The Transparency of Nudging: Evaluating Its Impact on Personal Autonomy

Sabina Pajmon
Center for Cognitive Science
University of Ljubljana
Ljubljana, Slovenia
sabina.pajmon@pef.uni-lj.si

Toma Strle
Center for Cognitive Science
University of Ljubljana
Ljubljana, Slovenia
toma.strle@pef.uni-lj.si

Abstract

Nudges are a strategic approach that shapes decision-making environments and the presentation of options to steer individuals toward certain behaviors while maintaining their freedom of choice. The ethical concerns surrounding nudges center on their potential to undermine personal autonomy, particularly when individuals are unaware of the influence exerted on them (i.e., covert or non-transparent nudges). The proposed solution for preserving autonomy is to increase transparency, which includes disclosing the presence and purpose of nudges to the people that are being nudged. There are various types of nudges and different types and levels of transparency associated with them. The most problematic in terms of violating personal autonomy are the non-transparent ones, those that exploit automatic cognitive mechanisms (Type 1 nudges), those that use type transparency and those that disclose their nature only after the fact (ex post). New approaches such as nudge plus approach seek to protect personal autonomy by involving citizens in the creation of nudges and enhancing reflectiveness during the nudging process.

Keywords

Nudge, transparency, autonomy, ethics of nudging, nudge plus approach

1 Introduction

Over the past thirty years, psychology and behavioral economics have highlighted how various contextual factors systematically influence our decision-making and behavior. In public policy-making, these insights are crucial for effectively addressing societal challenges like global warming, obesity, and poor economic decision-making. The groundbreaking paper [1] and the book that followed that brought the importance of decision architecture to the attention of academics, policymakers, and the general public was Thaler and Sunstein's "Nudge: Improving Decisions About Health, Wealth, and Happiness" [2]. In their work, they propose various ways in which government and private organizations could encourage or "nudge" individuals toward actions beneficial to them, while promoting a method that preserves a strong commitment to freedom of choice. Behavioral insights show that the context of decision-making can lead us to act inconsistently with our otherwise well-informed intentions [2]. The traditional approach to public policy assumes people are perfectly rational economic subjects ("econs") who act optimally with accurate information and clear rules. While this is an admirable goal, Thaler and Sunstein warn that basing public policy on this ideal often leads to failure. The authors introduce the concept of a "nudge" and propose its use as a policy-making approach that can influence citizens' behavior while avoiding the pitfalls and issues of traditional regulatory approaches, such as prohibitions and punishments. The advantage of this approach is that policymakers can influence our choices and behavior in a cost-effective and efficient manner without restricting us with prohibitions or interfering with our choices [3]. Despite the high effectiveness and utility of nudges, ethical concerns arise regarding the preservation of autonomy, especially with nudges that operate covertly and influence us without our awareness. This article investigates various types of nudges and levels of transparency, with a focus on their implications for personal autonomy. It begins by elucidating key concepts—nudges, autonomy, and transparency—before analyzing how different types of nudges, alongside varying types and levels of transparency, affect the preservation or violation of personal autonomy. Additionally, the article proposes criteria for determining which types of nudges are compatible with the preservation of personal autonomy. Finally, it explores potential strategies to mitigate adverse impacts on autonomy, including enhancing transparency, fostering citizen participation, and integrating reflective practices into the design of nudges.

2 Definition of a nudge

Thaler and Sunstein define a nudge as any element of choice architecture that influences behavior in a predictable way without restricting options or significantly altering economic incentives [2]. A nudge subtly guides individuals toward better decisions while preserving freedom of choice by adjusting how choices are presented [4]. Unlike prohibitions or penalties, nudges steer behavior without limiting options. An example is placing healthy snacks at eye level in stores to encourage healthier choices [5]. Hausman and Welch [6] add that nudges influence choices without increasing costs or limiting options, highlighting the potential for manipulation, which raises ethical concerns discussed in later chapters.

Information Society 2024, 7–11 October 2024, Ljubljana, Slovenia

https://doi.org/10.70314/is.2024.cog.8

3 How nudges work: leveraging heuristics and biases

To grasp how nudges impact behavior and decision-making, we must rely on insights from behavioral science, which reveal that nudges exploit inherent imperfections in human decision processes—leveraging cognitive heuristics and biases [4]. A key element of the nudge approach is that heuristics and biases, which often serve as mental shortcuts, are utilized to the advantage of the choice architect. While these mental shortcuts can sometimes lead to suboptimal decisions, a nudge aims to harness them to promote better decisions [3].

