

How autonomy is understood in discussions on the ethics of nudging

ANASTASIA VUGTS*

Chair group Strategic Communication, Wageningen University and Research, Wageningen, The Netherlands and

Chair group Philosophy, Wageningen University and Research, Wageningen, The Netherlands

MARIËTTE VAN DEN HOVEN

Department of Philosophy and Religious Studies, Utrecht University, Utrecht, The Netherlands

EMELY DE VET

Chair group Strategic Communication, Wageningen University and Research, Wageningen, The Netherlands

MARCEL VERWEIJ

Chair group Philosophy, Wageningen University and Research, Wageningen, The Netherlands

Abstract: Nudging is considered a promising approach for behavioural change. At the same time, nudging has raised ethical concerns, specifically in relation to the impact of nudges on autonomous choice. A complexity is that in this debate authors may appeal to different understandings or dimensions of autonomy. Clarifying the different conceptualisations of autonomy in ethical debates around nudging would help to advance our understanding of the ethics of nudging. A literature review of these considerations was conducted in order to identify and differentiate between the conceptualisations of autonomy. In 33 articles on the ethics of nudging, we identified 280 autonomy considerations, which we labelled with 790 unique autonomy codes and grouped under 61 unique super-codes. Finally, we formulated three general conceptualisations of autonomy. Freedom of choice refers to the availability of options and the environment in which individuals have to make choices. Agency involves an individual's capacity to deliberate and determine what to choose. Self-constitution relates to someone's identity and self-chosen goals. In the debate about the ethics of nudging, authors refer to different senses of autonomy. Clarifying these conceptualisations contributes to a better understanding of how nudges can undermine or, on the other hand, strengthen autonomy.

Submitted 24 October 2017; revised 16 January 2018; accepted
1 February 2018

* Correspondence to: Hollandseweg 1, 6706 KN, Wageningen, The Netherlands. Email: anastasia.vugts@wur.nl

Background

In 2008, Richard Thaler and Cass Sunstein proposed their theory of ‘nudge’ using previous work on bounded rationality as a basis for effective policies that aim to promote human well-being. Basically, they worked from the observation that human decision-making can be flawed and influenced by biases and heuristics, leading people to make decisions that are not in their own best interests. Human beings can, for example, be illogically optimistic (Sharot, 2011), which make them underestimate the chances of something bad happening to them (such as the chance of getting harmed in a car accident). Moreover, human beings easily overestimate the chances of positive events happening to them, such as winning a lottery. Human beings also tend to procrastinate, postponing their tasks at hand (such as studying for an exam) while distracting themselves with other activities (such as watching TV). Procrastination is considered a form of self-regulatory failure (Steel, 2007). Based on the dual process theory, Kahneman (2003, 2011) distinguishes heuristic, unconscious and fast or so-called ‘system I’ thinking from logical, conscious and slow ‘system II’ thinking. In a similar fashion, Thaler and Sunstein make an ideal-typical distinction between the so-called *Homo economicus* (or ‘econ’) and *Homo sapiens* (or ‘human’). Econs are perfect decision-makers who weigh pros and cons and always make rational decisions, but real humans make mistakes, are vulnerable to suggestion and may make decisions that are not beneficial for them. Nudges are based on these insights about human decision-making and aim to adjust the ‘choice architecture’ (the design of the physical or digital environment) in a way that induces better decisions and thus, for example, promote health. According to Thaler and Sunstein, “a nudge ... alters people’s behavior ... without forbidding any options or significantly changing [the nudges’] economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid” (2008, p. 6). By making small adjustments in the existing environment, nudges (mildly) steer individuals towards certain more desirable decisions. For example, in relation to food choices, there are various opportunities for health-promoting nudges. Subtle adjustments in the accessibility of food products could decrease food consumption (Rozin *et al.*, 2011). Supermarkets could use the fact that people often choose products that are placed at eye level (because it is easy and fast) and implement nudges that gently push people towards healthy product choices by placing healthier products at eye level (say, fruit and vegetables instead of candy and chocolate).

