (classical) stochasticity through what is known as decoherence.¹¹⁰ Tegmark explicitly advises neuroscientists to incorporate quantum indeterminacy into deterministic simulations "by including a random number generator in the simulation."¹¹¹ The upshot is that macroscopic physical systems behave largely like the deterministic systems of classical physics, but they are infused with a continuous stream of true quantum randomness, which interacts with e.g. thermal chaotic processes (see also appendix 1).

We will see how this helps to resolve Strawson's "great difficulty" in section 5. The key point of this section is to highlight that the sequential analytical approach predominant in the literature systematically obscures any phenomenon that occurs at the *interface* of deterministic and indeterministic processes.

4.2 The Reductionist Fallacy

Another fundamental obstacle to understanding free will, in our view, is a pervasive uncompromising metaphysical dedication to reductionism, which goes far be-

106 Sapolsky, *Determined: Life Without Free Will*.

107 Strawson, "Free Will."

108 Peter Jedlicka, "Revisiting the Quantum Brain Hypothesis: Toward Quantum (Neuro)Biology?", *Frontiers in Molecular Neuroscience* 10 (2017), <https://doi.org/10.3389/fnmol.2017.00366>.

109 Note that invoking quantum processes as a source of classical stochasticity is uncontroversial within physics and has nothing to do with a family of highly controversial theories speculating about macroscopic quantum effects. See e.g. Roger Penrose et al., *The Large, the Small and the Human Mind* (Cambridge: Cambridge University Press, 2000).

110 Decoherence, in simple terms, describes the translation from a quantum system (which is described as a probability distribution or "wave") to a classical system (where indeterminacy is broken by the realization of a specific "draw" from the probability distribution, often manifesting as a "particle"). The most famous experiment demonstrating the phenomenon is the double slit experiment. See Max Tegmark, "Importance of Quantum Decoherence in Brain Processes," *Physical Review E* 61, no. 4 (2000): 4194–206, <https://doi.org/10.1103/PhysRevE.61.4194>.

111 Tegmark, "Importance of Quantum Decoherence."

yond the standard “no miracles” clause mentioned above. It is part of a broader Zeitgeist promoting a “scientific,” “reductionistic” or “neurocentric” worldview.¹¹² Reductionism seems to prevail in contemporary thought in both ontology (the metaphysical question of *what really exists*) and epistemology (the question of *how we can know about things*). We treat each in turn.

According to ontological reductionism, things at smaller or lower levels of organization are somehow *more real* than things at larger or higher levels. An ontological reductionist might for instance argue that our conscious experience or sense of self is only an illusion and what is ‘real’ is instead a pattern of neural activity.¹¹³ Because only ‘real’ things can cause other things,¹¹⁴ the reductionist concludes that higher-level patterns, such as biologically stored information or “thoughts,” cannot have causal power in the world. Any perceived causal effect is cast as ‘really’ due to lower levels. This excludes the possibility of conscious choices causing changes in the physical world, and thus free will, by metaphysical assumption.

However, the notion that lower levels are somehow ‘more real’ than higher levels of organization is increasingly being recognized as problematic, both within the sciences¹¹⁵ and philosophy.¹¹⁶ The prevalence of unquestioned ontological reductionism seems particularly remarkable among natural scientists, who often stylize themselves as agnostic concerning metaphysical issues. Note that rejecting ontological reductionism, i.e. assuming that reality has no “preferred level”, in no way implies that higher level phenomena/entities cannot in principle be re-described as a collection of lower level phenomena/entities and their relations. The question is whether this speaks to the ‘reality’ or ‘causal efficacy’ of the phenomenon.

Epistemological reductionism refers to the notion that the best way to learn about the world is by a re-description at lower-levels. The value of the epistemologically reductive approach is powerfully evidenced by the successes in the natural sciences. However, there are now entire branches of science focused on elucidating where, and why exactly, epistemological reductionism fails in practice. Broadly speaking, these fall under the name of complexity science and investigate where and why higher-level entities can have ‘emergent’ properties, i.e. properties that cannot

112 Gabriel, *Ich ist nicht Gehirn*.

113 Notably, to be coherent, the reductionist would have to go on: patterns of neural activity are not as ‘real’ as the ion-flows that underlie them; and of course ions and atoms more generally are not as ‘real’ as quantum fluctuations. *Ibid*.

114 George F. R. Ellis, and Markus Gabriel, “Physical, Logical, and Mental Top-Down Effects,” in *Top-down Causation and Emergence*, ed. Jan Voosholz and Markus Gabriel (Cham: Springer International Publishing, 2021).

115 George F. R. Ellis, and Jonathan Kopel, “The Dynamical Emergence of Biology from Physics: Branching Causation via Biomolecules,” *Frontiers in Physiology* 9 (2019), <https://doi.org/10.3389/fphys.2018.01966>.

116 Gabriel, *Ich ist nicht Gehirn*; Gabriel, *Fields of Sense*.

be understood by studying their component parts.¹¹⁷ One particularly relevant advance in the present context has been a more rigorous mathematical understanding of the concept of *causation*.¹¹⁸ It describes precisely why higher-level entities, say algorithms or thoughts, are appropriately conceived as having causal power over lower-level entities, say resistors or hands, *without violating any physical laws*.¹¹⁹ This process has variably been referred to as ‘top down’ or ‘contextual’ causation,¹²⁰ ‘criterial causation,’¹²¹ ‘downward causation,’¹²² or indeed ‘mental causation.’¹²³

Particularly within biology, the traditional reductionist paradigm, which seeks to explain life without recourse to concepts like a conscious self, agency or goal-directed behavior, is increasingly being recognized as outdated.¹²⁴ Almost none of the defining properties of living things makes any sense or has any meaning in the reductionist physical conception of the world.¹²⁵ The key insight is that living things