4 Nudge and ethical issues

Although the theory of nudging presents a promising approach to public policy, it has faced significant criticism from both public and academic spheres. Over the past decade, a robust ethical debate has developed, featuring nuanced arguments both supporting and opposing the practice [3, 4, 6, 7, 8]. The primary critique centers on the idea that nudging involves manipulating choices, with concerns about potential misuse of power [3]. Critics argue that nudges can undermine free choice by subtly restricting rather than fostering individual decision-making. The core of nudging involves exploiting heuristics and biases, which often lead people to act in ways that deviate from their well-considered preferences. Bovens [8] contends that such mechanisms can compromise control over actions, raising worries that nudges might affect decision-making by diminishing rational or deliberate considerations. Additionally, he argues, the behavior change induced by nudges occurs, if not against citizens' will, then at least without their active consent and awareness; for broader discussion about this topic see also [9, Ch. “Avtonomija v svetu spodbud” (Autonomy in the World of Nudging), pp.. 81-100; 10].

4.1 Ethical dilemma of autonomy in nudge use

Although nudges have been shown to effectively influence behavior, critics argue they can be manipulative and threaten personal autonomy. Autonomy, a complex concept, is broken down by Schmidt and Engelen into four dimensions: the freedom to choose without external pressure, acting according to one’s desires and values (psychological autonomy), making rational decisions based on available information, and being free from domination or manipulation [5, 8]. Critics claim that nudges can undermine autonomy by subtly influencing behavior without explicit consent, raising concerns about democracy, especially if governments use nudges without informing citizens. Nudges that operate without notice are especially problematic, as they can influence decisions without individuals' awareness. As Ivanković and Engelen [11] argue, non-transparent nudges, which exploit less rational psychological mechanisms, undermine autonomy by denying people control and the ability to challenge, a right that should be protected in liberal democracies. Sunstein, however, argues that nudges maintain freedom by allowing people to opt out of the suggested behavior, a concept they call "libertarian paternalism" [12]. They believe nudges, unlike traditional regulations, don’t limit freedom but instead encourage choices that align with individuals' best interests.

4.2 Transparency as a solution to autonomy violation

A proposed solution is increasing nudge transparency, as it allows individuals to understand how nudges work and make autonomous decisions based on their values [8]. A transparent nudge is one where its purpose and the methods used to influence behavior are reasonably clear to the affected individual. Thaler and Sunstein moreover argue that nudges used by governments should be public and transparent, with officials ready to disclose their methods and motives. Sunstein further emphasizes that nudges must be visible, reviewed, and monitored to prevent violations of autonomy or dignity [10]. Transparency involves informing decision-makers about the presence and purpose of nudges, allowing individuals to remain aware of behavioral interventions, thus preserving their autonomy and freedom of choice [7]. [1]

5 Types of transparency in nudging

To better understand the impact of nudging on an individual's autonomy, it is crucial to first examine the different types of nudges, as they are not a uniform phenomenon; rather, they can be classified into various types [2]. Understanding these types is based on dual-process theory, which describes the two decision-making mechanisms that nudges can influence.

5.1 Dual process theory

Dual process theory, explored by Stanovich [13] and Kahneman [14], is key in Thaler and Sunstein’s work on nudges. It suggests the brain operates in two modes: fast, intuitive System 1 and slow, deliberate System 2. System 1 handles instinctive actions, while System 2 engages in reflective decision-making. Despite its acceptance, dual processing is contested, with some scholars arguing the differences are a matter of degree. De Neys [15], notes no conclusive evidence favors either model, and resolving this debate may not significantly enhance our understanding of human thinking mechanisms. In this article, we adopt the dual process theory model to categorize different types of nudges. This approach allows us to better understand and design interventions that leverage both intuitive and reflective processes.

