

On the emergence and design of AI nudging: the gentle big brother?

Dirk Nicolas Wagner¹ ✉

Abstract

For more than a decade, ideas and concepts from behavioural economics have been spreading and become more and more influential in politics, business and civil society. Decision-makers in all sectors have become choice architects who aim to nudge people to make supposedly better decisions. More recently, analogue techniques have largely been superseded by digital nudging supported by artificial intelligence, i.e. artificially intelligent digital nudging. This viewpoint sketches key concepts as well as opportunities and risks. It shows that to be able to exploit the underlying combination of social and technical innovations, organizations can develop the necessary key competencies through recruiting and training talent but also by adopting ideation techniques like design-thinking.

Keywords: AI nudging, artificially intelligent digital nudging, behavioural economics, choice-architecture, design-thinking, digital nudging, nudge

Type: Viewpoint

Citation: Wagner, D. N. (2021). On the emergence and design of AI nudging: the gentle big brother? *ROBONOMICS: The Journal of the Automated Economy*, 2, 18

¹ Karlshochschule International University, Karlstraße 36-38, 76133 Karlsruhe, Germany; email: dwagner@karlshochschule.org

✉ Corresponding author



© 2021 The Author(s)

This work is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0).

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

1. Introduction

This viewpoint takes a look at a recent and rising trend of artificially intelligent digital nudging, or in short AI nudging. It observes that across all sectors of the economy and society actors invest in a combination of behavioural economics and artificial intelligence to influence individual decisions which means that AI nudging is permeating business and society at an exponential rate. For example, nudging is used to achieve double-digit increases in customer retention (Irrational Labs, 2021), to improve occupational safety (Nioi et al., 2020), to enhance retirement saving rates (Benartzi & Thaler, 2007), or to improve hygiene behaviours during a pandemic (Weijers & de Koning, 2021).

One aim of this contribution is to outline the scientific relevance of the above named emerging trend and to point towards developments that need to be better understood from a social sciences and a software engineering point of view. Along with this, the practical relevance of this viewpoint is to conceptually show how recent social and technical innovations are combined to pursue artificially intelligent digital nudging. To achieve this, the text commences with a clarification of important notions before it goes on to describe how social and technical innovations are combined, which leads to basic explanations of risks and opportunities, how artificially intelligent nudging scales and how it can be designed.

2. Choice architecture, nudging and (artificially intelligent) digital nudging

Decision-makers are becoming choice architects. A choice architect is someone who influences the choice of others by “organizing the context in which people make decisions” (Thaler et al., 2013, 428). Part of this context can be so-called ‘nudges’. A ‘nudge’ stands for “an aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives” (Thaler & Sunstein, 2021, 6).

The first signs of the coming upheavals in AI nudging are already well known to knowledge workers and managers. They are no longer surprised when the computer asks in the evening if they want to consider sending the just formulated message at a later stage during normal working hours in the following morning. And shortly thereafter, the gaze naturally and yet curiously wanders to the fitness app on the smartphone to determine whether one has moved sufficiently today despite working from home. But only a few people are aware that these inconspicuous digital notifications are the harbingers of an avalanche of change, especially in the professional interaction between man and machine. And those who are in the know are probably keen to maintain their knowledge advantage.

This wave of change is about ‘digital nudging’ (Schneider, Weinmann and Vom Broke, 2018) and thus about nudging human decisions with the help of software and its user-interface design elements to gently steer them into certain directions without fundamentally restricting the individual's freedom of decision or even directly issuing instructions. Digital nudges can in principle reach the decision-maker via all conceivable channels, starting with emails and smartphone apps via internal company software but also social media platforms. And artificial intelligence (AI), which can be understood as a growing resource of interactive, autonomous, self-learning agency, that enables computational artefacts to perform tasks that otherwise would require human intelligence (McCarthy et al., 2006), is increasingly at work in the background.

3. Combination of social and technical innovations

The groundbreaking effect of digital nudging is based on the combination - or should we say the intersection - of a social science and a technical innovation of recent years. On the one hand, there are the insights of behavioural economics, which were quickly disseminated in politics and business through popular books by Richard Thaler and Cass Sunstein (2008, 2021), Daniel Kahneman (2015), and Dan Ariely (2010). In the beginning, there is the insight that humans do not always behave rationally. A human being simply is not a homo oeconomicus. Rather, when it comes to making decisions, humans predominantly follow their gut feeling and

simple rules of thumb. According to behavioural science, this is faster, often right or at least good enough. As a consequence, people's decisions are influenced by factors that would be ignored by a rational homo oeconomicus (Thaler & Sunstein, 2021). What is problematic about this is that human decision-makers are regularly biased in the process. This can lead to mistakes and to decisions that are later regretted. For example, it is typically human to be too optimistic or to procrastinate and to postpone things that would have been better done long ago. These are only two supposedly small peculiarities of human behaviour with which even the failure of major projects worth billions can be explained (Lovallo & Kahneman, 2003; Wagner, 2016).

