DIGITAL NUDGING: POTENTIAL AND PITFALLS

BY
AVISHALOM TOR

Professor of Law and Director, Research Program on Law and Market Behavior, Notre Dame Law School (ND LAMB). This article draws on my keynote lectures on "Digital Nudging" at the Lucerne conference on the *Law and Economics of the Digital Transformation* and on "The Costs of Personalized Digital Nudging" at the BIDEM conference on *Personalisation and Discrimination in a Fundamental Rights Perspective* at Copenhagen Business School, further developed in Avishalom Tor, *The Contours and Challenges of Digital Nudging*, in Law and Economics of the Digital Transformation (Klaus Mathis & Avishalom Tor eds., *forthcoming* 2023).

Visit www.competitionpolicyinternational.com for access to these articles and more!

DIGITAL NUDGING: POTENTIAL AND PITFALLS

By Avishalom Tor

Digital nudges — that is, significantly behavioral interventions that use software and its user-interface design elements — are an increasingly pervasive feature of online environments that can shape people's behavior both online (e.g. changing website cookie settings) and offline (e.g. taking a flu vaccine due to a text message reminder). While sharing many characteristics of offline behavioral interventions, digital nudges merit specific attention and analysis due to their growing ubiquity and potential potency, the opacity of their technological and behavioral mechanisms, and the central role of private actors in their implementation.

01
INTRODUCTION

To advance their policy goals, governments and other organizations have been employing behavioral instruments — also known as nudges — for some time now,[2] but the advent of digital nudges is more recent. Digital behavioral interventions are distinct from their offline counterparts in their deployment of software and its user-interface design elements and are an increasingly pervasive feature of online environments. These instruments can shape behavior online — e.g. when they encourage consumers to change their website privacy settings or to donate to a charity — as well as offline, as when people decide to take a flu vaccine at their annual medical checkup following a text message reminder from their health insurer.

Digital nudges share many features of offline behavioral interventions, yet merit particular attention and analysis due to their potential ubiquity across online platforms, social networks, other applications, and electronic devices, which brings into sharper relief the potential and pitfalls of nudges more generally. Moreover, digital nudging raises some unique — or at least qualitatively different — issues compared to offline nudging, because of its potentially greater potency (e.g. due to the possibility of personalized interventions using artificial intelligence, machine learning, and big data), the opacity of the technological and behavioral mechanisms through which it shapes people's judgments and decisions, and the central role of private intermediaries or independent private actors like internet platforms in its implementation.[3]

02
OFFLINE NUDGES: SOME BASICS

Behavioral regulation has been on the rise for some time now, beginning shortly after the publication of Thaler and Sunstein's 2008 book Nudge: Improving Decisions about Health, Wealth, and Happiness, which received widespread public attention.[4] Regulators and other policymakers increasingly turn to those significantly behavioral interventions as an integral part of their efforts to shape individual behavior in most major policy domains, including health, safety, education, finance, environmental protection, tax compliance, public service delivery and more.[5] Recent national responses to the coronavirus pandemic vividly illustrated this behavioral turn, with nudges employed to promote widespread vaccination, complement quarantine or masking mandates, or encourage social distancing practices.[6]

Nudging draws on behavioral science to inform policy design.[7] While traditional regulatory instruments affect behavior by imposing constraints (as mandates or bans do), using economic incentives (as in the case of taxes or subsidies), or disclosing unavailable or costly information, nudges rely on "softer" behavioral tools, like more effective or persuasive information presentation, the framing of the available choices, the selection of defaults, or the communication of social information.[8] Notably, while Thaler and Sunstein originally offered a somewhat narrow definition of nudging,[9] the expansive literature on behavioral regulation now encompasses a host of nudge usages, with most commentators

2 E.g. Klaus Mathis & Avishalom Tor (eds.), Nudging — Possibilities, Limitations and Applications in European Law and Economics (2016); Adam Oliver, The Origins of Behavioural Public Policy 110–11 (2017); Avishalom Tor, *The Law and Economics of Behavioral Regulation*, 18 Rev. L. & Econ. 1 (2022).