5.2 Type 1 and Type 2 nudges

According to Hansen and Jespersen [3], nudges can be categorized into two types based on dual process theory. Type 1 nudges target automatic, non-reflective thinking (System 1) and operate unconsciously, such as subliminal advertising or visual stimuli that influence behavior without conscious awareness. These nudges can be ethically problematic, as they often lack transparency and may lead to decisions misaligned with personal values. In contrast, Type 2 nudges engage reflective, deliberate thinking (System 2), promoting informed and thoughtful decision-making. These nudges are transparent and pose fewer ethical concerns regarding personal autonomy.

[1] Empirical evidence is inconsistent regarding the impact of transparency on the effectiveness of nudges. Transparency may: reduce their effectiveness (by prompting reflection), make nudges counterproductive (if people resist disliked nudges), enhance their effectiveness (if people understand and support the underlying goals), or have no significant impact at all [10].

5.3 Different types of transparency of nudges

The transparency of nudges plays a crucial role in safeguarding autonomy and freedom of choice, yet this concept itself is multifaceted. On one side of the spectrum, some nudges are explicitly transparent, functioning effectively because the individual is fully aware of the influence being exerted. Conversely, some nudges operate more subtly, relying on a lack of transparency to achieve their intended effect. To thoroughly assess which forms of transparency in nudges may raise ethical concerns, it is important to analyze the various ways in which transparency can manifest within nudges.

5.4 Type and token transparency

Bovens [8] introduces a crucial distinction between type and token transparency in nudges. Type transparency refers to when governments inform citizens about the general techniques they employ to intervene in decision-making contexts for the purpose of enhancing well-being. In this scenario, the government is open about the categories of measures it plans to implement. For example, when a government announces its intention to use specific psychological mechanisms to address social challenges, it demonstrates type transparency by clearly stating the kinds of interventions it will use to influence individuals' behavior and decision-making [16]. However, Bovens stresses that this is not enough. In his view, subliminal advertising does not become more acceptable simply because it is openly acknowledged [8]. On the other hand, token transparency requires that each individual instance of a nudge is clearly recognizable, including how it was implemented. This method, referred to as "here and now approach," aims to ensure that nudges are transparent to those encountering them at the moment of their decision-making [12]. However, even if this were feasible, it seems absurd to demand that every nudge be accompanied by a notice of its use. Since choice architecture is often unavoidable, token transparency may be too demanding, according to Bovens [8].

5.5 Levels of transparency

Transparent nudges differ also based on when they are noticed by the nudged individuals. With nudges that are transparent in advance (*ex ante*), the user can see the nudge beforehand and can avoid it if they choose. An example is traffic light labels (green, yellow, red) for healthy, less healthy, and unhealthy food products [17]. In contrast, a nudge is transparent afterward (*ex post*) if the target person only notices its influence after it has already affected them. Examples include fake cracks painted on the road to slow down drivers or the use of default options in certain contracts. Only after experiencing the effects do people realize they were influenced by a nudge [12]. Unlike the first category, the potential impact of such nudges on people's autonomy is more significant here. Ex post transparency may be insufficient to ensure autonomous action if it depends on individuals' ability to avoid the nudge. If transparency is meant to ensure that nudges do not deter people from achieving their goals and values, then, according to Ivanković & Engelen, ex post transparent nudges should either be excluded or efforts should be made to turn ex post transparency into ex ante transparency [11]. Occasionally, ex post transparent nudges become ex ante transparent through repeated exposure. For example, a fake speed bump may not have the same effect twice if the person learns when and where to expect it. With repeated exposure to such nudges, individuals may become more aware of their influence and may eventually avoid them altogether [12].