4. Cognitive biases motivate nudging

The list of cognitive biases is long (CEBM, 2021), now well researched, and it relentlessly reveals why no one is perfect and why to err is human. For some time now, behavioural economists and choice architects have systematically been countering them with the aforementioned nudges, thus ensuring desired behaviours or even desirable outcomes such as more organ donations, more environmentally friendly driving or healthier food choices in the cafeteria (Thaler & Sunstein, 2021; Ariely, 2010). Governments introduced so-called 'nudge units' (Halpern, 2019) to bring the necessary expertise to policymaking (Whitehead et al., 2014) and private organizations have gradually been increasing their use of nudging techniques both internally like for example in human resources management and occupational safety, as well as externally like in marketing and public relations (for more examples cf. Mele et al., 2021).

5. Machina Oeconomica always thinks along

To achieve nudging at scale is exactly where the announced technical innovation of AI fits in seamlessly. Because where man struggles, the computer is in its element. Machina oeconomica acts rationally, comes along with constantly growing computing power and tirelessly evaluates an even faster-growing sea of data. Across the economy, artificially intelligent agents are deployed to adopt highly specialized tasks which require the processing of large volumes of data including for example the processing of traces of human behaviours in digital environments. With the help of machine learning, artificial intelligence observes human behavioural patterns and then reports back to the choice architect in the nudge unit with pinpoint accuracy. This enables learning, both, at the level of the algorithm as well as at the human and thus architectural and social level. AI nudging is performed by artificially intelligent firms and organizations where humans and algorithms work hand in hand (Wagner, 2020) in the design and delivery of choice-architectures that nudge people in all sectors of the economy.

As a result, it is now increasingly rare for such decision architects to try to directly give us a nudge as a broad mass, for example with a billboard on the side of the motorway or writing and a picture on a packet of cigarettes. Instead, algorithms do the nudging. They are doing it more dynamically, unobtrusively, contextually and individually. People are already familiar with Google's autocompleting feature and with the effect of life-logging fitness apps (Lidynia, Offermann-van Heek & Ziefle, 2019). Amazon's logistics worker is guided by an electronic wristband, Deliveroo's bike courier is nudged to deliver faster (O'Byrne, 2019), and Uber drivers to work overtime (Scheiber, 2017).

At best, the influence of algorithms is about preventing people from making bad decisions and acting in their interests. However, as the examples show, digital nudging can be about decisions that are in the interest of others; in the interest of the political regime, the powerful retailer or the employer. In contrast to Thaler and Sunstein's well-meant original idea of libertarian paternalism put forward in the first edition of their now-classic book (Thaler & Sunstein, 2008), Puaschunder (2020) raises awareness for political and ethical concerns, mainly because AI-driven nudging enables the nudger to exploit uneducated citizens and this may increase social hierarchy and class division.

A broad classification of opportunities and risks of AI nudging is provided in Table I. The categories on the different levels can be combined in a multitude of ways which can be illustrated by the above-mentioned bike courier example: by deploying nudges, entrepreneurial managers of organizations exploit overconfidence bias on the level of the individual employee while politically maintaining the courier’s freedom of choice (libertarian paternalism). With increasing company success AI nudging here may contribute to scale effects leading to monopoly power and increased class division between the nudgers which are often represented the management of large scale tech-firms (organizations) and the nudged, i.e. employees and freelancers. The later symptoms are part of what has already been called „Uberisation of work“ (Webster, 2020) or „Taylorooism“ (McGaughey, 2018), where the artificially intelligent firm feature „the app as a boss?“ (Ivanova et al., 2018).