-
- 117 Casper Hesp et al., “A Multi-Scale View of the Emergent Complexity of Life: A Free-Energy Proposal,” *Springer Proceedings in Complexity* (2019): 195–227. https://doi.org/10.1007/978-3-030-00075-2_7; Harald Atmanspacher, “On Macrostates in Complex Multi-Scale Systems,” *Entropy* 18, no. 12 (2016): 426. <https://doi.org/10.3390/e18120426>; Gary William Flake, *The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems, and Adaptation* (Cambridge, MA: MIT Press, 1998). Also note that complexity scientists may still have a metaphysical conviction in ontological reductionism.
 - 118 Judea Pearl, and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect* (London: Penguin Books, 2019).
 - 119 Erik P. Hoel et al., “Can the Macro Beat the Micro? Integrated Information across Spatiotemporal Scales,” *Neuroscience of Consciousness*, no. 1 (2016), <https://doi.org/10.1093/nc/niw012>.
 - 120 Ellis, and Gabriel, “Physical, Logical, and Mental Top-Down Effects”; Mitchell, *Free Agents*; Raymond Noble, and Denis Noble, “Can Reasons and Values Influence Action: How Might Intentional Agency Work Physiologically?” *Journal for General Philosophy of Science* (2020): 1–19, <https://doi.org/10.1007/s10838-020-09525-3>.
 - 121 Peter U. Tse, “Two Types of Libertarian Free Will Are Realized in the Human Brain,” in *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience*, ed. Gregg Caruso, and Owen Flanagan (New York: Oxford University Press, 2018), 162–190, <https://doi.org/10.1093/oso/9780190460723.003.0010>.
 - 122 James Woodward, “Downward Causation Defended,” in *Top-Down Causation and Emergence*, ed. Jan Voosholz and Markus Gabriel (Cham: Springer International Publishing, 2021), https://doi.org/10.1007/978-3-030-71899-2_9.
 - 123 Ellis, and Gabriel, “Physical, Logical, and Mental Top-Down Effects.”
 - 124 David Haig, *From Darwin to Derrida: Selfish Genes, Social Selves, and the Meanings of Life* (Cambridge, MA: MIT Press, 2020); Oliver Braganza et al., “Teleonomy, Legibility, and Diversity: Do We Need More ‘Proxynomics’?” *Behavioral and Brain Sciences* 47 (2024): e88, <https://doi.org/10.1017/S0140525X24000050>; Kevin N. Laland et al., “The Extended Evolutionary Synthesis: Its Structure, Assumptions and Predictions,” *Proceedings of the Royal Society B: Biological Sciences* 282, no. 1813 (2015): 1–14, <https://doi.org/10.1098/RSPB.2015.1019>; Jonathan Bartlett, “Evolutionary Teleonomy as a Unifying Principle for the Extended Evolutionary Synthesis,” *BIO-Complexity*, no. 2 (2017): 1–7, <https://doi.org/10.5048/BIO-C.2017.2>.
 - 125 See for example Mitchell, *Free Agents*, xff.

can only be understood by considering the *information* stored in organisms, be it within DNA or conscious thought, and examining the ‘causal’ effects this information has on the world.¹²⁶

In sum, ontological reductionism is nothing more than a metaphysical assumption, while epistemological reductionism is nothing more than an often valuable methodological heuristic. The notion that information can have causal effects in the world conflicts with both forms of reductionism but is essential to an understanding of biological systems in general and agency in particular.

5 The Temporally Recursive Two-Stage Model of Free Will

In the following we introduce the *temporally recursive two-stage model* of free will. The model and its scientific foundation is essentially that laid out in book-length by Mitchell,¹²⁷ but variants thereof have been developed and promoted by many others.¹²⁸ The present contribution is to i) provide a synopsis of this model; ii) contribute a concise formalization;¹²⁹ iii) address some frequent objections (the problem of ‘infinite regress’ and the ‘timing objection’); and iv) highlight its relevance for digital humanism.

The basic model has been called the “two-stage model” of free will and was traced back to William James by Doyle.¹³⁰ It asserts that free will requires an interaction between deterministic and indeterministic processes. Within an organism, one might also think of the ‘stages’ as ‘levels of organization,’ where a ‘higher-level’ pattern, which is relatively deterministic (say a goal representation within the genes or brain of an organism), selects from a continuous stream of ‘lower level’ randomness.

126 Ellis, and Gabriel, “Physical, Logical, and Mental Top-Down Effects”; Ellis, and Kopel, “The Dynamical Emergence.”

127 Mitchell, *Free Agents*.

128 Doyle, “The Two-Stage Solution”; Björn Brembs, “Towards a Scientific Concept of Free Will as a Biological Trait: Spontaneous Actions and Decision-Making in Invertebrates,” *Proceedings of the Royal Society B: Biological Sciences* 278, no. 1707 (2010): 930–39, <https://doi.org/10.1098/rspb.2010.2325>; Martin Heisenberg, “Is Free Will an Illusion?” *Nature* 459, no. 7244 (2009): 164–65, <https://doi.org/10.1038/459164a>; Robert Kane, “The Complex Tapestry of Free Will: Striving Will, Indeterminism and Volitional Streams,” *Synthese* 196, no. 1 (2019): 145–60, <https://doi.org/10.1007/s11229-016-1046-8>; Noble, and Noble, “Can Reasons and Values Influence Action”; Tse, “Two Types of Libertarian Free Will”; Frankfurt, “Freedom of the Will.”

129 Following Hanaan Hashim, and R. Srikanth, “The Concept of Free Will as an Infinite Meta-theoretic Recursion,” *arXiv*, preprint, June 21, 2015, <http://arxiv.org/abs/1503.06485>.

130 Doyle, “The Two-Stage Solution.”

The core model is in fact better understood as a model of free agency,¹³¹ something that arguably all living things possess, and which does not require conscious reflection. Indeed, as Noble and Noble illustrate, the basic mechanism operates even within bodily subsystems, such as our immune system.¹³² Understanding how even such biological systems defy the standard assumptions of reductionism is essential. But, as argued by Mitchell,¹³³ agency only becomes something we would recognize as free will, once it enters into a recursive process of *conscious* reflection and deliberation.

5.1 Stage 1: Random Chance or 'Freedom'

In stage 1, stochastic processes are used to generate action options.

Let P_0 denote a (partially) random set of options:

$$P_0 = \{p_1, p_2, \dots, p_n\}.$$

In a brain, options would be instantiated as neural activity patterns. Stochasticity predominantly results from simple thermodynamic 'noise' and in fact is a hallmark feature of most neural activity (see appendix 1 for why it is appropriate to consider thermodynamic variations as random).¹³⁴ It is important to highlight that lower level randomness is ever present, and thus option generation does not need to be 'timed' to the moment of a decision (appendix 2).

5.2 Stage 2: Control or 'Will'

In stage 2, a 'will-function' λ_1 selects the option that best corresponds to an organism's goals, leading to an 'action' (or thought) α .

$$\alpha = \lambda_1(P_0)(1)$$

In a brain, the selection of options corresponds to the selection of certain activity patterns over others by dedicated brain areas which encode information about the organism's goals (e.g. dopaminergic systems). To reflect 'will,' λ_1 must reflect *existing* values and goals of an organism. Importantly, the values and goals defining λ_1 would

131 Note that reductionistic attack like that of Sapolsky would imply not only that free will, but also agency more generally, is an illusion (Mitchell, *Free Agents*, ix). See also, Heisenberg, "Is Free Will an Illusion?"; Brems, "Towards a Scientific Concept of Free Will."

132 Raymond Noble, and Denis Noble, "Harnessing Stochasticity: How Do Organisms Make Choices?" *Chaos* 28, no. 10 (2018): 1–7, <https://doi.org/10.1063/1.5039668>.

133 Mitchell, *Free Agents*, 68.

134 Paul W. Glimcher, "Indeterminacy in Brain and Behavior," *Annual Review of Psychology* 56, no. 1 (2005): 25–56, <https://doi.org/10.1146/annurev.psych.55.090902.141429>.

also be physically instantiated (e.g. in genes or brain circuits), but “only have causal power by virtue of what they mean,”¹³⁵ i.e. reflect ‘informational causation’ or ‘downward causation.’¹³⁶ This is why any metaphysical framework which obfuscates the causal power of biologically stored information, such as reductionism, precludes understanding the concept of agency.