6 Types of nudges and transparency: impact on personal autonomy

The debate over nudges centers on how different types of nudges as well as types and levels of transparency impact personal autonomy. As stated in the article, nudges are divided into two types: Type 1, which influence automatic, non-reflective behavior, and Type 2, which target reflective decision-making. Transparent Type 2 nudges, which engage reflective capacities, do not typically raise ethical concerns, as they allow for conscious and deliberate decision-making. In contrast, non-transparent Type 1 nudges, which act on automatic processes, can threaten autonomy by influencing behavior without the individual's awareness. This may lead to decisions misaligned with personal values or goals. Transparency is categorized into type transparency (general awareness of the nudge type) and token transparency (awareness of mechanisms of specific nudges). The former is particularly problematic, as it lacks disclosure of specific examples and mechanisms, leaving us potentially unaware of the influences on our behavior. Nudges can also be categorized by the level of transparency into two main groups. The first group includes nudges that are transparent in advance by design (ex ante). These nudges are openly presented, allowing users to consciously decide whether to respond to them. Such nudges generally do not threaten autonomy, as they encourage conscious and deliberate decision-making. The second group includes nudges that are only transparent afterward (ex post). These nudges can be problematic, as users may respond to them before realizing they have been nudged. Although information about the nudge is revealed later, it may already have influenced behavior in a way that threatens freedom of choice and autonomy [18]. In conclusion, the most problematic nudges, in terms of violating personal autonomy, are Type 1 nudges that exploit automatic cognitive mechanisms, lack transparency—where type transparency is more concerning than token transparency—or are only transparent afterward. Understanding and using nudges requires careful consideration of their transparency and impact on freedom of choice. While transparent nudges can serve as tools for encouraging thoughtful and autonomous decisions, non-transparent nudges, as well as Type 1 nudges, especially those with only type or post hoc transparency, must undergo thorough ethical scrutiny to prevent potential violations of personal autonomy.

Table 1: Classification of nudges based on their impact on personal autonomy

Nudges that violate autonomy	Nudges that do not violate autonomy
Type 1 nudges	Type 2 nudges
Type transparency	Token transparency
Ex post transparent nudges	*Ex ante* transparent nudges

This table helps determine whether a nudge preserves autonomy, but it's unclear how many criteria must be met to deem a nudge ethical or unethical. Further research is needed for clearer guidance.

6.1 Collaborative policy design: The nudge plus approach

The nudge plus approach extends beyond transparency by encouraging participatory engagement and reflection, viewing individuals as rational, reflective beings rather than passive agents. Unlike traditional nudging, which can influence behavior unconsciously, nudge plus focuses on democratic control and active collaboration between citizens and policymakers. Through methods like citizens' assemblies, participants are directly involved in policy design, contributing ideas that shape their environments. In the UK, medical sciences now require patient and public involvement in all research that includes patient populations. Similarly, adolescents are consulted in developing anti-bullying interventions [19]. These approaches foster mutual feedback and collaboration between policymakers and citizens, leading to more inclusive and transparent policies that respect community values. Nudge plus approach also refers to an intervention that has a reflective strategy embedded into the design of a nudge. Banerjee and John [20] state that this preserves personal autonomy while promoting pro-social interventions through active involvement by enhancing token transparency and decision-making autonomy. The nudge plus approach offers significant potential for enhancing public policy with maintaining individual autonomy. By embedding reflection, transparency and active citizen engagement, it encourages people to participate in decision-making rather than passively accepting nudge type interventions. This participatory approach builds trust, as individuals are more likely to embrace policies that respect their autonomy and align with their values.

7 Conclusion

This article reviewed nudges as tools for influencing decision-making and behavior, with a focus on their transparency and its influence on potential infringement of personal autonomy. We found that nudges vary in type and transparency, which significantly affects their ethical acceptability. Type 1 nudges, which target automatic decision-making mechanisms, can diminish personal autonomy by influencing behavior without conscious awareness. In contrast, Type 2 nudges, which encourage reflective decision-making, are less problematic as they support autonomous decision-making. Nudges that lack token transparency or are only transparent after the fact are more likely to infringe on autonomy. Conversely, when nudges are transparent in advance and individuals are informed about them, autonomy is better preserved. Additionally, the context in which nudges are implemented plays a critical role in their ethical assessment, as the goals and values of the intervention must align with those of the individuals affected. In conclusion, the ethical use of nudges in public policy requires focusing on preserving autonomy by choosing Type 2 nudges and ensuring high levels of transparency, especially regarding specific examples and advance notice. This approach allows nudges to support conscious decision-making rather than serving as tools for covert manipulation. The Nudge Plus approach, which adds an element of reflection, can enhance both the effectiveness and ethicality of interventions, empowering individuals to make more informed decisions.

Acknowledgments

This pilot research study was partly supported by The Green Nudge project (“UL za trajnostno družbo – ULTRA”) - European Union - NextGenerationEU, and Republic of Slovenia, Ministry of Higher Education, Science and Innovation.

