doi: 10.15678/ZP.2019.48.2.6

Zarządzanie Publiczne
Public Governance
No. 2(48)/2019
ISSN 1898-3529

Agnieszka Wincewicz-Price

Does behavioural economics equip policy-makers with a complete (enough) picture of the human: the case of nudging[1]

Abstract

Objectives: The article offers a critical discussion of the policy of nudging and suggests so far unexplored evaluation criteria for behavioural policy experts and practitioners.

Research design: A multi-disciplinary approach is taken here to fill out the thin anthropology of *homo economicus* – which is shown to inform the concept of nudging – with selected aspects of human agency which are commonly discussed in moral, political and economic philosophy. The aim of this paper is twofold: 1) to outline the conceptual shortcomings of the behavioural foundations of the nudge theory as it has been originally proposed by Thaler and Sunstein; 2) to suggest several non-behavioural aspects of human agency and action which extend the original concept of nudging and should be accounted for by policy-makers in their design of nudge-like behavioural interventions.

Findings: It is claimed that mere inclusion of cognitive biases and irrationalities in the behavioural approach to policy does not sufficiently extend the artificial concept of the rational agent; in particular this narrow understanding of human failure misses important aspects of the rich concept of well-being.

Implications: The use of nudges requires a comprehensive knowledge of the application context. In underspecified decision contexts, choice architects need to apply more care and critical reflection in order to prevent unintended or harmful consequences of nudging.

Contribution: It is rare for pragmatically oriented public policy research to focus on the philosophical concepts that inform its theory and practice. This paper is a philosophical reflection on some key elements inherent in nudging. It helps better to understand the ambiguous design, potential and limitations of nudge policy.

Article classification: theoretical/conceptual.

Keywords: nudges, behavioural economics, behavioural public policy, well-being, human choice, practical reason

JEL classification: B4, B5, D6, D9, H1, H8, I3

Introduction

Behavioural sciences, not least behavioural economics, have inspired policy-makers with a new approach to improve social welfare. The new approach suggests that policies can be designed to minimise the influence of various distorting – mostly cognitive – factors on people's choices, so that they make better decisions which increase their welfare. This approach has been popularised

[1] The views expressed in this paper are the views of the author and do not necessarily reflect the views of the Polish Economic Institute. I thank two anonymous referees for their constructive reviews and valuable suggestions.

Agnieszka Wincewicz-Price, Polish Economic Institute, Behavioural Economics Unit, Al. Jerozolimskie 87, 02-001 Warsaw; agnieszka.wincewicz@gmail.com; ORCID: 0000-0003-3336-1753.

by a bestselling 2008 book *Nudge: Improving Decisions about Health, Wealth, and Happiness* by behavioural economist Richard Thaler and legal philosopher Cass Sunstein. The book has provided an important argument for third-party intervention in the preferences and choice decisions of citizens and given rise to a rich academic and professional debate about the legitimacy and usefulness of nudges and other forms of non-coercive use of behavioural tools in policy.

Insights from the original nudge theory and its abounding later developments, broadly conceived of as behavioural policy tools, promise to enable policy-makers to improve the design of public policies. In essence, this is possible because they are said to provide a deeper understanding of human behaviour and how it can be influenced (Bubb and Pildes, 2014; European Commission, 2016). Whereas traditional policy instruments are based on the assumption that people make rational decisions, behavioural policy insights – informed by a broad array of findings from behavioural economics, psychology and neuroscience – rely on the understanding that most of the decisions people make are not rational, but subject to many different biases and heuristics. This creates new opportunities for influencing people's behaviour and, therefore, it is said to also increase effective decision-making. Not surprisingly then, many policy-makers have recently embraced this new approach internationally, with several forms of what are known as nudge units now featuring in national and international policy-making.

The scope of nudge-like solutions to public policy problems is very broad and the list of new forms and applications of nudges is constantly growing.[2] Some of these new policy tools are introduced to supplement conventional policy measures (e.g. SMS reminders about upcoming appointments at a doctor or behaviourally designed letters prompting payment of overdue tax), while others have started to replace ineffective conventional regulations which are often (mis)informed by the assumptions of rational behaviour (e.g. a pre-defined default option instead of a voluntary contribution to pension schemes). Nudges seem especially attractive in the context of pressing societal and economic problems, such as obesity or insufficient savings, which are encouraging governments to constantly rethink the ways in which to address these issues. While in general the idea of the state attempting to manipulate preferences via its institutions appears highly dubious, the discovery that people could, without coercion, be 'nudged' into certain beneficial behaviours has been welcome with enthusiasm.[3]

In what follows, the claim that the behaviourally-informed approach to policy is based on an accurate image of the human is challenged. It will be shown that insights from behavioural sciences and behavioural economics, in particular,[4] only partially reform the abstract image of the rational decision maker and that accounting for merely – or primarily – cognitive biases and irrationalities might not suffice as a reliable basis for policy making. It will be argued that the new perspective does not pay enough attention to some important aspects of human agency which are not captured or sufficiently accounted for by behavioural sciences but are thoroughly discussed by philosophers and social scientists. Since these neglected qualities are

[2] The extensive research literature on the subject together with elaborate policy reports and guidelines produced by the forerunners of the new approach show a wide array of policy dilemmas and challenges which are increasingly tackled by behavioural tools. The list of important publications on various aspects of behavioural policy in general and specific elements of nudging in particular is too long to mention here. Examples are included in the bibliography at the end.

[3] The growing scope of various attempts undertaken to this effect can be seen in annual and periodical publications of the British Behavioural Insights Team, the OECD, the World Bank and an increasing number of specialist journals dedicated to behavioural policy analysis.

[4] One ought to be mindful of the common mislabelling of applied behavioural sciences as behavioural economics (cf. Introduction in Shafir, 2013). The reason for putting a particular emphasis on behavioural economics in the argument of this paper is its concept of rationality which – as it is shown below – the nudge philosophy has confused, and not without consequences.

important factors with regard to human welfare, ignoring them in the behavioural approach to policy making risks bringing about unintended and adverse effects on individual and social well-being. "Choice architects" responsible for implementing nudges and similar tools as an innovative governance practice need to be mindful of this danger so as to avoid long-term financial and social costs of their innovations. This paper thus qualifies the premature enthusiasm for this new type of government intervention. While it relies on a richer model of human psychology than traditional policy tools, its lasting success requires more critical reflection and non-behavioural insights.