3 The preliminary assessment of digital nudges offered here focuses on the welfare effects of these instruments — namely, their private benefits and costs — though nudges raise other legal questions and normative concerns. See, e.g. the sources referenced in Avishalom Tor, *Nudges that Should Fail*, 4 Behav. Pub. Pol'y 316, n. 1 (2020).

4 Richard H. Thaler & Cass R. Sunstein, Nudge: Improving Decisions about Health, Wealth, and Happiness (2008). The book already sold over 2 million copies before the recent publication of an updated version as Richard H. Thaler & Cass R. Sunstein, Nudge: The Final Edition (2021).

5 For instance, in his front-cover praise of the Final Edition, Nobel prize winner Daniel Kahneman states: "Few books can be said to have changed the world, but *Nudge* did." This reality is reflected, for instance, in the OECD's Behavioral Insights web page reporting that there are 202 "institutions around the world applying behavioural insights to public policy" at https://www.oecd.org/gov/regulatory-policy/behavioural-insights.htm.

6 See generally Doron Teichman & Kristen Underhill, *Infected by Bias: Behavioral Science and the Legal Response to COVID-19*, 47 Am. J. of L. & Med. 205 (2021).

7 Brigitte C. Madrian, Applying Insights from Behavioral Economics to Policy Design, 6 Ann. Rev. of Econ. 663 (2014); Avishalom Tor, *The Critical and Problematic Role of Bounded Rationality in Nudging*, in Nudging — Possibilities, Limitations and Applications in European Law and Economics 3 (Klaus Mathis & Avishalom Tor eds., 2016).

8 Thaler and Sunstein, *supra* note 2; Avishalom Tor, *The Target Opportunity Costs of Successful Nudges*, in Consumer Law and Economics 3 (Klaus Mathis & Avishalom Tor eds., 2021).

9 Sunstein and Thaler, 2003, p. 120; Thaler and Sunstein, 2008.

using the term broadly, as a loose shorthand for policies with some behavioral component or connection.[10]

The popularity of offline nudging owes, in large part, to the perception that it offers a more palatable and cost-effective form of regulation.[11] Policy makers may believe that nudges are politically more feasible than traditional regulation, since large segments of the public — often a majority — in many democratic nations appear to find some common nudges acceptable.[12] Regulators in democratic societies may also prefer non-coercive behavioral interventions that leave citizens with greater freedom of choice than some forms of traditional regulation.[13] In addition, there is a widespread view that nudges make cost-effective policy instruments is due to their low implementation costs — that is, nudges do not require resource-intensive enforcement efforts as mandates or bans and do not otherwise burden public budgets as do some financial incentive polices (e.g. subsidies).[14]

Notwithstanding the benefits of behavioral interventions, however, more recent scholarship also highlights some of their limitations and costs. For one, empirical studies of nudge efficacy suggest that while nudges can be efficacious the absolute magnitude of their effects is often modest, with the notable exception of defaults that commonly have substantial effects on choice.[15] A recent meta-analysis (that excluded defaults) further found that the effect sizes of actual real-world interventions deployed by major governmental nudge units are substantially smaller than those reported in the academic literature and of limited absolute magnitude.[16]

> ***The popularity of offline nudging owes, in large part, to the perception that it offers a more palatable and cost-effective form of regulation***

These findings show that although nudges can produce behavior change at scale their real-world efficacy frequently may be limited. At the same time, the results of behavioral interventions in the academic literature reveal that some nudges — most notably, but not only, defaults — are capable of producing substantially larger effect sizes, with the efficacy of specific interventions depending on myriad factors of the particular nudge, including its specific features, the behaviors it targets, whether it complements a traditional intervention or substitutes for it, and more.[17]

Beyond concerns about nudge efficacy, current research further reveals that these policies can be much costlier than they appear. Specifically, nudges that entail only limited direct implementation costs can generate significant private costs, particularly when they are efficacious. These costs include direct cognitive, emotional, or monetary costs to some of the individuals targeted by behavioral policies, as well as the costs borne by private third parties due to behavior changes brought about by successful nudging.[18]

However, the most significant costs of most behavioral regulation typically are the private opportunity costs to individuals whose behavior it successfully changes.[19] All successful interventions, including those that make their targets better off on balance, entail opportunity costs — namely, the now-forgone benefits these individuals obtained from their former course of action. Yet, successful nudges are capable imposing even greater opportunity costs on people

10 See, e.g. Anne-Lise Sibony & Alberto Alemanno, *The Emergence of Behavioural Policy-Making: A European Perspective*, *in* Nudge and the Law: A European Perspective 1, 2 (Alberto Alemanno & Anne-Lise Sibony, eds., 2015).