It is worth pausing, to consider the remarkable result of this conjunction of deterministic and indeterministic processes: The precise action (or thought) ultimately chosen is *in principle unpredictable*—it can entail genuine novelty and defies determinism. Yet it must simultaneously accord to the agent’s goal or selection criteria (what we have called λ_1)—it is not random. Since this result defies traditional analytical patterns of thought,¹³⁷ it may feel strange at first. But it is important to note that the underlying mechanism is increasingly recognized to pervade biology at every level (see appendix 2).¹³⁸

5.3 Enter Schopenhauer

While the determinism problem is eliminated at stage 1, it can be partially resurected at stage 2. This was perhaps most famously expressed by Schopenhauer, who claimed that while we may be free to do what we will, we are not free to will what we will. Hashim and Srikanth suggest this problem can be solved by introducing an additional “deeper” stage 3, which chooses λ_1 in a similar manner as the original two-stage model:

$$\lambda_1 = \lambda_2(P_1)(2)$$

where λ_2 reflects a meta-preference, which selects λ_1 out of a set of possible ‘will functions’ P_1 .¹³⁹ Harry Frankfurt has famously called this “second order volition,” and argued that it is what makes someone a “person.”¹⁴⁰ Importantly, the generation of the will-functions must again entail some small random element, in order for it to qualify as free. In practice, we might again imagine a set of will functions generated as stochastic neural spiking patterns. These would likely be slight random variations of previously present will-functions. Substituting, the action selection can be written as

$$\alpha = [\lambda_2(P_1)](P_2) \equiv \lambda_2(P_1, P_0)(3)$$

135 Mitchell, *Free Agents*, 196.

136 Ellis, and Gabriel, “Physical, Logical, and Mental Top-Down Effects.”

137 For example: Strawson, “Free Will.” (See section 4)

138 Noble, and Noble, “Harnessing Stochasticity.”

139 Hashim, and Srikanth, “The Concept of Free Will.”

140 Note that Frankfurt is operating in a compatibilist regime and thus does not require randomness at any level. See Frankfurt, “Freedom of the Will.”

Again, the deeper level (λ_2) is seen as causally preceding more shallow levels. Up to this point, the model remains fairly intuitive, and indeed reminiscent of folk psychology: We might have a 'meta-preference' (stage 3, to be a healthy person) which causes a preference (stage 2, to like salad).

5.4 The Problem of Infinite Regress

As becomes immediately clear though, the problem of determinism can again be resurrected at stage 3. Where did our meta-preference come from? Hashim and Srikanth¹⁴¹ go the logical next step. They argue that a truly free-willed agent must possess an "infinite-stage" depth-recursion:

$$\alpha = \lim_{n \rightarrow \infty} \lambda_n(P_{n-1}, \dots, P_0)(4)$$

However, this is where Hashim and Srikanth's model of free will becomes implausible. Indeed, the problem of infinite regress has been forcefully pointed out as the key problem with Frankfurt's hierarchical theory¹⁴² of free will.¹⁴³ In a real organism, representing a 'will function' requires physical space, for instance because any neural network can store only a finite number of representations. It also conflicts with introspective intuition. We might have a conscious meta-preference over, say, an immediate consumption preference (e.g., I want the burger and I feel bad about it).¹⁴⁴ But we clearly cannot take this to infinite depth.

At this point it is worth noting that, because we do not assume determinism,¹⁴⁵ even a finite number of steps would yield something new, reminiscent of genuine free agency. The actions resulting from such a process would be neither determined nor random. They would be oriented towards the higher-level goal, but they would be impossible to predict even in principle.

141 Hashim, and Srikanth, "The Concept of Free Will."

142 Frankfurt, "Freedom of the Will."

143 For example: Christian F. Rostbøll, "Freedom as Satisfaction? A Critique of Frankfurt's Hierarchical Theory of Freedom," *SATS* 5, no. 1 (2004): 131–46, <https://doi.org/10.1515/SATS.2004.131>. While Frankfurt recognized the issue of infinite regress and dedicated two subsequent essays attempting to address it, Rostbøll convincingly argues that he ultimately fails. Specifically, Rostbøll highlights that the criteria Frankfurt uses to label someone as lacking free will (of being a "wanton") are difficult to distinguish from those used to break the infinite regress.

144 Some may even have conscious meta-meta-preferences (e.g., I'm happy that I feel bad about wanting the burger, say because it will help me want it less in the future.).

145 See for example Frankfurt, "Freedom of the Will."

5.5 Free Will as a Temporal Recursion

The next key step is to recognize that the recursion we outlined above can occur not only across hierarchical levels of organization, but also *in time*.¹⁴⁶ We can, for instance, define $\lambda_{2 \text{ to } n}$ in eq. 4 as reflecting choices at *previous points in time*. Say one had previously chosen to try to be a healthier person, and has since cultivated a preference for salad. In fact, every choice we make can be thought to influence habit and preference formation,¹⁴⁷ suggesting that every single choice is simultaneously a choice about our future preferences (also see appendix 2).

This understanding suggests a present choice is indeed influenced by myriad previous choices and external influence factors, reaching into the distant past.¹⁴⁸ But, it is important to reiterate, that for each choice, higher-level (and/or earlier) goals select from a not-fully determined option set, meaning the entire sequence is far from deterministic. More importantly, if goals and option-sets are at least partially subject to *conscious deliberation*, then it is reasonable to state that our conscious selves can genuinely guide which kind of person we come to be and which kinds of actions we come to take. In other words, the present will λ_1 is in part the product of conscious deliberations and choices over the entire course of one's life.¹⁴⁹ In the words of Carstensen "agents with the capacity for recursive decision-making, i.e. decisions that target and alter the agent's own motivational structure, have an important kind of control over their own internal randomness."¹⁵⁰

The final step is to turn the above reasoning on its head, and consider the implications of conscious choice *for the future*. Every choice we make will impact both our will and choices into the indefinite future—it will shape who we come to be. If the above image of a retrospective temporal recursion suggested limits to responsibility, then surely recognizing the *unquantifiably large* impact of a choice on our future implies significant responsibility. In fact, taking the retrospective understanding seriously, as Sapolsky clearly does, by necessity implies that our choices will impact not only our future selves, but also who our children come to be, and how society more generally develops.¹⁵¹

146 Kane, "The Complex Tapestry of Free Will."

147 Matija Franklin et al., "Recognising the Importance of Preference Change: A Call for a Coordinated Multidisciplinary Research Effort in the Age of AI," *The AAAI-22 Workshop on AI For Behavior Change*, held at the Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22), 2022.

148 Consistent with Sapolsky, *Determined: Life Without Free Will*.

149 It will also, in part, be the product of choices made by our parents and grandparents, and indeed society and political decision makers more generally (say the presence of a humanistic education system).

150 Carstensen, "Freedom and Chance."

151 Sapolsky, *Determined: Life Without Free Will*.

We would argue that this understanding of free will as a temporal recursion of a two-stage model is not only what the best scientific evidence suggests, it is also the appropriate scientific foundation for digital humanism. It means we have the freedom and the responsibility to shape our personal and our collective future. But it simultaneously acknowledges that any decision is influenced by a myriad of factors in our biology, our history, and increasingly our digital choice environments. Free will, in this model, would be what unfolds when we do the hard work of learning about and understanding these influences and choosing whether to follow, to nurture or to suppress their role in our individual and collective decisions.