A number of typologies of behavioural interventions in public policy have been proposed since the publication of *Nudge*, and the eponymous nudge itself is now widely considered to be merely one among several other possible forms of behavioural policy. It is, however, the original concept of nudging that is put under special scrutiny here to show how the economic vision of agency which takes on a new role in this approach – one of a normative standard – narrows the applicability of nudges and of analogous behavioural tools in policy.

Some of the arguments brought up here have been discussed more or less broadly by other authors. The controversies regarding nudging in particular have received an exceptionally rich treatment in the literature.[5] What is distinctive in the approach of this paper is the focusing of these various sources of critical reflection in the perspective of human practical reasoning and the recognition of the undetermined character of many of human goals.

The argument advanced in this paper necessarily crosses the boundaries of several disciplines. It tackles the important problem of conceptualisation of the human in economic and political theory with a mix of tools borrowed from economic methodology as well as political and moral philosophy. Such a multi-disciplinary approach seems necessary, given the narrow confines in which modern social sciences (including its behavioural branch) picture human choice and action. Selected explanatory limitations of this narrow framework are identified and important qualifications which should be considered in its application in behaviourally informed policy are suggested. Simplistic moral philosophy of well-being is given special attention in questions of policy objectives which are ethical in nature and need to be considered in a broader context than that of instrumental means-ends reasoning. Thus, reintegration of the economic image of the human, which has arguably been retained as a normative standard in behavioural policy of nudging, into the philosophical and moral discussion about human actions and human goals has been the proposed method of analysis.

The structure of this paper is as follows. Section 1 reviews the definitional characteristics of nudges. Section 2 discusses normative and epistemic foundations on which the original nudge theory rests. Section 3 challenges the limitations of the image of human agency painted by the authors of nudging by contrasting it with some constitutive aspects of human personhood. Section 4 provides cursory policy recommendations. The final section concludes this paper.

What is nudging?

For decades public policy has relied on the economic assumptions about human behaviour. It is not surprising, then, that refutation of these assumptions by empirical findings of behavioural sciences and development of behavioural economics in particular have also affected the perception of human decision making and choice behaviour in public policy. Contrary to the assumptions of the rational choice theory, behavioural research has shown that individuals often do not come to decision problems with pre-existing preferences. They rather form those only when confronted with particular problems, and they are sensitive to

[5] For example, Grüne-Yanoff (2012), Rebonato (2012), Sugden (2008) and White (2013).

details of 'framing'.[6] Thus, the authors of the nudge strategy claim that the decision-making situation can be designed to improve choices, so that they are closer to what the individual would choose in a situation free from obstacles. They have concluded that the findings of behavioural science justify policies which "nudge" individuals towards those choices that are in their best interests.

Although the image of nudging does a lot of work, the concept itself, originally, was not clearly defined.[7] Thaler and Sunstein define nudges mainly by example. The most unified account of the concept they propose describes nudge as "any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their (…) incentives" (2008, p. 6). In a later formulation Sunstein defines nudges as "initiatives that maintain freedom of choice while also **steering** people's decisions in the **right** direction (**as judged by people themselves**)" (2014, p. 17; emphasis added). That is vague in all the right ways so as to remain open for policy uses.

Thaler and Sunstein demonstrate how policy-makers can assume the role of choice architects and make major improvements to the lives of others by designing "user-friendly environments". Examples of nudge policies range from simple techniques, such as serving drinks in smaller glasses in order to reduce unhealthy consumption, designating sections of supermarket trolleys for fruit and vegetables or redrawing lines on roads to prevent speeding, to requirements that household energy bills contain comparative consumption information (e.g. in period X you have consumed n% more energy than your neighbours) and default enrolment in pension plans. In cafeterias a clever positioning of food – with the less healthy choices placed further away – could make it less likely that individuals choose unhealthy food, say in a sudden moment of weak will. Some nudges work because they inform people, other nudges work because they make certain choices easier, still others work because of the power of inertia or procrastination. By presenting information in particular ways, the state can nudge people towards being more sensitive to salient aspects of a situation. Such policies promise to reduce our exposure to misinformation or offer helpful suggestions of ways to achieve our goals. They do not merely simplify technicalities of a given decision process, but also streamline it so that the beneficial goal is accomplished more efficiently, without much ado.

To count as a **mere** nudge, as opposed to coercion, the intervention must be easy to avoid and avoiding it must not incur the chooser any serious costs. The set of available options should remain "essentially unchanged". The possibility for a person to make his/her own decision must remain. Choice architecture should be primarily intended to facilitate an individual's pursuit of his/her own goals. So, a subsidy is not a nudge, a tax is not a nudge, a fine or a jail sentence is not a nudge.

What is particularly important in the context of this paper, the authors of nudging claim to be offering policy ideas that are "informed by a more accurate conception of choice, one that reflects a better understanding of human behaviour and its wellsprings" (Sunstein 2000, p. 13). They ground their fresh understanding of human behaviour in the psychologically based assumptions about human decisions and choice behaviour. These assumptions are claimed to be more accurate, since they are backed by sound results of scientific experiments. What they hope to gain through this approach is improvement of "law's ability to move society toward desired outcomes" (Sunstein 2000, p. 38).

[6] Kahneman and Tversky (1979)

[7] Many attempts have been made to narrow down and specify the original concept (e.g. Hansen (2016), Hausman (2018)). The analysis of this paper focuses on Thaler and Sunstein's definition, since theirs best reveals its roots in standard economic theory and its account of rationality. It is also this concept which sparked the subsequent debate on behavioural policy. Its critical assessment can thus be useful for the analysis of modified concepts of nudging and other behavioural tools.

Thaler and Sunstein's concept of nudging is but one account of how to use behavioural insights in policy, as well as how to understand those uses normatively. Theirs, however, is the first and best-known policy approach of its kind, and the most widely discussed in current debates. Many of its distinct features have been pointed out and elaborated on. Various qualifications and typologies have been put forward in the debate about legitimacy of behavioural policy tools, not least nudges. The following sections present one more critical approach which contributes, so far unexplored, evaluation criteria for behavioural policy experts and practitioners. What has not been dealt with is a theoretical issue which has important practical implications for the policy of nudging and related behavioural tools, namely the concept's reliance on a confused idea of rationality. I take the original account of nudging as the case-in-point illustrating why philosophical complexities of human choice may deserve more attention in policy application of behavioural findings. Before explaining why these findings should not uncritically increase our enthusiasm for government intervention, I will offer a more in-depth analysis of the normative and epistemic underpinnings of the nudge approach.