11 For other attractions of behavioral regulation see Tor, *supra* note 2.

12 E.g. Janice Jung & Barbara Mellers, *American Attitudes Toward Nudges,* 11 Judgment & Decision Making 62 (2016); Sunstein et al. *Trusting Nudges? Lessons from an International Survey*, 26 J. Eur. Pub. Pol'y 1417 (2019).

13 Cass R. Sunstein & Lucia A. Reisch, Trusting Nudges: Toward A Bill of Rights for Nudging (2019).

14 E.g. Sibony and Alemanno, *supra* note 7.

15 Dennis Hummel & Alexander Maedche, *How Effective is Nudging? A Quantitative Review on the Effect Sizes and Limits of Empirical Nudging Studies*, 80 J. Behav. & Experim. Econ. 47 (2019).

16 Stefano DellaVigna & Elizabeth Linos, RCTs to Scale: Comprehensive Evidence from Two Nudge Units (2021) (SSRN working paper, https://www.nber.org/papers/w27594).

17 E.g. Hummel & Maedche, *supra* note 15; Katherine L. Milkman et al., *A Megastudy of Text-Based Nudges Encouraging Patients to Get Vaccinated at an Upcoming Doctor's Appointment*, 118 Proceed. Nat'l Acad. Sci., e2101165118 (2021).

18 Avishalom Tor, *The Private Costs of Behavioral Interventions*, 72 Duke L. J. (*forthcoming* 2023).

19 Avishalom Tor, *The Target Opportunity Costs of Successful Nudges*, in Consumer Law and Economics 3 (Klaus Mathis & Avishalom Tor, eds.).

when causing them to make personally detrimental behavior changes.[20] This is apparent, for instance, when regulators concerned with public welfare seek to reduce harmful environmental externalities by nudging consumers to conserve energy (e.g. by mailing them Home Energy Reports — letters that compare their consumption to that of their neighbors and imply a social norm favoring energy conservation.[21] All successfully nudged households inevitably forgo the benefits of their previous, higher energy usage (e.g. greater indoor comfort). Moreover, at least some energy consumers — like those who reduce usage only to avoid the "moral tax" aspect of a purported social-norm nudge — can end up bearing opportunity costs that exceed their benefits from lower energy consumption.[22]

As with traditional regulation, the behavior changes produced by nudging can impose economic costs on non-consumer third parties. To illustrate, Home Energy Reports that lead consumers to reduce their energy consumption produce net revenue losses for energy retailers due to their diminished sales.[23] From the perspective of energy retailers, in fact, the losses from reduced use are the same irrespective of the mechanism employed to change consumer behavior.

Of course, the often-substantial opportunity costs and other private costs that accompany successful behavioral interventions do not necessarily render these policies undesirable. Nudges increase social welfare when their overall benefits exceed their overall costs, and they are particularly attractive when they produce net private benefits — that is, when they improve individual well-being on balance enough to also make up for any attendant costs to third parties or to the public. Nonetheless, the prevalence and magnitude of private costs militate for requiring a demonstration that proposed behavioral interventions offer society net benefits, just as expected of traditional regulation.[24]

03
DIGITAL NUDGES: POTENTIAL

The more recent and ongoing development of digital nudges is at the intersection of behavioral science and technological innovation in digital environments. These behavioral instruments can be distinguished from their offline predecessors by the unique medium they use to deliver interventions. Specifically, digital nudges employ software and its user interface design elements[25] — those aspects of computer systems with which humans interact — to shape the behavior of the individuals they target.[26]

In our current technological environment, people spend a large portion of their time interacting with sophisticated computer systems, from personal computers, through smartphones, to countless other digital devices that pervade daily life at home, work, or other public and private settings. This reality increases both the opportunities for and the incidence of digital nudging. Local governments can nudge residents to pay their taxes on time by highlighting social norms of tax payment or presenting the penalties for overdue payments as psychologically painful losses; charitable organizations can nudge individuals for higher donations by offering donation menus that lead people more often to select favored options or triggering emotional reactions; social media platforms can nudge individuals to follow news from media outlets those platforms deem reliable; and even private email providers may nudge their customers with a simple reminder to follow up on an email they sent five days ago that received no reply.