6 Summary and Conclusion

We have argued that increasingly sophisticated forms of digital nudging reveal that a robust conception of digital humanism requires updating our understanding of free will. The reason is that digital nudging has increasing power and sophistication but defies traditional normative categories because it often arises without intention (via proxy-optimization). This has furnished two seemingly irreconcilable visions for digital humanism, in particular concerning the normative question whether we should or should not emphasize free will in order to further humanist goals. One side argues we should *deemphasize*, or perhaps even reject, free will in order to further humanist goals. The other side insists that free will and the idea that we can be the authors of our own decisions are the very foundation of humanist notions such as personhood and dignity. We suggested that the normative arguments of both sides have merit, i.e., that both over- and underemphasizing free will can undermine humanist goals: Asserting free will where it is illusory leads to unjustified praise and blame; asserting that there is no free will undermines human dignity itself. But of course the desirability of free will (the normative question) does not speak to whether or not it actually exists (the positive question).

We thus moved to the positive question, i.e., whether or in which way free will actually exists. We first highlighted two entrenched obstacles to understanding within the free will debate, namely a false dichotomy between determinism and indeterminism and a strong metaphysical commitment to reductionism. Addressing these obstacles laid the groundwork for our presentation of an emerging modern, scientifically grounded model of genuine libertarian free will. Briefly, it construes free will as a continuous interaction between deterministic and indeterministic 'stages' within a temporal recursion. Importantly, this understanding implies that we are making not just a normative but also a positive error (i.e. we are misconstruing reality) when we over- or underemphasize free will.

6.1 General Implications

What are the general positive implications of the *temporally recursive two stage model* of free will?

- a) Free will *exists*. The human capacity for genuine self-authorship is consistent with, and supported by, a modern scientific understanding of the world. We can reject the tacit (or overt) nihilism of those who argue that scientific evidence implies that free will is merely an illusion.¹⁵²
- b) Free will is a *graded* phenomenon.¹⁵³ To assume any choice can be *fully* free from external (prior) influences would indeed be an illusion. Understanding, and scientifically studying, the full role and impact of influences should not be misconstrued to be in itself undermining free will, and indeed *is necessary* to develop genuine freedom.
- c) Free will is a *temporal* phenomenon—it unfolds and cumulates over time. This, e.g., directly motivates why we regard a child as less responsible than an adult.
- d) Free will is in part a *social* phenomenon. It is not fully reducible to an individual but instead requires consciously shaping the influence factors and nudges we are exposed to. Not recognizing this, and not participating in shaping, amounts to relinquishing part of one's freedom.¹⁵⁴

The reader might notice the resonance of these conclusions with traditional humanist views. In particular, it seems to us that humanists have not only generally emphasized the *existence* of free will (a), but also recognized it to be *graded* (b), to unfold *over time* (c), and occur *within a social context* (d). In fact, all this is implicit in what is (not coincidentally) known as a 'humanistic education.'¹⁵⁵ Humanistic education is designed precisely to nurture our deliberative capacities and ability for genuine self-authorship, implying not only existence but also gradation. Moreover, already the earliest Renaissance humanists recognized that such self-authorship unfolds *over time*—Pico della Mirandola speaks of “seeds” which we must “cultivate” in order to

152 Such as Sapolsky, *Determined: Life Without Free Will*.

153 Carstensen, “Freedom and Chance.”

154 For an individual to shape its future, it needs to in part shape the choice environment it lives in, e.g. via democratic institutions. In other words, a part of free will requires a shift from the individual-frame to a system-frame. See Nick Chater, and George Loewenstein, “The I-Frame and the s-Frame: How Focusing on Individual-Level Solutions Has Led Behavioral Public Policy Astray,” *Behavioral and Brain Sciences* 46 (2023): 1–60, <https://doi.org/10.1017/S0140525X22002023>.

155 Martha C. Nussbaum, *Not for Profit: Why Democracy Needs the Humanities* (Princeton, NJ: Princeton University Press, 2012), <https://doi.org/10.1215/0961754X-1305490>.

become who we want to be.¹⁵⁶ And of course, education and character-formation are overtly *social*. If education matters in any way for our character formation, then self-authorship is necessarily *in part* a social project. It then necessarily entails socially deliberating and designing, e.g., our system of education. Similarly, if our prefrontal cortex strength (and ability for self-control) is in part determined by childhood deprivation, then promoting free choice requires social policy to prevent deprivation.

6.2 Implications for Digital Humanism

What are the concrete implications for digital humanism? First of all, the two normative stances (deemphasizing or emphasizing free will) must both be considered. We treat each in turn.

Deemphasizing free will: We should accept that our digital environments will inevitably influence behavior and reject the notion that merely understanding how this happens in and of itself undermines free will (as Zuboff seems to do¹⁵⁷). Instead, understanding digital nudging is a *fundamental prerequisite to taking conscious control* over how we are nudged. In this sense, the reductionist research program of “social physics,”¹⁵⁸ whereby we learn exactly when and how digital infrastructure shapes our thoughts and choices without us even noticing, must be an essential part of digital humanism. However, the emphasis that such knowledge is key to establishing genuine freedom in no way denies that it simultaneously creates unprecedented dangers of being misused.¹⁵⁹ Indeed the key conclusion from such knowledge may well be to monitor and restrict its application (including restrictions on “digital surveillance”).

Emphasizing free will: We should emphasize the genuine freedom and right of every individual in choosing how we want to be nudged. Wherever possible, we should promote autonomy at the individual level, say by providing options that allow us to explicitly control how we are digitally nudged.¹⁶⁰ However, it is again essential to appreciate that nudging is in part irreducibly social (e.g., in recommender systems). In such cases, the only opportunity to consciously control how we are nudged is as a community. This means we must promote social (e.g. democratic) deliberation about how exactly we want to be nudged. As a result, we may choose to promote

156 Pico della Mirandola, *Pico Della Mirandola: Oration on the Dignity of Man*, A New Translation and Commentary by Francesco Borghesi, Michael Papio and Massimo Riva (New York: Cambridge University Press, 2012), 547.

157 Zuboff, *The Age of Surveillance Capitalism*.

158 Pentland, *Social Physics*.

159 Treen et al., “Online Misinformation about Climate Change”; Zuboff, *The Age of Surveillance Capitalism*.

160 Robertson et al., “Changing the Incentive Structure of Social Media.”

some types of nudging (say against conspiracies or for sustainability¹⁶¹) and restrict other types (say clearly deceptive nudges¹⁶²).

To summarize, asserting humanism and free will in a digital era will increasingly come to mean that we must ‘*judge the nudges*.’ It is no longer sufficient to simply assume our choices are free unless someone (demonstrably) manipulates us. Indeed, overemphasizing free will in this way systematically obfuscates the myriad factors that subconsciously shape our choices, beliefs and identities, and thus paradoxically undermines genuine free will. Instead, we can assert genuine free will by understanding and emphasizing how digital environments shape our choices and developing mechanisms which allow to continually subject this shaping to individual and democratic control.