The normative and epistemic foundations of nudge theory

Nudging is often referred to as soft or libertarian paternalism, for it claims to draw on the liberal vision of mature and enlightened individuals who are free to act in accordance with their interests. They make their own decisions in accordance with those interests and the available alternatives. They do all that autonomously. Followers of the liberal tradition, according to which people themselves know best what is good or bad for them and even a democratically legitimised government does not have the right to judge their convictions (Kirchgässner 2017), often reject paternalism. They are also sceptical about its softened version manifest in the policy of nudging, for by harnessing the modified vision of a mistake-prone individual, nudge advocates seem to be undermining this traditional liberal vision. Advocates of libertarian paternalism claim that once preferences are recognized to be context-dependent and tainted with "behavioural anomalies", the notion of individual sovereignty appears to be not well-defined (Brennan and Lomasky, 1983). Research shows that it is not uncommon for people to mispredict or overlook what is good for them. Human proclivity to choose smaller immediate rewards over larger delayed gratification (a phenomenon known as hyperbolic discounting) is a case in point.[8] This recognition has become the foundation of Thaler and Sunstein's innovative policy of nudging, as they say:

> … we emphasize the possibility that in some cases individuals make inferior choices, choices that they would change if they had complete information, unlimited cognitive abilities, and no lack of willpower (Thaler and Sunstein, 2003, p. 175).

While people no doubt often do make choices they later regret, the belief that they would change those choices in an ideal world of full information, perfect knowledge and unfailing character is not very well founded. In fact, it looks a lot like the assumptions made about choice in the standard neoclassical approach to economic agency which behaviouralists famously question. Thaler and Sunstein recognise that on the behavioural level people are not like the theoretical "homo economicus", but they nonetheless seem to propose this ideal as a normative benchmark for the making of superior choices.

The apparent reliance on homo economicus is even more visible in one of Sunstein's articles (2012). He quotes Rebonato's sceptical characterisation of libertarian paternalism:

> "Libertarian paternalism is the set of interventions aimed at overcoming the unavoidable biases and decisional inadequacies of an individual by exploiting them in such a way as to influence her

[8] See, for example, Ainslie and Monterosso (2003) and Scharff (2009).

> decisions (in an easily reversible manner) towards choices that she herself would make if she had at her disposal unlimited time and information, and the analytic abilities of a rational decision maker (more precisely, of Homo Economicus)." (Rebonato, 2012, quoted in Sunstein, 2013, p. 1860).

Interestingly, Sunstein does not argue with this description of his theory. But this makes the theory all the more debatable, because it implies that a particular – idealised – model of human decision is granted an epistemic priority that is far from being self-evident. It seems, therefore, that libertarian paternalists take behavioural economics seriously in their description of human behaviour, but they otherwise ignore it by normatively adhering to the overly demanding rationality principles endorsed in standard economics.

In what follows I argue that the normative reliance on the homo economicus view of choice is a problematic element in Thaler and Sunstein's theory of nudging and likely the source of much criticism professed against it. It downplays the role of human agency in discovering and learning one's true ends over time. It also fails to appreciate the value of that very process for humans. It overlooks the not uncommon possibility that the chooser's personal experience of his/her choice situations might over time enable him/her to develop insights into what is worthwhile and what is not. In short, this behavioural approach can have an adverse effect on a person's well-being, broadly understood, especially if it aims to affect the person's good merely by influencing causes beyond his/her control without engaging his/her conscious deliberation and action.

What is missing in the behaviourally accurate picture of the human? Three problems behind the theoretical foundations of nudging

The implicit normative foundation of nudging defined by an idealised, static picture of homo economicus can be challenged on (at least) three important notions underlying the theory of nudging: mistaken or inferior choice; good or welfare-increasing choice; and the optimistic status of nudge which can only increase or, at worst, be neutral, but never impair one's welfare. These notions acquire different meaning when seen from the position of a choosing person rather than an artificially constructed rational agent.

A choosing person is a practical reasoner who usually lacks full knowledge of his/her chosen end(s). He/she chooses things under some aspect of the good, and without knowing what those choices really entail until after enacting the choice in his/her life. I chose to marry, not realising all it entailed. My choice might not have been better (or even possible) if I had had full knowledge of the entailment. A model of choice in partial knowledge of what is chosen, and in hope of becoming or remaining a certain kind of person, is a better model of how ordinary moral persons choose, than that proposed either by rational choice theory or by the nudgers. Whereas the former assumes full knowledge of one's preferences, the latter accepts this is not always the case, but focuses merely on psychological factors which make the perfect knowledge impossible. The following three subsections provide examples of alternative non-behavioural explanation of why people tend to make "irrational choices" and show why these cases might not be good candidates for nudging.

Preventing mistakes

Many of Thaler and Sunstein's examples of nudges suggest that the problem they want to address consists of bad choices resulting in decreased well-being of the chooser. Their famous cafeteria example targets the mistake of eating too much unhealthy food. The other often quoted example of automatic enrolment in a pension scheme addresses the fault of not saving enough money for retirement. Other mistaken choices they list can result from unrealistic optimism (e.g. in starting a business), status quo bias,

loss aversion, and other cognitive or volitional factors of choice which are hard to control. More generally, bad decisions are understood here as decisions people "would not have made if they had paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control" (2008, p. 5).

A serious limitation of Thaler and Sunstein's approach in this regard is that it does not consider the possibility that what might appear to be poor or mistaken decisions are not necessarily or merely results of cognitive or psychological biases. There are many other factors which can lead to mistakes in human judgment and decision-making. While – as empirical experimentation demonstrates[9] – quite often mistaken choices are indeed the result of cognitive biases and heuristics, decision-making processes (for there are many) are influenced by a wide array of motivations and considerations which are neglected in the framework postulated by nudge advocates. Poor choices could, for instance, reflect one's incomplete understanding of what one should value or which of one's values should be pursued. In short, the "mistakes" might result from bad reasoning or poor judgment, which is not easily reversible by the application of a nudge.