Of course, while digital nudging occurs online, its behavioral effects are not limited to digital environments. Offline effects may occur incidentally, because the online behaviors that people are nudged towards have offline parallels: The nudged tax payment may be done with a physical check or even in person; the emotionally-triggered donation may take place at a local charity; successfully-nudged social media consumers may subscribe to a physical edition of a favored newspaper; and even the automated email nudge may lead one to knock on a colleague's office door to follow up in person on that answered email.

20 Tor, *supra* note 18.

21 Hunt Allcott, *Social Norms and Energy Conservation*, 95 J. Pub. Econ. 1082 (2011).

22 E.g. Hunt Allcott & Judd B. Kessler, *The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons*, 11 Am. Econ. J.: Applied Econ. 236 (2019).

23 *Id.*

24 E.g. Cass R. Sunstein, The Cost-Benefit Revolution (2018).

25 Markus Weinmann et al., *Digital Nudging*, 58 Bus. Inf. Syst. Eng. 433 (2016).

26 E.g. Debbie Stone et al., User Interface Design and Evaluation 4 (2005).

In such cases, the offline effects of digital nudging are mere byproducts of online interventions whose main goal is the shaping of online behavior. Yet digital nudges are frequently implemented specifically to change offline behavior. This is the case, for instance, with health apps or gadgets that nudge individuals to increase their physical activities; with text messages that nudge people to engage in targeted real-world health behaviors, like taking a flu vaccine; or with websites or mobile phone apps that closely resemble those physical Home Energy Reports and seek to lower consumers' home energy use.

Digital nudges also vary in the degree to which they are uniquely digital. Some nudges are digital primarily in the sense that they operate through some digital medium, as when citizens receive a "reminder" to pay taxes on time via email or text rather than by a physical letter in the mail. Other behavioral interventions are more uniquely digital in that they exploit features of the digital environment that are unavailable to their offline counterparts. For example, when originally conceived as an accompaniment to physical utility bills, Home Energy Reports could only offer monthly feedback on a household's overall consumption over a previous month, while the digital version of the same reports or similar apps are capable of providing ongoing, immediate, and far more granular, energy-use or social comparison information.

The advantages of digital nudges are not limited to their potentially rapid response times or their access to current information. For one, the flexible and technologically advanced nature of common digital interfaces allow nudgers to use a wide array of visual and auditory effects to direct attention, emphasize or deemphasize information, or trigger affective or intuitive psychological reactions, in ways that are usually unavailable to offline nudges. Additionally, digital nudges can use software that benefits from machine learning, big data analytics, and more to track and evaluate individual behavior and develop more nuanced and personalized interventions, with rapid content modification as new information is obtained. Studies show, for instance, how data on Facebook "likes" can predict different personal characteristics, such as demographics or even personality traits (e.g. extraversion or openness), with some accuracy.[27] Such predictions, in turn, can form the basis of more effective behavioral interventions that target these characteristics.[28]

Beyond the technological strengths of digital interfaces and the software underlying them, which may enable more effective behavioral interventions, digital environments also yield novel opportunities for nudging that do not exist offline. Online social networks (e.g. Facebook or Instagram) and other online social groups (such as gaming communities) are a familiar case on point. Such networks and communities are uniquely online fora, with no direct offline counterparts. They allow for data collection, analysis, and use in the service of behavioral policy interventions, just as they do for commercial interests (e.g. through advertising).

In one clever example of leveraging digital social interaction and technology to shape offline behavior, a Japanese COVID-19 contact-tracing mobile phone app included both a gaming element (getting a "fortune slip" when checking in at a new physical location where social interaction could take place) and a socially displayed digital art that becomes increasingly elaborate with each additional user who is physically present.[29] The latter element in particular aimed at producing social recognition effects that single out app users — as well as those who do not use the app — in social settings and might further the creation of social norms that favor contact tracing.