Appendix 1: Debunking Determinism

Natural science, it might be argued, is the study of how exactly causes *determine* their effects.¹⁶³ The remarkable success of the sciences in recent centuries is a testament to both the power and accuracy of deterministic thinking. As science progresses, it uncovers more and more previously hidden cause-effect relationships, i.e. ways in which things were determined by causes in the past. So, we can confidently assert that the world is *partially deterministic* or contains many macroscopically deterministic processes.

However, this does not in itself support a widespread metaphysical assumption: That the world is a *fully deterministic* “clockwork universe”, a single great chain of microscopic causes and effects, where every action anyone has ever taken was predetermined at the beginning of the universe. Of course, this metaphysical assumption is much older than modern science and can be questioned on purely philosophical grounds.¹⁶⁴

But it is important to note that “full determinism” has in fact also been at odds with scientific understanding since the beginning of the 20th century, when it was fundamentally undermined by the finding of quantum randomness.¹⁶⁵ While some still postulate that quantum randomness could be compatible with some hidden

161 C. Tyler DesRoches et al., “When Is Green Nudging Ethically Permissible?” *Current Opinion in Environmental Sustainability* 60 (2023), <https://doi.org/10.1016/j.cosust.2022.101236>.

162 Jamie Luguri, and Lior Strahilevitz, “Shining a Light on Dark Patterns,” *Journal of Legal Analysis* 13, no. 1 (2021): 43–109, <https://doi.org/10.2139/ssrn.3431205>; Simms, and Murray, *Badvertising*.

163 Glimcher, “Indeterminacy in Brain and Behavior.”

164 Markus Gabriel, *Warum Es Die Welt Nicht Gibt* (Berlin: Ullstein, 2013); Ellis, and Gabriel, “Physical, Logical, and Mental Top-Down Effects.”

165 Jedlicka, “Revisiting the Quantum Brain Hypothesis.”

form of “superdeterminism,”¹⁶⁶ our best scientific evidence suggests that the world at the quantum level is fundamentally indeterministic.

Despite this apparent indeterminism at the most fundamental level of physical reality, many macroscopic (or ‘classical’) physical processes are “effectively determinate”¹⁶⁷ because impacts of quantum fluctuation often ‘average out.’¹⁶⁸ Importantly, this kind of ‘statistical’ determinism is an emergent phenomenon. It in no way implies that say the trajectories of individual molecules in a fluid or ion channels within our brains are 100% predetermined, only that randomness should *mostly* cancel out when a system is examined at a larger scale or described statistically.¹⁶⁹

What is more, many of the systems in question are *chaotic*, i.e. they by no means tend to average out minuscule changes of micro-level components, but can instead translate such changes to large scale randomness (a phenomenon popularly known as the butterfly effect). Even if such a chaotic system were only ‘infused’ with true quantum randomness once in 100 years, say because some radioactive decay process slightly altered a trajectory of a water molecule, many of the subsequent chaotic molecular trajectories should arguably be considered random for all intents and purposes.

So we can also confidently assert that the world is *partially indeterministic*, where even partially deterministic higher-level processes may actually occur on top of a layer of lower-level indeterminism and chaos. Our best scientific understanding thus suggests the world is a complex mesh of deterministic and indeterministic processes, which constantly interact.

Appendix 2: The Timing Objection

One objection to two-stage models concerns the problem of timing.¹⁷⁰ Dennett argues, that the time point of a random quantum event would have to coincide pre-

166 Giacomo Andreoletti, and Louis Vervoort, “Superdeterminism: A Reappraisal,” *Synthese* 200, no. 5 (2022): 361, <https://doi.org/10.1007/s11229-022-03832-6>.

167 Glimcher, “Indeterminacy in Brain and Behavior.”

168 Christof Koch, “Free Will, Physics, Biology, and the Brain,” in *Downward Causation and the Neurobiology of Free Will*, ed. Nancey Murphy, George F. R. Ellis, and Timothy O'Connor (Berlin, Heidelberg: Springer, 2009), https://doi.org/10.1007/978-3-642-03205-9_2.

169 As an aside, it is worth noting that biology is filled with cases where random quantum processes clearly affect macroscopic outcomes, such as when radioactive decay causes a genetic mutation, or when individual photons are detected by a human eye. See Jedlicka, “Revisiting the Quantum Brain Hypothesis.” Within the brain, a crucial stochastic process that could be partially affected by quantum indeterminacy is the opening and closing of ion-channels, which forms the bases of all neural activity. See Tse, “Two Types of Libertarian Free Will.”

170 For example, Alfred R. Mele, *Free Will and Luck* (New York: Oxford University Press, 2008).

cisely with the time point of a decision, something that he believes is difficult to accomplish.¹⁷¹ The same intuition, likely underlies Kane's caution, when identifying free will with the occurrence or *very rare* "self-forming acts", which need occur only few times in life.¹⁷² In our view, the notion that there is a problem of timing belies an inaccurate understanding of complex biological systems, which essentially reflect an 'emergent' form of deterministic higher-level structure which is built upon a continuous stream of lower-level randomness.

For instance, Doyle highlights that the organism has *continuous access* to stochastic events (which need not all be based in quantum dynamics, see appendix 1), and can *control* when to harness or ignore them.¹⁷³ The organism can for instance generate action options, deliberate if they are sufficient, and if not, decide to generate additional options. Notably, such an account is fully consistent with biological evidence. Noble and Noble¹⁷⁴, as well as Mitchell¹⁷⁵ review the biological evidence for such controlled usage of stochasticity at numerous levels in biology. It is perhaps best explored in the immune system, where exposure to novel antigens triggers controlled hypermutation *specifically* at antigen binding sites.¹⁷⁶ More generally, an increasing number of evolutionary biologists is emphasizing that the processes of genetic mutation more generally are far from random, but are tightly controlled and directed. For instance, when bacteria are subjected to stress, they can transiently increase mutation rates, thus for instances generating resistance to antibiotics.¹⁷⁷ Mitchell reviews how empirically observed neural activity, across species, fits perfectly into this picture.¹⁷⁸ In each case, it is misguided to think of stochasticity being generated at a certain point in time. Instead, biological systems exist on top of constant stochastic flows of matter and energy, which can be amplified or ignored according to the organism's goals.

171 Daniel C. Dennett, "On Giving Libertarians What They Say They Want," in *Brainstorms* (Cambridge, MA: MIT Press, 1978), <https://direct.mit.edu/books/edited-volume/3637/chapter/121522/On-Giving-Libertarians-What-They-Say-They-Want>.

172 Kane, "The Complex Tapestry of Free Will."

173 Doyle, "The Two-Stage Solution."

174 Noble, and Noble, "Harnessing Stochasticity."

175 Mitchell, *Free Agents*.

176 Noble, and Noble, "Harnessing Stochasticity."

177 Laland et al., "The Extended Evolutionary Synthesis." Incidentally, this literature emphasizes that randomness is harnessed towards teleonomic (i.e. apparently goal-oriented) behavior. See Bartlett, "Evolutionary Teleonomy."