Curiously, the architects of nudging do not conceive mistakes as deviations from some objective notion of the good. Instead, they understand mistakes as decisions that people themselves regret upon reflection. Nudging is therefore supposed to help people make choices they will not regret. Its apparent role is to correct or prevent people's mistakes and thus help them to achieve their "true" or more "authentic" ends.

This approach suggests a fairly strong epistemic and normative conclusion, namely that the phenomena observed by behavioural economists and described as deviations, bias, or anomalies can and should be counteracted. When the experiments in psychology and social sciences on which these ideas are based are interpreted from the perspective of standard rationality theory – as is arguably done by the authors of *Nudge* – they are thought to reveal mistakes people would not make if they were like homo economicus. They accordingly show the psychological heuristics and biases people exhibit as incidental and correctable rather than as fundamental to their nature, or even as essential to how people make choices. But if we take the human tendency to value the present more strongly than the future, for example, it is far from obvious that this is an error in decision-making, rather a feature of human nature. Nudges that are supposed to correct for this "mistake" are based on a somewhat arbitrary decision to value long-term preferences more highly than short-term preferences. Since it can be predicted that a person might regret his/her choice in the future, the "mistaken" choice needs to be prevented.

It is, however, not always the case that increasing someone's well-being in the future at the expense of today is the right thing to do. As Sunstein himself admits, "there can be a thin line between a self-control problem and a legitimate focus on short-term pleasure" and, continuing, that "no choice architect should engage in a program of nudging that disregards the importance of short-term pleasures, or pleasures in general, which are of course crucial parts of good lives" (2016, p. 47).[10]

An overtly simplistic diagnosis of a mistaken choice entails one other important difficulty for the original nudge theory. From their understanding of mistaken decisions Thaler and Sunstein infer that the value of nudges can be found in that nudges allow individuals to overcome their various biases and blunders that affect their everyday behaviour, and help them resist temptations. It is hard to see, however, how a nudge which relies on an automatic and essentially unconscious psychological mechanism (e.g. increasing the salience of healthy food at a cafeteria by exposing it more than unhealthy sweets) has anything to do with a demanding and effortful conscious process of overcoming one's weakness. The latter requires an act of will and

[9] AIKhars et al. (2019)

[10] On the importance of purpose and pleasure see Dolan (2014) and Gilbert (2006).

usually a long-term endeavour, sometimes thwarted by failures. Nudging which harnesses cognitive biases and gets around human consciousness has little to do with self-aware attempts to correct one's bad habits.

Lastly and related to the next section, the approach does not account for the possibility that people form their values and ideas of the good in part *through learning* about their biases and weaknesses. Their real or apparent mistakes might, in fact, help them understand what is good for them by providing a formative experience. Sometimes people only arrive at their judgments of well-being and learn to appreciate those in their efforts to overcome difficulties, resist temptations, juggle priorities, etc. A reluctant long-term smoker who makes repeated resolutions to give up his/her addiction is a case in point. An obvious consideration in this regard is that the learning experience may be quite costly for the person concerned and even for the whole society, especially when the mistaken decision is irreversible or the long-term consequences are realised too late. Two commonly discussed examples are poor health decisions and insufficient retirement savings. A nudge put in place to help avoid severe consequences in these and similar circumstances is therefore often seen as a boon.[11]

While this is a complex issue which would benefit from separate treatment, it seems relatively evident that preventing people's mistakes can be less controversial in cases in which it is obvious or generally agreed what is good for people (crossing the street properly would be an obvious example). That the "mistake" is correctly identified and understood seems therefore crucial for the legitimacy and effectiveness of nudges, so that they can indeed help people achieve ends which are good for them. In other cases, where it is not obvious what a person's good consists of, nudges become problematic, since there is a danger that they arbitrarily impose ideas of the good or make discovery of the good for oneself impossible.[12]

Nudging for the good

Nudging is intended to improve welfare of the persons concerned. In order to succeed, choice architects need to know which choice reflects the person's well-being. Is it better for the person if he/she spends or if he/she saves some part of his/her income (and what part in each case)? Is it better for him/her if he/she buys insurance or if he/she does not? Is it better for him/her if he/she sticks to his/her diet or enjoys a family meal, indulging in food he/she would otherwise deny himself/herself? It is hard to see how these questions could be answered merely by simulating what the person in question would choose if he/she had been free from temptation and reasoning imperfections.

This is, however, how Sunstein and Thaler propose to resolve such problems when the authors say that people are not acting in their own best interests if their decisions are ones "they would change if they had complete information, unlimited cognitive abilities, and no lack of willpower". It is difficult to see how such an idealised criterion can be treated as empirical. For example, it is not obvious how we can determine what complete information, unlimited cognition, or complete willpower entail without making normative judgements in relation to specific circumstances. This counterfactual reasoning advocated in *Nudge* has been rightly criticised as (yet another) a "view from nowhere".[13] The argument relies on a purely hypothetical basis: if individuals were fully rational and were choosing according to their informed preferences, they would do *X*. Of course, one may wonder what is it to be "fully rational", and to have "informed preferences". Similarly, how can one know (or even

[11] For this reason default opt-in in pension schemes is gaining popularity around the world (including Poland). An argument to the contrary has been put forward by van Aaken (2016, p. 95).

[12] On the state's responsibility for nudging see Hansen (2016).

[13] See, for example, Sugden (2008), Hédoin (2015).

make sense of) what choice someone endowed with "unlimited cognitive abilities" will make? The "view from nowhere" is a perspective typical of homo economicus, which has no history, no future, no commitments, and no context-specific considerations. This is not the perspective that the practically reasoning moral person assumes in his/her decisions.

In asking after what a person would choose if he/she experienced no problems of cognition or self-control, nudge advocates imply that there is a true choice one would make if one had appropriate conditions for choice. A nudge is in place because it ostensibly creates the appropriate conditions. This perspective entirely neglects the fact that as long as the person has not fully formed the idea of the good he/she wants to pursue, it is hard to say what decision framework can help him/her achieve the end. It does not seem possible to design a choice framework to help realise an unknown good as an end. Therefore, constructing a "proper" decision framework cannot be seen as a sufficient factor determining the proper choice for that person. It will not be sufficient for the person to make a choice he/she will not regret.