04
DIGITAL NUDGES: PITFALLS

The same factors that render digital nudges potentially more effective policy tools, however, also bring with them attendant risks and costs. Most importantly, digital behavioral interventions can generate opportunity costs and other private costs that are similar in kind but substantially greater in magnitude than the comparable costs of offline nudges. The greater magnitude of digital nudge costs is due to a combination of factors: First, more effective interventions usually generate higher private costs (irrespective of their benefits), and those unique characteristics of digital nudges that render them potentially more effective — such as their employment of AI, big data, rapid and dynamic personalization, or engaging and multi-sensory interfaces — are especially capable of leading people to make personally detrimental behavior changes; second, the opacity of the algorithms on which digital nudging relies means that it is

27 E.g. Michal Kosinski et al., *Private Traits and Attributes are Predictable from Digital Records of Human Behavior*, 110 Proceed. Nat'l Acad. Sci. 5802 (2013).

28 E.g. Sandra C. Matz et al., *Psychological Targeting as an Effective Approach to Digital Mass Persuasion,* 114 Proceed. Nat'l Acad. Sci. 12714 (2017).

29 Yuji Kanamitsu et al., *Using Interaction as Nudge to Increase Installation Rate of COVID-19 Contact-Confirming Application*, Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (2021).

more difficult to determine whether they or the behavioral processes through which they change individuals' judgments and decisions are detrimental; and, third, because most digital nudging takes place on private online platforms and websites, private intermediaries and other private actors have much greater influence on the goals and characteristics of these interventions than in the case of their offline counterparts, thereby increasing the likelihood that they will harm their targets.

The potentially greater private costs of digital nudges are nicely illustrated by the Japanese contact-tracing app described above, which uses both a gaming element and social recognition effects to encourage people to download and use the app. This app could be particularly powerful because it provides individuals with a strong social recognition signal, with publicly displayed digital art becoming visibly more elaborate whenever a person with an active app joins the social group. Notably, this public signal not only demonstrates to others that one is using the contact-tracing app, but also identifies those who are avoiding it by the lack of a change in the public digital art when then join others in a social setting. This powerful nudge may well pressure individuals who do not wish to surrender their privacy to nevertheless use the app to their personal detriment.

More generally, as this example demonstrates, the same tools and data that increase the potential effectiveness of digital nudging also tend to increase its private costs. The social recognition aspect of the Japanese contact-tracing app could not work without constantly tracking of its users' whereabouts, collecting, and using personal information, benefiting from rapid feedback, employing a multi-sensory interface, and so on. Yet by building upon such data and technology, this digital nudge may be particularly capable of leading individuals to accede to a contact-tracing method they would have otherwise refused.

As this example demonstrates, the same tools and data that increase the potential effectiveness of digital nudging also tend to increase its private costs

Similar technologies and data similar can underpin digital nudges that seek to promote other public welfare goals (e.g. a reduction in household energy consumption) or individual well-being (such as a more healthful diet or an increased rate of saving for retirement). In the case of many such common interventions, therefore, digital nudges may impose substantial private opportunity costs on many and occasionally also entail private costs to third parties (e.g. the net revenue losses to energy providers or less-healthful food sellers due to diminished consumption).[30] As with offline nudges, these private costs may not render digital behavioral interventions altogether unappealing, but they must be weighed, together with all other policy costs, against whatever private or public benefits these nudges provide.

Beyond concerns about the private costs that follow their greater efficacy, digital nudges typically also rely on opaque algorithms that make it exceedingly difficult to determine their benefits and costs. Like other current uses of artificial intelligence ("AI") in the commercial sphere, nudges can employ AI and machine learning ("ML") systems to identify their targets and determine when and how to approach them. These software systems are trained on and learn from a great deal of individual-level data (e.g. online behavior) that allow algorithms to predict which outcome would optimize a set of parameters.[31] Importantly, AI/ML systems are can act in ways that are not strictly pre-programmed and to adapt their actions to changing environments; once trained, they rely on recursive feedback to organically continue learning from new information to improve their predictions.[32]

Importantly, the adaptive nature of AI/ML systems can make it particularly difficult to determine precisely why they nudged a given individual. An AI/ML system designed to encourage household energy conservation, for instance, may seek individuals whose preferences it predicts to favor energy conservation, people it predicts to consume more energy irrespective of whether their preferences favor conservation, or simply consumers whom the system estimates to be most susceptible to a particular nudge based on their personality characteristics. Yet we may not be able to ascertain for which of these reasons an individual was selected for nudging, only that the AI/ML system predicted that nudging that person will best optimize its energy conservation parameters.