178 Mitchell, *Free Agents*.

Bibliography

- Andreoletti, Giacomo, and Louis Vervoort. "Superdeterminism: A Reappraisal." *Synthese* 200, no. 361 (2022). <https://doi.org/10.1007/s11229-022-03832-6>.
- Atmanspacher, Harald. "On Macrostates in Complex Multi-Scale Systems." *Entropy* 18, no. 12 (2016): 426. <https://doi.org/10.3390/e18120426>.
- Bartlett, Jonathan. "Evolutionary Teleonomy as a Unifying Principle for the Extended Evolutionary Synthesis." *BIO-Complexity*, no. 2 (2017): 1–7. <https://doi.org/10.5048/BIO-C.2017.2>.
- Blüher, Matthias. "Obesity: Global Epidemiology and Pathogenesis." *Nature Reviews Endocrinology* 15, no. 5 (2019): 288–98. <https://doi.org/10.1038/s41574-019-0176-8>.
- Bond, Robert M., Christopher J. Fariss, Jason J. Jones, et al. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489 (2012): 295–98. <https://doi.org/10.1038/nature11421>.
- Braganza, Oliver. "Proxyeconomics, a Theory and Model of Proxy-Based Competition and Cultural Evolution." *Royal Society Open Science* 9, no. 2 (2022): 1–41. <https://doi.org/10.1098/RSOS.211030>.
- . "Nudging Society to Consume—the Aggregate Consequences of Consumption Nudges." *Review of Evolutionary Political Economy*, 6 (2025): 753–783. <https://doi.org/10.1007/s43253-024-00144-y>.
- Braganza, Oliver, Yohan J. John, Leigh Caldwell, and Dakota E. McCoy. "Teleonomy, Legibility, and Diversity: Do We Need More 'Proxynomics'?" *Behavioral and Brain Sciences* 47 (2024): e88. <https://doi.org/10.1017/S0140525X24000050>.
- Braganza, Oliver, and Jakob Kapeller. "Reappraising Consumption Nudging – on Liberty in the Age of Climate Catastrophe." *Humanities and Social Sciences Communications* 12, no. 251 (2025).
- Brembs, Björn. "Towards a Scientific Concept of Free Will as a Biological Trait: Spontaneous Actions and Decision-Making in Invertebrates." *Proceedings of the Royal Society B: Biological Sciences* 278, no. 1707 (2010): 930–39. <https://doi.org/10.1098/rspb.2010.2325>.
- Calo, Ryan. "Digital Market Manipulation." *George Washington Law Review* 82, no. 4 (2013): 995–1051.
- Carstensen, Mark Wulff. "Freedom and Chance: Towards a Graded Approach to Free Will." PhD. diss., Ludwig-Maximilians-Universität München, 2021. https://edoc.ub.uni-muenchen.de/28639/1/Carstensen_Mark_Wulff.pdf.
- Caruso, Gregg D. *Public Health and Safety: The Social Determinants of Health and Criminal Behavior*. UK: ResearchersLinks Books, 2017.
- . "Rejecting Retributivism: Free Will, Punishment, and Criminal Justice, A Précis." *Journal of Practical Ethics* 10, no. 2 (2023). <https://doi.org/10.3998/jpe.3544>.
- Chater, Nick, and George Loewenstein. "The I-Frame and the s-Frame: How Focus on Individual-Level Solutions Has Led Behavioral Public Policy Astray." *Be-*

- havioral and Brain Sciences* 46 (2023): 1–60. <https://doi.org/10.1017/S0140525X22002023>.
- Cohen, Deborah A. “Neurophysiological Pathways to Obesity: Below Awareness and beyond Individual Control.” *Diabetes* 57, no. 7 (2008): 1768–73. <https://doi.org/10.2337/db08-0163>.
- Conway, Erik M, and Naomi Oreskes. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. New York: Bloomsbury Press, 2011.
- Dawkins, Richard. *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design*. Reissue Edition. New York: W. W. Norton & Company, 2015.
- Dennett, Daniel C. “On Giving Libertarians What They Say They Want.” In *Brainstorms*. Cambridge, MA: MIT Press, 1978. <https://direct.mit.edu/books/edited-volume/3637/chapter/121522/On-Giving-Libertarians-What-They-Say-They-Want>.
- . *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA, London: MIT Press, 1984. <https://doi.org/10.7551/mitpress/10470.001.0001>.
- Dennett, Daniel C., and Gregg D. Caruso. “On Free Will: Daniel Dennett and Gregg Caruso Go Head to Head | Aeon Essays.” *Aeon* (2018). <https://aeon.co/essays/on-free-will-daniel-dennett-and-gregg-caruso-go-head-to-head>.
- DesRoches, C Tyler, Daniel Fischer, Julia Silver, et al. “When Is Green Nudging Ethically Permissible?” *Current Opinion in Environmental Sustainability* 60 (2023): 101236. <https://doi.org/10.1016/j.cosust.2022.101236>.
- Doyle, Robert O. “The Two-Stage Solution to the Problem of Free Will.” In *Is Science Compatible with Free Will?* Edited by Antoine Suarez and Peter Adams. New York: Springer, 2013.
- Ellis, George F. R., and Markus Gabriel. “Physical, Logical, and Mental Top-Down Effects.” In *Top-down Causation and Emergence*, edited by Jan Voosholz and Markus Gabriel. Cham: Springer International Publishing, 2021.
- Ellis, George F. R., and Jonathan Kopel. “The Dynamical Emergence of Biology From Physics: Branching Causation via Biomolecules.” *Frontiers in Physiology* 9 (2019). <https://doi.org/10.3389/fphys.2018.01966>.
- Enders, Adam M., Joseph E. Uscinski, Michelle I. Seelig, et al. “The Relationship Between Social Media Use and Beliefs in Conspiracy Theories and Misinformation.” *Political Behavior* 45, no. 2 (2023): 781–804. <https://doi.org/10.1007/s11109-021-09734-6>.
- Esposito, Fabrizio, and Thaís Maciel Cathoud Ferreira. “Addictive Design as an Unfair Commercial Practice: The Case of Hyper-Engaging Dark Patterns.” *European Journal of Risk Regulation* 15, no. 4 (2024): 999–1016. <https://doi.org/10.1017/err.2024.8>.