This conceptual basis of nudging seems to share the more profound problem entailed in liberalism. That is the assumption that one can be allowed to pursue one's idea of the good within an ethically neutral institutional framework, which itself respects each individual's judgment and does not impose any value criteria on him/her. In a libertarian spirit (here meaning just hyper-liberal and morally libertarian, rather than anti-government) Sunstein and Thaler recommend an unassuming policy tool. It is not to contain a doctrine about what constitutes people's welfare.[14] It merely suggests ways to improve people's own decisions regarding what is good for them.

In this, they want to be able to say that their proposals will steer each individual in the direction that he/she would have chosen for himself/herself under ideal circumstances. In their counterfactual reasoning the authors seem to rely on the idea that anything increasing the extent to which people make choices that they regret less than they otherwise would is respectful of their internal sense of what is important. However, it is one thing to determine what people's preferences would be if they were free of rational foibles, and it is a different thing to determine what is good for people. The approach of *Nudge* seems to suggest that one's decision situation can be modified to help one achieve an outcome that one finds agreeable, without the need to define that outcome in advance. The decision framework itself is meant to be neutral ethically. It is as if one's ends could exist or be formed in isolation from one's decision situation. But in order to determine what is good for people "some grounds independent of their distorted preferences" (Hausman 2012, p. 101) are needed. It is hard to imagine how policy advice could be formulated without a concept of well-being or welfare. What is deliberately undertaken must have a goal. Nudges are no exception. If policy makers can determine what is truly good for individuals and society, then they can devise policies that will lead people to make better choices. A substantive theory about which choices actually make individuals better off, is ultimately needed. When a strong position cannot be worked out, all possible trade-offs should be recognised and weighed against one another. Such might be the decision about using default options in pension schemes and organ donations.

In avoiding taking a position on what is good for persons, nudging offers a procedural approach to welfare. What is not captured in this ostensibly neutral approach is the possibility that one's **formation of ends and ideas of the good itself is**

[14] Sunstein and Thaler avoid defining welfare, claiming that they "are not attempting to say anything controversial about welfare, or to take sides in reasonable disputes about how to understand that term" (Sunstein and Thaler 2003, p. 1163). The abstract character of this concept has been criticised by Sobel (2016, p. 51), who argues that the "the notion of a fully informed self is a chimera", because the great variety of possibilities, choices, and lives an agent might lead are not available to a single consciousness.

an important element of choice and important factor of welfare. The idea of nudging does not account for this possibility. It locates well-being in given end states, outcomes of choices which one would make, if one were a fully-knowledgeable, rational agent. The nature of this outcome is seen as mere utility, or as reducible to utility. As moral persons, however, we do not act to maximise some well-defined utility. Rather, we often act to realise some under-specified value, often in order to learn what that value is and what it means for us. Action is not always or merely outcome-oriented. It could (also) be an expression of one's values (Hargreaves Heap, 1989). Thus the end outcome is not the only criterion defining the right and unmistaken action.

Skipping levels

The third important assumption which is implicit in nudging holds that human limitations are best corrected by a third party through a properly designed nudge. This is so, apparently, because people, in their fallibility and susceptibility to bias, cannot be trusted to correct themselves. The choice architect responsible for the design of choice environment is seen as an impartially benevolent spectator, who helps people to achieve their own ends without influencing them normatively. If therefore, he/she nudges a person into ways of acting which he/she finds in the person's best interest, and it turns out to be in that person's best interest, we should be inclined to welcome nudging as a true improvement to decision making. There are at least two reasons making this optimism unjustified.

Firstly, as argued in the previous section, it is rarely obvious what one's true interests might be. People not only are uncertain whether their actions serve their goals, but in some important circumstances they are also uncertain of the goal itself. When it is not clear what the goal is, the choice architect needs to make a value judgment that is not necessarily in line with what the person's judgment would be if performed in an ideal environment. This creates a danger that the choice architect will be relying on his/her own values and beliefs instead of the chooser's. Since a nudge cannot be neutral among available alternatives, in practice it steers people toward choices that the architect expects are going to be seen as beneficial by the chooser himself/herself. Since there is no 'neutral' frame, any way of presenting a situation of choice will necessarily make an option more salient than another one (by defining, say, a default option, a first option in a list of alternatives, a set of possible options, etc.).

But even if nudges are designed to further ends which choosers approve of, there is a second, more profound, problem with this strategy for welfare-enhancing. Due to its over-reliance on behaviourism, nudging takes people to be (mostly) passive, plastic, malleable entities. It is understood in this approach that humans respond to external modifications in their environment, but underappreciated that they can also influence it themselves. Given that the capacity to form one's ends is an element in a person's well-being, when a person's value formation is replaced by a third party's valuation (which may happen in the process of nudging), he/she can be justified in feeling deprived of his/her inherent capacity to determine his/her own ends based on his/her own values.

As rational agents, in Christine Korsgaard's words we "are faced with the task of making something of [ourselves]" (2009, xii). With the passive, atemporal homo economicus as its implicit normative standard, nudging neither seems interested in nor capable of helping people develop capacities or abilities for this kind of transformative choice (Wartenberg, 1990).

Nudges can alter the behaviour of individuals to coincide with those who accept certain norms, but it cannot provide **the reasons** necessary to alter people's behaviour in the long run. Thus, on their own, successful nudges merely lead people to act **as if** they had a well-established notion of the good they pursue, without attempting to engage with the individual at that deeper level where he/she might be, actually only, trying to establish what end to pursue.

Insofar as nudges are designed to select the "prudent" or "best" alternative without having to invest the otherwise necessary careful deliberation, they do not engage the chooser's key capacities and potential. One could go as far as to claim that nudging could stand in the way of a person making efforts entailed in that person's self-formation. This is especially true if it prevents the person from acquiring the kind of experience in which he/she gains knowledge about what is and is not good for him/her.[15] Imagine a video game in which a player was allowed to win a game **by** skipping a few, perhaps the most difficult, levels. The general concept of nudging does not seem to appreciate this heuristic aspect of personal identity. In their original publication the authors of nudging fail to consider human agency and well-being from a broad enough perspective. They do not account for the fact that welfare can be found in the capacity for self-formation, which homo economicus does not have. In short, thus understood, behaviourally informed attempts to influence and steer behaviour can have an adverse effect on a person's well-being. That is to say, if it aims to affect the person's good merely by influencing causes beyond his/her control without engaging his/her conscious deliberation and action.