Authors’ statement

ChatGPT-4 was used for improving language of this paper.

References

[1] R. H. Thaler and C. R. Sunstein, “Libertarian Paternalism,” American Economic Review, vol. 93, no. 2, pp. 175-179, May 2003.

[2] R. H. Thaler and C. R. Sunstein, Nudge: Improving Decisions about Health, Wealth, and Happiness. New Haven, CT: Yale University Press, 2008.

[3] P. G. Hansen and A. M. Jespersen, "Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy," 2013.

[4] A. T. Schmidt and B. Engelen, "The ethics of nudging: An overview," DOI: 10.1111/phc3.12658, 2019.

[5] T. Bucher et al., “Nudging consumers towards healthier choices: a systematic review of positional influences on food choice,” British Journal of Nutrition, vol. 115, no. 12, pp. 2252–2263, 2016. doi:10.1017/S0007114516001653

[6] R. A. Abumalloh, O. Halabi, R. Ali, and D. Al-Thani, “Nudging Techniques: Design, Theoretical Grounds, and Ethical View,” Journal of Knowledge Economy, 2024. [Online]. Available: https://doi.org/10.1007/s13132-024-02219-x.

[7] D. M. Hausman and B. Welch, "Debate: To Nudge or Not to Nudge," The Journal of Political Philosophy, vol. 18, no. 1, pp. 123–136, 2010.Sam An

[8] L. Bovens, "The Ethics of Nudge," in T. Grüne-Yanoff and S. O. Hansson, Eds., Preference Change: Approaches from Philosophy, Economics and Psychology, Theory and Decision Library A 42, Springer Science+Business Media B.V., 2009, ch. 10.

[9] T. Strle in O. Markič, O odločanju in osebni avtonomiji, 1. izd., let. 20. Maribor: Aristej, 2021, str. 145.

[10] Kognitivna znanost, Kognitivna znanost: zbornik 22. Mednarodne multikonference Informacijska družba - IS 2019, 10. oktober 2019: zvezek B = Cognitive Science. Ljubljana: Institut „Jožef Stefan“, 2019. [Na spletu]. Dostopno na: http://library.ijs.si/Stacks/Proceedings/InformationSociety

[11] V. Ivanković and B. Engelen, “Nudging, Transparency, and Watchfulness,” Social Theory and Practice, vol. 45, no. 1, pp. 43-73, Jan. 2019, doi: 10.5840/soctheorpract20191751.

[12] C. R. Sunstein, “The Ethics of Nudging,” Yale Journal on Regulation, vol. 32, no. 2, pp. 413-450, 2015.

[13] K. E. Stanovich, Who is Rational?: Studies of Individual Differences in Reasoning. Milton: Psychology Press, 1999.

[14] D. Kahneman, Thinking, Fast and Slow. London: Penguin, 2012.

[15] W. De Neys, “On dual-and single-process models of thinking,” Perspectives on Psychological Science, vol. 16, no. 6, pp. 1412-1427, 2021.

[16] K. Dowding and A. Oprea, "Nudges, Regulations and Liberty," British Journal of Political Science, vol. 53, pp. 204–220, 2023, doi: 10.1017/S0007123421000685.

[17] A. Arno and S. Thomas, “The efficacy of nudge theory strategies in influencing adult dietary behaviour: a systematic review and meta-analysis,” BMC Public Health, vol. 16, no. 676, pp. 1-13, 2016, doi: 10.1186/s12889-016-3272-x.

[18] J. Wachner, M. Adriaanse, and D. De Ridder, "The influence of nudge transparency on the experience of autonomy," Comprehensive Results in Social Psychology, vol. 5, no. 1-3, pp. 49-63, 2021, doi: 10.1080/23743603.2020.1808782.

[19] J. K. Madsen, L. de-Wit, P. Ayton, C. Brick, L. de-Moliere, and C. J. Groom, "Behavioral science should start by assuming people are reasonable," Science & Society, vol. 28, no. 7, pp. 583-585, Jul. 2024. doi: 10.1016/j.tics.2024.04.010.

[20] S. Banerjee and P. John, "Nudge plus: incorporating reflection into behavioral public policy," Behavioural Public Policy, vol. 8, pp. 69-84, 2024. doi: 10.1017/bpp.2021.6.