

Nudging: To know ‘what works’ you need to know why it works

Pelle Guldborg Hansen^{1*}

Abstract

Nudging is widely portrayed as a purely inductive approach to influencing human behavior using insights from the behavioral sciences to learn what works. However, as this paper argues, to understand ‘what works’, requires not only scientists, but also policy-makers as well as practitioners to understand what cognitive mechanisms brings behavior change about as well as under what conditions. This is argued by explicating how the concept of nudge itself identifies the main condition for the efficacy of nudging as well as calls for considering what specific mechanisms mediate a nudge and its behavioral effects. The practical implications are illustrated relative to the intuitively appealing policy application of nudging people into becoming organ donors by changing the default from an opt-in to an opt-out system; and in turn reveals why prominent scientists in the field believe this policy application to be a bad idea.

Keywords

nudge — definition — what works — organ donation

¹Roskilde University & iNudgeyou – The Applied Behavioural Science Group

*Corresponding author: pgh@ruc.dk

Introduction

In recognizing the potential of applying insights from the behavioral sciences to policy, Thaler and Sunstein (2008, p. 6) coined their concept of ‘nudge’:

“A nudge, as we will use the term, is any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives”.

Various revisions have since then been provided in order to clarify issues of precision and consistency for reconciling the definition with its theoretical underpinnings in the behavioral sciences (see Hausman & Welch, 2010; Hansen, 2016).

Here we follow the definition suggested by Hansen (2016):

“a *nudge* may be regarded as any function of (1) an attempt at influencing people’s judgment, choice or behavior in a predictable way, that is (2) made possible because of cognitive boundaries, biases, routines, and habits in individual and social decision-making posing barriers for people to perform rationally in their own self-declared interests, and which (3) works by making use of those boundaries, biases, routines, and habits as integral parts of such attempts”.

This revised definition implies Thaler and Sunstein’s definition since it follows that nudges work due to reasons that

go beyond (i) forbidding or adding any rationally relevant choice options, (ii) changing incentives, whether regarded in terms of time, trouble, social sanctions, economics and so forth, or (iii) the provision of factual information and rational argumentation (*ibid*).

Theoretical considerations like these are often treated as secondary by policy-makers and practitioners. A contributing factor to this is the main methodological characteristic of nudging, i.e. its use of the experimental methods drawn from the behavioral sciences, which has led to a widespread portrayal of nudging, and the grander paradigm of Behavioral Insights (BI), as a radical empirical paradigm relying on a purely inductive experimental approach to understanding ‘what works’ (Haynes et al, 2013; Lunn, 2014; Halpern, 2015). Consequently, considerations about the ways in which behaviorally informed policies affect behavior (3) and under what conditions (2), are often neglected.

However, for any policy paradigm making claims to be evidence based, theorizing is never, nor primarily, of mere academic interest. To work effectively and responsibly within any applied field requires one to have a profound theoretical understanding allowing for the consistent interpretation of real-world phenomena observed as well as attempts at influencing these (Robson & McCartan, 2016).

Practical implications of theory

The simplest way to show this relative to nudging is by observing how the revised definition itself calls attention to this point. Thus, (1) in the above definition emphasizes the inten-

tional nature of nudges. This prevents one from using the fact that ‘people are always being influenced by contextual factors anyways’ to mistakenly conclude that there is *carte blanche* to nudge people and ignore ethics. Intent is what separates nudging from accidental influence; enter ethics (Hansen & Jespersen, 2013). Yet, this mistake is quite widespread in the field even amongst its most prominent proponents (see Osman, 2016).

In turn, (2) identifies the main condition under which nudges in general may be expected to work; namely when cognitive boundaries, biases, routines, and habits in individual and social decision-making prevents people to perform rationally in their own self-declared interests. Thus, nudges apply to behaviors that occur despite people having good subjective reasons to act otherwise. As emphasized by Hausman and Welch (2010, p. 126) nudges are called for “because of flaws in individual decision-making”. This main condition for nudges to work is crucial as it follows, that nudges should not be expected to work, or at least not stick, unless they target behaviors that satisfy this condition.

Finally, (3) defines the broad categories of cognitive features that nudges make use of, *viz* the cognitive boundaries, biases, routines, and habits – i.e. the same type of features that causes the targeted behaviors. In so far as the efficacy of the specific cognitive features utilized by nudges and referred to by (3) are not coincidentally related to the cognitive features involved in the targeted behavior referred to by (2), this points to the crucial issue that nudges cannot be effectively applied without understanding *why* (mechanisms) as well as under *what circumstances* (specific conditions) their application may be expected to cause behavior change (Marchionni & Reijula, 2018; Hansen, 2019).

To illustrate, we may look at the so-called ‘default effect’ widely referenced in the nudging and BI literature in relation to its intuitively appealing policy application to get more people to register as organ donors.

Nudging defaults in organ donation

Since the publication of *Nudge*, policy makers around the world have been attracted to the idea of enrolling citizens as organ donors by default. By changing the default, the expectation is that more people will end up as potential organ donors due to the default effect. Despite Thaler and Sunstein themselves rejecting this policy proposal, policy-makers are often found arguing for adopting the policy on the grounds that it is a ‘nudge’ that ‘alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives’. Adopting this approach also appears attractive as it allows policy-makers to argue for influencing citizens’ decision-making, while at the same time respecting the generally accepted ethical position that organ donation is a private, personal and important issue, where no one can know what’s right or should try to pressure people into making a particular decision.

However, this is where the practical implications of theory become important.