Table I. *Opportunities and risks of AI nudging*

Level	Opportunities	Risks
Political (macro)	Enhanced freedom of choice (libertarian paternalism)	Increased social hierarchy (class division)
Organizational (meso)	More entrepreneurship	Increased monopoly power
Individual (micro)	De-biasing of decisions	Exploitation of biases

6. Exponential development

The exponential unfolding of the intersection of nudging and AI is, like all exponential developments, difficult to grasp. Already today, the average computer and smartphone user has several thousand contacts a week with their employer's servers and with those of the major tech corporations Google, Amazon, Facebook, Microsoft, and Apple (Hill, 2019). These digital ecosystems now have all the prerequisites to digitally nudge their users and change, ideally improve, their decisions. And behavioural economists have long since begun to design industry-specific concepts for digital nudging. This is not just about delivery services, but also about the financial market, for example, when AI questions the decisions of investment fund managers, or about healthcare when AI prompts the treating doctor to consider alternative diagnoses to avoid confirmation bias, as well as about many other markets and spheres of life including the political arena and civil society.

7. Aiming for the adjacent possible

Regardless of industry and field, the question arises how the respective domain experts, i.e. company directors, politicians, volunteers, activists can come to grips with these two demanding fields of innovation, behavioural economics and artificial intelligence? The pioneers in the field, Thaler and Sunstein, suggest that leaders of organizations need to consider recruiting specialists with dedicated knowledge in behavioural economics as well as providing in-house training in behavioural science (Fusaro and Sperling-Magro, 2021). Yet, Caraban, Konstantinou and Karapanos (2020) bring in another social innovation that has recently spread to promote new ideas and organizational change: design thinking (Plattner, Meinel & Weinberg, 2009). This means pursuing what Koppl et al. (2015) call the “adjacent possible”: design-thinking methods like the so-called ‘nudge deck’ lower the threshold for decision-makers, managers, and software engineers who have had little contact with behavioural economics and AI to become choice architects for digital nudging (Caraban, Konstantinou & Karapanos, 2020). The combination of three recent innovations is an adjacent possible and enables choice architects to devise nudges at a pace, scale and scope that would have been unthinkable a decade ago. Which leads to the ultimate question: Are you still being nudged or are you already nudging?

8. Conclusion

This viewpoint took an interdisciplinary approach to summarise technological, economical and social developments that are in the process of being well researched within their respective fields, whose overall impact, however, is not yet well covered, neither in the scientific literature nor by practitioners in politics, business and civil society. It was shown that artificially intelligent digital nudging can be interpreted as an important but also controversial factor in the field of robonomics (Ivanov, 2021) which comes with a complex

variety of opportunities and risks. A conceptual outline linking theoretical building blocks from behavioural economics, institutional economics, artificial intelligence, and management science was provided and presented with a case-based approach. Ultimately, this viewpoint can be seen as a 'nudge' towards further research as well as political and managerial consideration.