More recently, Sunstein himself took up the issue of the threats that nudging might pose to agency (Sunstein 2017). He distinguishes non-educative nudges, which rely on unreflective behaviour and harness cognitive biases (default options, framing, use of emotionally charged graphics), from educative ones, which are devised to increase people's reflective capacities and help them make more informed decisions (disclosure requirements, warnings, reminders). The latter type of intervention arguably respects and perhaps even enhances human agency.[16] Even so, Sunstein still argues in favour of non-educative nudges, mostly on the grounds of their value in simplifying our increasingly complex life. Thus, for instance, nudges might be considered to be less of a threat to agency overall when it comes to basic retirement savings or severe health risks. In a similar vein, Valdman (2010) suggests that nudging could be seen as the product of an act of voluntary partial "outsourcing of self-government" to some external body. He offers a view that there are certain domains where it is plausible to presume that a large majority of citizens benefits from contextual support – from partially "outsourcing their agency" – in order to minimize the risk of severe distress later in life. Similarly, Conly (2013) seems to think that most of us would gladly outsource choices associated with giving up smoking or examining nutritional content which we do not enjoy making. On this reasoning, the public body would "ease us of the responsibility of doing what we would rather not do on our own". Far from finding such restriction of autonomy objectionable, we might welcome such laws and policies as "unburdening" us with regard to doing the things in life we would rather do. It should be noted, however, that the examples offered by Valdman and Conly, respectively, illustrate an important contrast. For there seems to be a significant difference between outsourcing one's agency in order to avoid severe distress in the future and letting a third party decide for one for the sake of being unburdened of difficult choice-making. Conley's line of argument seems to value convenience in a way that is not as explicit in Valdman's analysis. That that convenience may become dangerously "excessive" (Korsgaard 2009) is not considered in Sunstein's analysis of non-educative nudges.

[15] Goodwin (2012) argues that alongside these traditional restrictions on freedom, the process of self-realisation and overcoming internal obstacles to action (e.g. addictions, phobias, aversion and prejudices) may also be important, I would say, as part of positive freedom.

[16] Some authors place non-educative nudges in an altogether different category of so called boosts (e.g. Yanoff & Hertwig (2016)). The educational value of non-coercive policy interventions as opposed to the manipulative character of nudging is also endorsed by John et al. (2013). Distinction that is closest to the argument developed in this paper has been made by Niker (2017).

Cursory lessons for policy

The above discussion exhibits some key assumptions which are implicit in the notion of nudging. It shows that insofar as the design of nudges relies on the conceptual framework which is analogous to that in which the abstract, atemporal homo economicus operates, it is not necessarily fit for furthering well-being broadly understood. Nudging, at least in Thaler and Sunstein's construction, is informed by the counterfactual reasoning of the rational economic man and thus relies on a partially misjudged normative ideal. Its objective is to prevent people from making mistakes they will regret without due consideration for the larger context of mistake-making and its meaning for the individual. What follows from this is that uncritical support for nudging fails to account for a realistic picture of human agency. Trying to influence people's choice behaviour rather than their reasons for choosing neglects that real persons, unlike the rational homo economicus, are often not able to specify or define what they truly want, at least not in an(y given) instant. What they truly want often depends on values which they form over time, and not merely on a choice framework. And those values are valued for reasons unrelated to achieving some specific goal. In other words, the original notion of nudging – and related behavioural policy instruments – neglects the human capacity for practical reasoning and focuses merely on the instrumental aspects of decision making. It thus also fails to account for the broader concept of human welfare, which is in part constituted by one's ability to exercise practical reason, without which rational agency would be inconceivable.

The foregoing analysis should not be read as uncompromising criticism of the entire approach but rather as a call for a deeper reflection on the conceptual construction and the use of nudges and their possible extensions. It also suggests a framework of rarely considered evaluation criteria which should be useful for the assessment of the legitimacy of a particular nudge-like tool. For the nature of nudges is such that they need to be designed and applied in the context of a particular case, with an understanding of all relevant institutional and behavioural particularities. Because nudges vary in form, mechanism and complexity, and their long-term consequences are not always obvious or agreed upon, it is imperative for policy-makers to have a nuanced grasp of how they are meant to work and what they can and cannot be expected to achieve. It is also important, if challenging, to be able to fit nudges into the larger picture of a given policy aside other tools which may or may not be more applicable in a given case. With all its promising qualities, nudging should be seen as an addition to the existing policy practice and not as replacement of traditional regulatory tools. And thus it is the bigger picture of a given policy which should serve as an evaluation framework for most instances of nudging.

On the practical level it is crucial to remember that nudges might help mitigate adverse consequences of myopic or uninformed decisions only if the perceived failures of cognition or attention are indeed the root cause of the problem in question. When the problem is caused by poor motivation, lack of education or infrastructure, behavioural tools should not be employed as a primary remedy. All in all, behavioural insights need to be interpreted with caution so that turning away from the ideal of homo economicus as major reference point for policy is not replaced by an equally distorted perception of people as hyper-irrational creatures. The following theoretical scenario illustrates the danger of excessive focus on people's perceived behavioural biases in policy making.

Take the example of entrepreneurs. From behavioural point of view, it would not be difficult to highlight possible biases inherent in entrepreneurial activity: overestimation of abilities, overconfidence, the desire for self-direction and self-determination which may lead to unpredictable consequences (Astebro et al., 2014). If we accepted the position of nudge advocates who want to use nudges to debias people's excessive optimism, nudges might need to be introduced to prevent entrepreneurs

from initiating risky ventures. Their creative ideas might be viewed as the product of biases that are reported to be destructive to the individual and that lead to behaviours that proliferate market choices or disorder decision-making environments for the rest of society. In the Schumpeterian view, for example, the entrepreneur disrupts the prevailing equilibrium by introducing some new innovation that completely alters existing patterns of supply and demand. Overtly cautious nudge advocates might also say that entrepreneurs are constantly disrupting the stasis of the marketplace, adding and replacing products, services, businesses, jobs, etc., from the market, which simply contributes to the proliferation of choices available to us, thus making new and changing cognitive demands on us. Nudges designed to prevent such "unwelcome" consequences would surely curtail the ability of entrepreneurial individuals to act upon their initiatives. This could be problematic for the otherwise entrepreneurial society and for the individual entrepreneurs.