First, we observe that a *default* is defined as an aspect of choice architecture, where one particular choice option is chosen as the pre-set choice such that people have to make an active decision to choose an alternative choice option (Johnson et al, 2002, 6). A *default effect* is then defined as the change in likelihood that a particular alternative is chosen when designated as the default versus a control condition when no default is designated (Brown and Krishna, 2004, 530). However, the definition of the default effect does not point to the specific cognitive mechanisms that might mediate a stimulus and its resulting behavioral effect. Yet, this is important to consider as a default effect may be brought about by several different mechanisms four of which are: *the default effect by inattention* (Johnson et al, 2002); *the default effect by recommendation* (McKenzie et al, 2006; Gigerenzer 2008: 24); *the default effect by loss aversion* (Tversky & Kahneman, 1991: 1040); and *the default effect by cognitive avoidance* (Johnson & Goldstein, 2003: 1338).

Relative to Condition (2) and (3), this implies that policy-makers and practitioners should only expect a change of default to be effective and stick in so far as people are either inattentive, uncertain about what they believe or what they prefer, or want to avoid the cognitive effort involved in making and sticking to the decision. Observing these special conditions should guide policy-makers and practitioners when forming hypotheses and interpretations. In particular, if expecting citizens to be inattentive to a change of default, the default may be expected to translate directly into a behavioral effect. Further, if one expects that citizens are uncertain about what to believe relative to organ donation, one may expect the default effect to be moderated by trust in the choice architect’s judgment, as for instance has happened in Brazil (Csillag, 1998). Also, if expecting citizens to be uncertain about what they prefer relative to organ donation, one should expect the default effect to depend on, amongst other factors, the reference point defined by the default choice relative to alternatives in terms of loss aversion (Sallis, 2018). Finally, if expecting that citizens are insufficiently motivated to go through the procedure of registering as an organ donor, one should expect the effect to be moderated by motivation, i.e. they should expect the least motivated to be affected most by the change of default.

This in turn is not only of importance relative to efficacy. Returning to condition (1), if a change of default is intended to cause its effect by *inattention*, *recommendation*, or *loss aversion* it renders the policy proposal ethically inconsistent; a change of default intended to work by inattention does not treat the decision as personal and important; recommendation, does not treat the decision as one where no one can know what’s right; and loss aversion implies people are pressured into making the decision favored by these policy-makers. Only the adoption of an opt-out system based on a *default effect by cognitive avoidance* appears to be reconcilable with the

ethical position. In practice, however, one effect can rarely be obtained without the others. It is for this reason that many prominent scientists in the field believe this policy to be a bad idea.

These considerations also reveal the more general point, that if expectations about what works are not based on the mechanisms involved as well as the main and specific circumstances under which they are supposed to work, policy-maker and practitioners will not only be incapable of nudging effectively and responsibly. They will also be incapable of interpreting the effect of tests and interventions (Marchionni & Reijula, 2018; Grüne-Yanoff, 2016) – that is they will be incapable of determining 'what works'.

Conclusion

When it comes to nudging, to know 'what works' you need to know why it works.

References

- Brown, C. L., & Krishna, A. (2004). The Skeptical Shopper: A Metacognitive Account for the Effects of Default Options on Choice. *Journal of Consumer Research*, 31(3), 529–539.
- Csillag, C. (1998). Brazil Abolishes 'Presumed Consent' in Organ Donation. *The Lancet*, 352(9137), 1367–1379.
- Gigerenzer, G. (2008). Why Heuristics Work. *Perspectives on Psychological Science*, 3(1), 20–29.
- Grüne-Yanoff, T. (2016). Why behavioural policy needs mechanistic evidence. *Economics and Philosophy*, 32(3), 463–483. doi.org/10.1017/S0266267115000425.
- Halpern, D. (2015). *Inside the Nudge Unit: How small changes can make a big difference*. WH Allen.
- Hansen, P.G. (2016). The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove? *European Journal of Risk Regulation*, 7(1), 155–174.
- Hansen, P.G. (2019). *BASIC: Behavioural Insights Toolkit and Ethical Guidelines for Policy Makers, Organisation for Economic Cooperation and Development*. OECD Publishing.
- Hansen, P., & Jespersen, A. (2013). Nudge and the Manipulation of Choice. *European Journal of Risk Regulation*, 4(1), 3–28.
- Hanyes, L., Service, O., Goldacre, B., & Torgerson, D. (2013). *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*. The Behavioural Insights Team.
- Hausman, D., & Welch, B. (2010). Debate: To Nudge or Not to Nudge. *Journal of Political Philosophy*, 18(1), 123–136.
- Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). Defaults, Framing, and Privacy: Why Opting In ≠ Opting Out. *Marketing Letters*, 13(1), 5–15.
- Johnson, Eric J., & Goldstein, D.G. (2003). Do Defaults Save Lives? *Science*, 302, 1338–1339.
- Lunn, P. (2014). *Regulatory Policy and Behavioural Economics*. OECD Publishing.
- Marchionni, C., & Reijula, S. (2018). What is mechanistic evidence, and why do we need it for evidence-based policy? *Studies in History and Philosophy of Science*, 73, 54–63.
- McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, 17(5), 414–420.
- Osman, M. (2016). Nudge: How Far Have We Come? *AEconomia – History, Methodology, Philosophy*, 6(4), 557–570.
- Robson, C., & K. McCartan (2016). *Real world research* (4th ed.). Wiley.
- Sallis, A., Harper, H., & Sander, M. (2018). Effect of persuasive messages on National Health Service Organ Donor Registrations: a pragmatic quasi-randomised controlled trial with one million UK road taxpayers. *Trials*, 19, 513. doi.org/10.1186/s13063-018-2855-5.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: improving decisions about health, wealth, and happiness*. Yale University Press.
- Tversky, A., & Kahneman, D. (1971). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics*, 106(4), 1039–1061.