References

- Ariely, D., & Jones, S. (2010). *Predictably irrational*. New York, NY: Harper Perrenial
- Benartzi, S., & Thaler, R. (2007). Heuristics and biases in retirement savings behavior. *Journal of Economic Perspectives*, 21(3), 81-104. DOI: 10.1257/jep.21.3.81
- Caraban, A., Konstantinou, L., & Karapanos, E. (2020). The Nudge Deck: A design support tool for technology-mediated nudging. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 395-406.
- Fusaro, R., & Sperling-Magro, J. (2021). Much anew about nudging. In *McKinsey Quarterly*, August 06. Retrieved 01.09.2021 from <https://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/much-anew-about-nudging>
- Centre for Evidence-Based Medicine -CEBM (2021). *Catalogue of bias*. Retrieved 01.09.2021 from <https://catalogofbias.org/>
- Halpern, D. (2019). *Inside the nudge unit: How small changes can make a big difference*. London: W.H. Allen.
- Hill, K. (2019). I cut Google out of my life. It screwed up everything. In *gizmodo.com*, January 29. Retrieved 01.09.2021 from <https://gizmodo.com/i-cut-google-out-of-my-life-it-screwed-up-everything-1830565500>
- Irrational Labs (2021). How can we reduce customer churn? In *irrationallabs.com*. Retrieved 01.09.2021 <https://irrationallabs.com/case-studies/google-how-can-we-reduce-customer-churn/>
- Ivanov, S. (2021). Robonomics: The rise of the automated economy. *ROBONOMICS: The Journal of the Automated Economy*, 1, 11-11
- Ivanova, M., Bronowicka, J., Kocher, E., & Degner, A. (2018). Foodora and Deliveroo: The App as a Boss? Control and autonomy in app-based management-the case of food delivery riders. Working Paper No. 108. *Hans-Böckler Stiftung, Forschungsförderung*. Retrieved 01.09.2021 from <http://hdl.handle.net/10419/216032>
- Kahneman, D. (2015). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux
- Lidynia, C., Offermann-van Heek, J., & Ziefle, M. (2019). Nudging vs. budging—users' acceptance of nudging for more physical activity. In *International Conference on Applied Human Factors and Ergonomics* (pp. 20-33). Cham: Springer. DOI: 10.1007/978-3-030-20476-1_3
- Koppl, R., Kauffman, S., Felin, T., & Longo, G. (2015). Economics for a creative world. *Journal of Institutional Economics*, 11(1), 1-31. DOI: 10.1017/S1744137414000150
- Lovallo, D., & Kahneman, D. (2003). Delusions of Success. How Optimism Undermines Executives' Decisions. *Harvard Business Review*, 81(7), 56-63.
- Nioi, A., Wendelboe-Nelson, C., Cowan, S., Cherrie, M., Rashid, S., Cowie, H., Davis, A., Ritchie, P., Lansdown, T., & Cherrie, J. W. (2020). Nudging construction workers towards better sun-safety behaviour: summary of the evidence for practitioners. *Policy and Practice in Health and Safety*, 18(1), 25-33. DOI: 10.1080/14773996.2019.1708614
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, 27(4), 12-12. DOI: 10.1609/aimag.v27i4.1904
- McGaughey, E. (2018). Taylorooism: when network technology meets corporate power. *Industrial Relations Journal*, 49(5-6), 459-472. DOI: 10.1111/irj.12228
- Mele, C., Spena, T. R., Kaartemo, V., & Marzullo, M. L. (2021). Smart nudging: How cognitive technologies enable choice architectures for value co-creation. *Journal of Business Research*, 129, 949-960. DOI: 10.1016/j.jbusres.2020.09.004
- O'Byrne, E. (2019). Slave to the algorithm? What it's really like to be a Deliveroo rider. *Irish Examiner*, February 20. Retrieved 01.09.2021 from <https://www.irishexaminer.com/lifestyle/arid-30905680.html>
- Plattner, H., Meinel, C., & Weinberg, U. (2011). *Design-thinking. Understand – Improve – Apply*. Berlin Heidelberg: Springer. DOI: 10.1007/978-3-642-13757-0
- Scheiber, N. (2017). How Uber Uses Psychological Tricks to Push Its Drivers' Buttons. *New York Times*, April 02. Retrieved 01.09.2021 from <https://www.nytimes.com/interactive/2017/04/02/technology/uber-drivers-psychological-tricks.html>
- Puaschunder, J. (2020). Artificial Intelligence and Nudging. In *Behavioral Economics and Finance Leadership*. Cham: Springer. 101-150. DOI: 10.1007/978-3-030-54330-3_6
- Schneider, C., Weinmann, M., & Vom Brocke, J. (2018). Digital nudging: guiding online user choices through interface design. *Communications of the ACM*, 61(7), 67-73. DOI: 10.1145/3213765
- Thaler, R., & Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. London: Penguin.
- Thaler, R., & Sunstein, C. (2021). *Nudge: The final edition*. London: Penguin.
- Thaler, R., Sunstein, C. R., & Balz, J. P. (2013). Choice architecture. In E. Shafir (Ed.), *The behavioral foundations of public policy*. Princeton, NJ: Princeton University Press, 428-439.
- Wagner, D. (2016). Breakin' the Project Wave - Understanding and avoiding failure in project management.: *PM World Journal*, 5 (1). <https://pmworldlibrary.net/article/breakin-the-project-wave-understanding-and-avoiding-failure-in-project-management/>
- Wagner, D. (2020). The nature of the Artificially Intelligent Firm - An economic investigation into changes that AI brings to the firm. *Telecommunications Policy*, 44(6), 101954. DOI: 10.1016/j.telpol.2020.101954
- Webster, E. (2020). The Uberisation of work: the challenge of regulating platform capitalism. A commentary. *International Review of Applied Economics*, 34(4), 512-521. DOI: 10.1080/02692171.2020.1773647

- Weijers, R. J., & de Koning, B. B. (2021). Nudging to increase hand hygiene during the COVID-19 pandemic: A field experiment. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 53(3), 353. DOI: 10.1037/cbs0000245
- Whitehead, M., Jones, R., Howell, R., Lilley, R., & Pykett, J. (2014). Nudging all over the World: Assessing the Impacts of the Behavioural Sciences on Public Policy. *ESRC Negotiating Neuroliberalism Project Report*. Swindon/Edinburgh: ESRC. Retrieved 01.09.2021 from <http://changingbehaviours.wordpress.com>

Received: 01/09/2021

Revised: 17/10/2021

Accepted: 17/10/2021