This uncertainty is exacerbated by the further challenge of identifying the specific behavioral processes an algorithm recruited to cause individuals to conserve more energy. After all, even offline a single intervention may recruit mul-

30 Avishalom Tor & Jonathan Click, *When Should Governments Invest More in Nudging? Revisiting Benartzi et al. (2017),* Rev. L. & Econ. (*forthcoming* 2023).

31 Karni Chagal-Feferkorn & Niva Elkin-Koren, *LEX AI: Revisiting Private Order by Design*, Berkeley Tech. L. J. (*forthcoming*).

32 *Id.*

tiple behavioral processes — as when the traditional Home Energy Reports included a combination of social comparison information (comparing the household to a group of "efficient neighbors"), purported injunctive social norms favoring energy conservation (with smiley faces for those who conserve more than average), and energy conservation tips.[33] Consequently, It may be unclear which specific behavioral process led to a behavior change in any given case. After all, the energy conservation nudge may have assisted some who already wished to conserve more energy than their peers to follow through (e.g. by providing a social comparison benchmark), in which case they were likely made better off. But it may have led others to reduce energy consumption for fear of violating a purported social norm (such as through smiling/unsmiling face icons) or even just because they were susceptible to the pressure of repeated reminders (e.g. of their daily energy consumption). In the latter cases, however, those who conserved more energy have been made worse off.[34]

Beyond its tendency to produce higher private costs that may be especially difficult to identify, the great majority of digital nudging occurs on private platforms and websites, which renders private intermediaries essential to their delivery. This reality also provides these intermediaries, as well as any other private actor who uses their services, with the ability to influence or even determine the goals and characteristics of digital nudges. Furthermore, unlike the legal scrutiny and limits of governmental regulatory interventions, privately initiated or executed digital nudges are subject to few constraints, particularly in the United States.

Of special concern is the fact that private online intermediaries already possess vast troves of personal data and sophisticated tools that use this data to great effect in their commercial dealings with consumers. The same capabilities that enable Google to personalize its search results or Facebook to provide its users with a personalized feed, for instance, can be used to nudge their users towards online and offline behaviors that they or their private customers (e.g. a non-profit environmental protection organization) wish to promote. In addition, more effective, privately initiated, digital nudging that is subject to little scrutiny is all the more concerning given its ability unobtrusively to advance controversial policy goals (e.g. encouraging Covid booster shots for young children or discouraging abortion) outside established legal institutions or public political debate.

05
CONCLUSION

All in all, it is apparent that digital nudges can offer more effective means for shaping people's judgments and decisions than offline behavioral instruments. However, this greater efficacy, together with the increasing ubiquity of digital nudging, the opaque means it employs in the service of behavior change, and the key role of online platforms and other private actors in delivering or commissioning digital interventions raise significant concerns that merit further critical evaluation. While digital nudging that draws on the same capabilities that successfully advance the commercial interests of private industry is here to stay, the various costs and risks associated with it should be weighed against its benefits, and appropriate responses — legal or technological — may well be needed to address its more egregious instances.

All in all, it is apparent that digital nudges can offer more effective means for shaping people's judgments and decisions than offline behavioral instruments

33 Tor, *supra* note 2.

34 *Cf.* Tor & Klick, *supra* note 30.

CPI SUBSCRIPTIONS

CPI reaches more than **35,000 readers** in over **150 countries** every day. Our online library houses over **23,000 papers**, articles and interviews.

Visit **competitionpolicyinternational.com** today to see our available plans and join CPI's global community of antitrust experts.

CPI COMPETITION POLICY™ INTERNATIONAL