- Faddoul, Marc, Guillaume Chaslot, and Hany Farid. "A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos." *arXiv*, March 6, 2020. <http://arxiv.org/abs/2003.03318>.
- Flake, Gary William. *The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems, and Adaptation*. Cambridge, MA: MIT Press, 1998.
- Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68, no. 1 (1971): 5–20. <https://doi.org/10.2307/2024717>.
- Franklin, Matija, Hal Ashton, Rebecca Gorman, and Stuart Armstrong. "Recognising the Importance of Preference Change: A Call for a Coordinated Multidisciplinary Research Effort in the Age of AI." *The AAAI-22 Workshop on AI For Behavior Change*, held at the Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22), 2022.
- Frick, Vivian, Maike Gossen, and Sara Elisa Kettner. "Does online advertising stimulate overconsumption?" *Ökologisches Wirtschaften – Fachzeitschrift* 37, no. 1 (2022). <https://doi.org/10.14512/OEW370146>.
- Gabriel, Markus. *Warum Es Die Welt Nicht Gibt*. Berlin: Ullstein, 2013.
- . *Fields of Sense: A New Realist Ontology*. Edinburgh: Edinburgh University Press, 2015.
- . *Ich ist nicht Gehirn: Philosophie des Geistes für das 21. Jahrhundert*. Berlin: Ullstein Ebooks, 2015.
- Glimcher, Paul W. "Indeterminacy in Brain and Behavior." *Annual Review of Psychology* 56, no. 1 (2005): 25–56. <https://doi.org/10.1146/annurev.psych.55.090902.141429>.
- Haig, David. *From Darwin to Derrida: Selfish Genes, Social Selves, and the Meanings of Life*. Cambridge, MA: MIT Press, 2020.
- Hall, Kevin D. "Did the Food Environment Cause the Obesity Epidemic?" *Obesity* 26, no. 1 (2018): 11–13. <https://doi.org/10.1002/oby.22073>.
- Harris, Tristan. "How a Handful of Tech Companies Control Billions of Minds Every Day." 2017. https://Opgaver.Roedkilde-Gym.Dk/Engelsk/Engelsk%20B%20stx.%2031.%20maj%202021/Files/193760_non_fiction_4.Pdf.
- Hashim, Hanaan, and R. Srikanth. "The Concept of Free Will as an Infinite Metatheoretic Recursion." *arXiv*, Preprint, June 21, 2015. <http://arxiv.org/abs/1503.06485>.
- Heisenberg, Martin. "Is Free Will an Illusion?" *Nature* 459 (2009): 164–65. <https://doi.org/10.1038/459164a>.
- Henrich, Joseph Patrick. *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton, NJ: Princeton University Press, 2016.
- Hesp, Casper, Maxwell Ramstead, Axel Constant, Paul Badcock, Michael Kirchhoff, and Karl Friston. "A Multi-Scale View of the Emergent Complexity of Life: A Free-Energy Proposal." *Springer Proceedings in Complexity* (2019): 195–227. https://doi.org/10.1007/978-3-030-00075-2_7.

- Hoel, Erik P., Larissa Albantakis, William Marshall, and Giulio Tononi. "Can the Macro Beat the Micro? Integrated Information across Spatiotemporal Scales." *Neuroscience of Consciousness*, no. 1 (2016). <https://doi.org/10.1093/nc/niw012>.
- Ichihashi, Shota, and Byung-Cheol Kim. "Addictive Platforms." *Management Science* 69, no. 2 (2023): 1127–45. <https://doi.org/10.1287/mnsc.2022.4392>.
- Ivanković, Viktor, and Bart Engelen. "Market Nudges and Autonomy." *Economics & Philosophy* 40, no. 1 (2022): 138–65. <https://doi.org/10.1017/S0266267122000347>.
- Jedlicka, Peter. "Revisiting the Quantum Brain Hypothesis: Toward Quantum (Neuro)Biology?" *Frontiers in Molecular Neuroscience* 10 (2017). <https://doi.org/10.3389/fnmol.2017.00366>.
- Jesse, Mathias, and Dietmar Jannach. "Digital Nudging with Recommender Systems: Survey and Future Directions." *Computers in Human Behavior Reports* 3 (2021). <https://doi.org/10.1016/j.chbr.2020.100052>.
- John, Yohan J., Leigh Caldwell, Dakota E. McCoy, and Oliver Braganza. "Dead Rats, Dopamine, Performance Metrics, and Peacock Tails: Proxy Failure Is an Inherent Risk in Goal-Oriented Systems." *Behavioral and Brain Sciences* 47 (2024): 1–56. <https://doi.org/10.1017/S0140525X23002753>.
- Kane, Robert. "The Complex Tapestry of Free Will: Striving Will, Indeterminism and Volitional Streams." *Synthese* 196, no. 1 (2019): 145–60. <https://doi.org/10.1007/s11229-016-1046-8>.
- Kaplan, Joel. "More Speech and Fewer Mistakes." *Meta Newsroom*, January 7, 2025. <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>.
- Kirkpatrick, Jerry. *In Defense of Advertising: Arguments From Reason, Ethical Egoism, and Laissez-Faire Capitalism*. Claremont, CA: Kirkpatrick Books, 1994. <https://philpapers.org/rec/KIRIDO>.
- Koch, Christof. "Free Will, Physics, Biology, and the Brain." In *Downward Causation and the Neurobiology of Free Will*, edited by Nancey Murphy, George F. R. Ellis, and Timothy O'Connor. Berlin, Heidelberg: Springer, 2009. https://doi.org/10.1007/978-3-642-03205-9_2.
- Laland, Kevin N., Tobias Uller, Marcus W. Feldman, et al. "The Extended Evolutionary Synthesis: Its Structure, Assumptions and Predictions." *Proceedings of the Royal Society B: Biological Sciences* 282, no. 1813 (2015): 1–14. <https://doi.org/10.1098/RSPB.2015.1019>.
- Lee, Patrick, and Robert P. George. "The Nature and Basis of Human Dignity." *Ratio Juris* 21, no. 2 (2008): 173–93. <https://doi.org/10.1111/j.1467-9337.2008.00386.x>.
- Luguri, Jamie, and Lior Strahilevitz. "Shining a Light on Dark Patterns." *Journal of Legal Analysis* 13, no. 1 (2021): 43–109. <https://doi.org/10.2139/ssrn.3431205>.
- Mann, Michael E. *The New Climate War: The Fight to Take Back Our Planet*. London: Scribe Publications, 2022.
- Mele, Alfred R. *Free Will and Luck*. New York: Oxford University Press, 2008.

- Mesoudi, Alex, and Alex Thornton. "What Is Cumulative Cultural Evolution?" *Proceedings of the Royal Society B: Biological Sciences* 285, no. 1880 (2018). <https://doi.org/10.1098/rspb.2018.0712>.
- Mills, Stuart. "Nudge/Sludge Symmetry: On the Relationship between Nudge and Sludge and the Resulting Ontological, Normative and Transparency Implications." *Behavioural Public Policy* 7, no. 9 (2020): 309–32. <https://doi.org/10.1017/bpp.2020.61>.
- Mirandola, Pico della. *Pico Della Mirandola: Oration on the Dignity of Man*. A New Translation and Commentary by Francesco Borghesi, Michael Papio and Massimo Riva. New York: Cambridge University Press, 2012.
- Mitchell, Kevin J. *Free Agents*. Princeton: Princeton University Press, 2023. <https://press.princeton.edu/books/hardcover/9780691226231/free-agents>.
- Montag, Christian, Bernd Lachmann, Marc Herrlich, and Katharina Zweig. "Addictive Features of Social Media/Messenger Platforms and Freemium Games against the Background of Psychological and Economic Theories." *International Journal of Environmental Research and Public Health* 16, no. 14 (2019). <https://doi.org/10.3390/ijerph16142612>.
- Mora-Salgueiro, Javier, Aitana García-Estela, Bridget Hogg, et al. "The Prevalence and Clinical and Sociodemographic Factors of Problem Online Gambling: A Systematic Review." *Journal of Gambling Studies* 37, no. 3 (2021): 899–926. <https://doi.org/10.1007/s10899-021-09999-w>.
- Nida-Rümelin, Julian, and Dorothea Winter. "Humanism and Enlightenment." In *Introduction to Digital Humanism*, edited by Hannes Werthner et al. Cham: Springer 2024. <https://doi.org/10.1007/978-3-031-45304-5>.
- Noble, Raymond, and Denis Noble. "Harnessing Stochasticity: How Do Organisms Make Choices?" *Chaos* 28, no. 10 (2018): 1–7. <https://doi.org/10.1063/1.5039668>.
- . "Can Reasons and Values Influence Action: How Might Intentional Agency Work Physiologically?" *Journal for General Philosophy of Science* (2020): 1–19. <https://doi.org/10.1007/s10838-020-09525-3>.
- Nussbaum, Martha C. *Not For Profit: Why Democracy Needs the Humanities*. Princeton, NJ: Princeton University Press, 2012. <https://doi.org/10.1215/0961754X-1305490>.
- Papineau, David. "Naturalism." *The Stanford Encyclopedia of Philosophy*, Fall 2023, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2023/entries/naturalism/>.
- Pearl, Judea, and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. London: Penguin Books, 2019.
- Penrose, Roger, Abner Shimony, Nancy Cartwright, and Stephen Hawking. *The Large, the Small and the Human Mind*. Cambridge: Cambridge University Press, 2000.