While the above might not seem like a plausible policy scenario, it does illustrate some conceptual deficiencies in the theory of nudging by taking it to its logical extreme. Yet, an analogous case could be considered in which unintended consequences of business activity are prevented by a slightly different nudge, one that is rooted in a richer normative framework than that proposed by Thaler and Sunstein. A view of entrepreneurial decision making as neoclassical (instrumental) rationality has little to say about an important dimension of entrepreneurship, namely the uncertainty inherent in the development of new business ventures and openness to learning. From the policy perspective, the major problem with entrepreneurship is the fact that a large percentage of newly incorporated companies do not succeed and those which fail cannot therefore provide the benefits they promise to their owners, shareholders and the society. Mere tax incentives and better funding options might not be enough to rescue vulnerable enterprises. A simple nudge aimed at identifying risky cases of vulnerable companies and enable managing their problems at an early stage could be of more help. Here is how it could work.

In an annually filed tax statement, underneath the section in which the business owner states the amount of his/her profit/loss made in a given year, he/she is asked the following question: "Given your performance last year, would you like to take advantage of a free-of-charge consultation/mentoring session in which you can discuss your major ideas, needs and challenges and work out possible ways to improve your track record in the next 12 months?". Entrepreneurs with little or no success in the initial phase of their activity are often discouraged from continuing. Lack of persistence is one of the main reasons for start-up failure. It is conceivable that an optional meeting with a good mentor who can help identify areas of possible improvement – both technical, related to the product itself and personal, aimed at improving the entrepreneur's leadership skills – result in working out realistic milestones for the next 6–12 months and possibly help the business stay on the course longer than would otherwise be possible.

The above example points out an important difference between creating environments that have general characteristics and manipulating environments to create particular outcomes. The paternalism that (at least some of) the nudge advocates endorse is a paternalism that aims to direct individuals toward certain actions in aid of a determinate view of their well-being. This is much different than simply advocating general conditions in which individuals might flourish in the ways they see fit and according to their own decisions. No one would disagree with the critics that how we frame choices reveals our preferences, morals, priorities, and no one would disagree that framing choices in some general way is inescapable. But given that constant, it does not follow that we have no principled objections available to us to avoid endorsing an increasing number of policies that direct individuals to particular ends. The external "scaffolding" which Clark (1997) recommends in the form of institutions

that support practical reasoning and that focus our attention on better comprehension of our ends might be one that helps one along the way of an uncertain and perhaps risky undertaking (be it setting-up a new business venture, starting a career in X, changing jobs, etc.) rather than one which is designed to set us on a pre-defined track or prevent one from making mistakes. This way of support could be exercised explicitly and transparently, unlike automatic nudges, which do not allow for deliberation and are criticised as manipulative. Moreover, it would offer a more direct causal link between affecting individual well-being and contributing to overall social welfare which is not obvious in many "standard nudges".[17]

Concluding remarks

Critical reflection on the conceptual deficiencies inherent in the original theory of nudging reveals potential risks which could affect some but not all areas of its application. It cannot be denied that there are some areas in which the policy of nudging emerges as a convenient, attractive, and innovative regulatory tool designed to help public policy objectives be achieved more effectively and sometimes less expensively. Its helpful potential lies in the recognition that human cognitive biases are not merely an obstacle but also an opportunity for public regulation. It identifies and addresses important sources of many present-day social problems and endeavours to help "busy people trying to cope in a complex world in which they often do not have the time to think deeply about every choice they have to make" (Thaler & Sunstein, 2008, p. 95). Simplifying complex administrative procedures and forms as well as clever communication of professional jargon (financial, legal, technical, etc.) to non-expert clients are examples of areas where un-intrusive and non-manipulative nudging could be welcome and needed. In such cases, its overall intention to make people's lives easier and help them achieve their true interests seems uncontroversial, if optimistic. Whether reducing bias and preventing mistakes by defaulting people into a scheme they do not have sufficient knowledge of or framing their choice situation in a way that hinders deliberation should be treated as analogously beneficial is far from obvious.

The above considerations expound the increasingly recognised fact that there is nothing direct, straightforward and automatic in the use of nudging. Neither is nudging normatively neutral. That nudges can be adapted to a given situation is probably their greatest advantage. The difficulty remains in dissecting the important factors that matter in a given policy context, both functionally and normatively for unambiguous theoretical evaluation of the use of nudges is not possible. The experimental methods which are increasingly applied to serve this objective are a promising source of much needed data.

On a more fundamental level, the foregoing adds to the debate on the role of the state in increasing the well-being of individual citizens. Even in rare conditions of a general agreement that the state does in fact have a great role to play in this regard, the above analysis suggests that special caution is required in the design and implementation of specific nudges so that their proximate success in deterring someone from a harmful choice does not come at a long-term cost of hindering that person's decision abilities. For people might not always benefit from the type of security from mistakes and comfort of having their ends (co-) decided for them which some forms of nudging promise to provide.

The perspective taken in this paper suggests that the tool of nudging could be better used to facilitate people's formation of their goals instead of steering them towards pre-defined ends. Policy-makers could then shift their focus from the micro-level of identifying 'true' preferences to the meso-scale creation of conditions where such preferences stand a better chance of being

[17] Recent research shows that individually targeted nudges can seriously backfire on the collective level and result in decreasing social welfare instead of helping it (see, for example, Bolton, Dimant & Schmidt, 2018).

developed by the person herself. The short case study of entrepreneurship and the alternative form of nudge it proposes shows that there are unexplored ways in which this approach might be possible and useful.

In conclusion, it appears that the full potential of nudging as a policy instrument and means to strengthen public policies will remain underutilised and its long-term consequences far from understood, unless a more comprehensive picture of human choice is considered by nudge experts. As indicated in this paper, such broader perspective is needed for a more nuanced view of the workings of nudging as well as for its normative appraisal.

References

van Aaken, A., (2016). Judge the nudge: In search of the legal limits of Paternalistic Nudging in the EU, In: A. Alemanno & A.L. Sibony, *Nudge and the Law: A European Perspective*. Oxford: Hart Publishing.

AIKhars, M., Evangelopoulos, N., Pavur, R. & Kulkarn, S. (2019). Cognitive biases resulting from the representativeness heuristic in operations management: An experimental investigation. *Psychology Research and Behavior Management*, *12*: 263–276.

Ainslie, G. & Monterosso, J. (2003). Hyperbolic discounting as a factor in addiction: A critical analysis. In R. E. Vuchinich & N. Heather (Eds.), *Choice, Behavioural Economics and Addiction* (pp. 35–69). Amsterdam: Pergamon/Elsevier Science Inc.

Alemanno, A. & Sibony, A.L. (2016). *Nudge and the Law: A European Perspective*. Oxford: Hart Publishing.

Astebro, T., Herz, H., Nanda, R. & Weber, R. (2014). Seeking the roots of entrepreneurship: Insights from behavioral economics. *Journal of Economic Perspectives*, *28*(3): 49–70.

Bolton, G., Dimant, E. & Schmidt, U. (2018). When a nudge backfires: Using observation with social and economic incentives to promote pro-social behavior. *PPE Working Papers 0017, Philosophy, Politics and Economics*, University of Pennsylvania, https://ideas.repec.org/p/ppc/wpaper/0017.html

Brennan, G. & Lomasky, L. (1983). Institutional aspects of "merit goods" analysis. *FinanzArchiv/Public Finance Analysis*, *41*(2): 183–206.

Bubb, R. & Pildes, R.H. (2014). How behavioral economics trims its sails and why. *Harvard Law Review*, 127(6): 1593–1678, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2331000

Clark, A. (1997). Economic reason: The interplay of individual learning and external structure. In J. Drobak & J. Nye (Eds.), *The Frontiers of the New Institutional Economics* (pp. 269–290). Cambridge: Academic Press.

Conly, S. (2013). *Against Autonomy. Justifying Coercive Paternalism*. Cambridge: Cambridge University Press.

Dolan, P. (2014). *Happiness by Design*. New York: Avery Publishing Group.

European Commission (2016). *Behavioural Insights*, https://ec.europa.eu/jrc/en/research/crosscutting-activities/behavioural-insights

Gilbert, D. (2006). *Stumbling on Happiness*. New York: Alfred A. Knopf.

Goodwin, T. (2012). Why we should reject 'nudge'. *Politics*, *32*(2): 85–92.

Grüne-Yanoff, T. (2012) Old wine in new casks: Libertarian paternalism still violates liberal principles. *Social Choice and Welfare*, *38*(4): 635–645.

Halpern, D. (2015). *Inside the Nudge Unit: How Small Changes Can Make a Big Difference*. London: Random House.

Hansen, P. (2016) The definition of nudge and libertarian paternalism: Does the hand fit the glove? *European Journal of Risk Regulation*, *7*(1): 155–174.

Hargreaves Heap, S. P. (1989). *Rationality in Economics*. Oxford – New York: Basil Blackwell.

Hausman, D. M. (2012). *Preference, Value, Choice, and Welfare*. New York: Cambridge University Press.

Hausman, D. M. (2018). Nudging and other ways of steering choices. *Intereconomics*, *53*(1): 17–20.

Hédoin, C. (2015). From utilitarianism to paternalism: When behavioral economics meets moral philosophy. *Revue de philosophie économique*, *16*(2): 73–106.

John, P., Cotterill, S., Richardson, L., Moseley, A., Smith, G., Stoker, G. & Wales, C. (2013). *Nudge, Nudge, Think, Think: Using Experiments to Change Civic Behavior*. London – New York: Bloomsbury Academic.

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk, *Econometrica*, *47*(2): 263–292.

Kirchgässner, G. (2017). Soft paternalism, merit goods, and normative individualism. *European Journal of Law and Economics*, *43*(1): 125–152.

Korsgaard, C.M. (2009). *Self-constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Niker, F. (2017). Policy-led 'ecological' virtue-cultivation: Can we 'nudge' citizens towards developing virtues? In T. Harrison and D. Walker (Eds.), *Theory and Practice of Virtue Education*. London: Taylor & Francis.

Rebonato, R. (2012). *Taking Liberties: A Critical Examination of Libertarian Paternalism*. Basingstoke: Palgrave Macmillan.

Shafir, E. (2013). *The Behavioral Foundations of Public Policy*. Princeton, NJ: Princeton University Press.

Sharff, R. (2009). Obesity and hyperbolic discounting: Evidence and implications. *Journal of Consumer Policy*, *32*(1): 3–21.

Sobel, D. (2016). *From Valuing to Value: Towards a Defense of Subjectivism*. Oxford: Oxford University Press.

Sugden, R. (2008) Why incoherent preferences do not justify paternalism. *Constitutional Political Economy*, *19*(3): 226–248.

Sunstein, C. R. (Ed.) (2000). *Behavioural Law and Economics*. Cambridge: Cambridge University Press.

Sunstein, C. R. (2013). The Storrs lectures: Behavioral economics and paternalism. *Yale Law Journal*, *122*(7): 1826–1899.

Sunstein, C. R. (2014). *Why Nudge?: The Politics of Libertarian Paternalism*. New Haven: Yale University Press.

Sunstein, C. R. (2016). Fifty shades of manipulation. *Journal of Behavioral Marketing*, *1*(3–4), 213–244.

Sunstein, C. R. (2017). *Human Agency and Behavioral Economics. Nudging Fast and Slow*. New York: Palgrave Macmillan.

Sunstein, C. R. Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, *70*(4): 1159–1202.

Thaler, R. H., Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.

Valdman, M. (2010). Outsourcing self-government. *Ethics*, *120*(4): 761–790.

Wartenberg, T. E. (1990). *Forms of Power: From Domination to Transformation*. Philadelphia: Temple University Press.

White, M. (2013). *The Manipulation of Choice: Ethics and Libertarian Paternalism*. New York: Palgrave Macmillan.

Yanoff, T. G., Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory? *Minds and Machines*, *26*(1–2): 149–183.