- Pentland, Alex. *Social Physics: How Social Networks Can Make Us Smarter*. New York: Penguin Books, 2015. <https://www.amazon.de/Social-Physics-Networks-Make-Smarter/dp/0143126334>.
- Pereboom, Derk, and Michael McKenna. "Manipulation Arguments Against Compatibilism." In *The Oxford Handbook of Moral Responsibility*, 1st Edition, edited by Dana Kay Nelkin and Derk Pereboom. New York: Oxford University Press, 2022. <https://doi.org/10.1093/oxfordhb/9780190679309.013.38>.
- Rizzo, Mario J., and Glen Whitman. *Escaping Paternalism: Rationality, Behavioral Economics, and Public Policy*. Cambridge, New York: Cambridge University Press, 2019. <https://www.cambridge.org/core/books/escaping-paternalism/6CC068636EB534E04792F2AC41F2D4D9>.
- Robertson, Claire E., Kareena del Rosario, Steve Rathje, and Jay J. Van Bavel. "Changing the Incentive Structure of Social Media May Reduce Online Proxy Failure and Proliferation of Negativity." *Behavioral and Brain Sciences* 47 (2024). <https://doi.org/10.1017/S0140525X23002935>.
- Rostbøll, Christian F. "Freedom as Satisfaction? A Critique of Frankfurt's Hierarchical Theory of Freedom." *SATS* 5, no. 1 (2004): 131–46. <https://doi.org/10.1515/SATS.2004.131>.
- Sacco, Pier Luigi. "It's the Biology, Stupid! Proxy Failures in Economic Decision Making." *Behavioral and Brain Sciences* 47 (2024). <https://doi.org/10.1017/S0140525X23002947>.
- Sapolsky, Robert M. *Determined: Life Without Free Will*. London: Bodley Head, 2023.
- Sartre, Jean-Paul. *Existentialism Is a Humanism*. New Haven: Yale University Press, 1996.
- Silva Pinho, Ana da, Violeta Céspedes Izquierdo, Björn Lindström, and Wouter van den Bos. "Youths' Sensitivity to Social Media Feedback: A Computational Account." *Science Advances* 10, no. 43 (2024). <https://doi.org/10.1126/sciadv.adp8775>.
- Simms, Andrew, and Leo Murray. *Badvertising: Polluting Our Minds and Fuelling Climate Chaos*. London: Pluto Press, 2023.
- Smaldino, Paul E., and Peter J. Richerson. "Human Cumulative Cultural Evolution as a Form of Distributed Computation." In *Handbook of Human Computation*. New York: Springer, 2013. https://doi.org/10.1007/978-1-4614-8806-4_76.
- Strawson, Galen. "Free Will." In *Routledge Encyclopedia of Philosophy*. London, New York: Routledge, 1998. <https://doi.org/10.4324/9780415249126-VO14-1>.
- Stuckler, David, Martin McKee, Shah Ebrahim, and Sanjay Basu. "Manufacturing Epidemics: The Role of Global Producers in Increased Consumption of Unhealthy Commodities Including Processed Foods, Alcohol, and Tobacco." *PLoS Medicine* 9, no. 6 (2012). <https://doi.org/10.1371/journal.pmed.1001235>.

- Sunstein, Cass R., and Richard H. Thaler. "Libertarian Paternalism Is Not an Oxymoron." *University of Chicago Law Review* 70, no. 4 (2003): 1159–202. <https://doi.org/10.2307/1600573>.
- Tay, Li Quian, and K. H. Ecker. "Continued Influence of Disinformation and the Information Disorder." In *The Remaking of Memory in the Age of the Internet and Social Media*, edited by Qi Wang and Andrew Hoskins. New York: Oxford University Press, 2024.
- Tegmark, Max. "Importance of Quantum Decoherence in Brain Processes." *Physical Review E* 61, no. 4 (2000): 4194–206. <https://doi.org/10.1103/PhysRevE.61.4194>.
- Thaler, Richard H., and Cass R. Sunstein. *Nudge: The Final Edition*. 1st Edition. London: Penguin Books, 2021.
- . *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New York, London: Penguin Books, 2008. <https://psycnet.apa.org/record/2008-03730-000>.
- Treen, Kathie M. d'I., Hywel T. P. Williams, and Saffron J. O'Neill. "Online Misinformation about Climate Change." *WIREs Climate Change* 11, no. 5 (2020): e665. <https://doi.org/10.1002/wcc.665>.
- Trzaskowski, Jan. "Manipulation by Design." *Electronic Markets* 34, no. 1 (2024): 14. <https://doi.org/10.1007/s12525-024-00699-y>.
- Tse, Peter U. "Two Types of Libertarian Free Will Are Realized in the Human Brain." In *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience*, edited by Gregg Caruso, and Owen Flanagan. New York: Oxford University Press, 2018. <https://doi.org/10.1093/oso/9780190460723.003.0010>.
- Van Satrien, Marij. "The Vienna Circle against Quantum Speculations." *HOPOS: The Journal of the International Society for the History of Philosophy of Science* 12, no. 2 (2022): 359–94. <https://doi.org/10.1086/721137>.
- Van Tulleken, Chris. *Ultra-Processed People: The Science Behind Food That Isn't Food*. New York: W. W. Norton & Company, 2023.
- Vienna Manifesto on Digital Humanism*. DIGHUM. 2019. <https://caiml.org/dighum/dighum-manifesto/>.
- Werthner, Hannes, Erich Prem, Carlo Ghezzi, et al., eds. *Introduction to Digital Humanism*. Cham: Springer, 2024. <https://doi.org/10.1007/978-3-030-86144-5>.
- Wilkinson, Richard G, and Kate Pickett. "The Spirit Level: Why More Equal Societies Almost Always Do Better." *Allen Lane* 330 (2009).
- Wolf, Susan. "The Importance of Free Will." *Mind* 90, no. 359 (1981): 386–405.
- Woodward, James. "Downward Causation Defended." In *Top-Down Causation and Emergence*, edited by Jan Voosholz and Markus Gabriel. Cham: Springer International Publishing, 2021. https://doi.org/10.1007/978-3-030-71899-2_9.
- Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books, 2019.