Since its first publication in 2008, *Nudge* has received a lot of attention, especially in the field of (public) health policy and practice. Several governmental behavioural insight teams (BITs) are exploring the potential of nudges for

public policies, and psychologists are investigating whether and how nudges could work for promoting a variety of behaviours, such as healthy eating or sustainable living. Early on in their book, Thaler and Sunstein address the question of whether governments should be allowed to stimulate the behaviour of citizens via nudging, and they argue that within their liberal paternalist framework it is indeed acceptable. Thaler and Sunstein argue that their framework for nudges will always leave a free choice for individuals. A central feature of nudging is that nudges do not forbid or limit the choice set; the choice set is simply arranged in such a way that the most desirable option is most likely selected. This means that individuals can always decide to choose alternative options. Yet, criticism has been raised against its legitimacy (Bovens, 2009; Selinger & Whyte, 2011; Blumenthal-Barby & Burroughs, 2012; Hansen & Jespersen, 2013). Nudges might steer behaviour in covert ways, diminish freedom of choice or even be manipulative (e.g. because they exploit flaws in human reasoning or decision-making). In other words, nudges threaten individual autonomy. To us, it is an open question as to whether individual autonomy is infringed and threatened, as autonomy is a complex notion that can be understood in different ways. Autonomy can, among others, be understood as self-government, independence, free will or self-rule. In fact, in the literature, several different concepts of autonomy are distinguished. Philosopher Joel Feinberg (1986), for example, distinguishes between: (a) autonomy as a capacity that people may have; (b) the actual condition or state of autonomy; (c) autonomy as an ideal to strive for; and (d) autonomy as a right that should be respected. Other philosophers describe personal autonomy as self-governance (like neo-Kantian philosophers) or as relational autonomy in feminist philosophical literature. It matters which conceptualisation of autonomy one uses to determine whether it will be infringed, as an infringement on rights differs greatly from an infringement on an ideal or self-governance. Autonomy itself is an overarching concept that is quite abstract and can be filled and interpreted in different ways (conceptualisations). In this review, we are interested in what conceptualisations of autonomy are assumed in the literature on nudging.

We hypothesise that, indeed, quite different understandings of autonomy are at play in the debate. This might help us to understand why some authors claim that a nudge threatens autonomy, while others will argue that it instead strengthens autonomy. In discussions on the acceptability of nudging, this seems relevant, especially in the context of liberal democratic societies and in relation to (public) health. It might also be that – possibly useful – conceptualisations of autonomy are overlooked or one view is dominating the debate. We therefore performed a systematic review on nudging, focusing on the question of what considerations of autonomy are assumed

or reflected in the ethical debate. The articles used in this literature review were scanned for considerations regarding autonomy and nudging. A basic assumption behind this approach is that authors rarely present a specific, well-defined concept of autonomy, but that this remains largely implicit. For example, if someone discusses autonomy only via the notion of informed consent, this assumes a view on autonomy that focuses primarily on the decision-making capacities of individuals. The review explores considerations that explicitly or implicitly appeal to autonomy. This could refer to various sorts of phrases. Some authors might simply offer a thought or present what is stated in the literature. We wanted to include the greatest possible number of quotes in the literature in order to grasp what underlying views on autonomy are presupposed and used. The aim of this review was to systematically map the various considerations concerning autonomy in the literature on nudging, so that different autonomy conceptualisations could be identified. Such a differentiation could provide directions for evaluating the legitimacy of nudges by suggesting how nudges may differentially affect autonomy depending on the assumed notion of autonomy.

Methods

Systematic review of considerations

We agree with Strech and Sofaer that “traditional systematic reviews traditionally answer an empirical question based on an unbiased assessment of all the empirical studies that address it” (2012, p. 121). They developed an alternative review of argument-based literature that enquires as to which reasons have been given when discussing an ethically relevant topic. The benefit of such reviews, contrary to what is often found in bioethics debate, is that they offer a more systematic evaluation of relevant considerations and can highlight blind spots and reduce the risk of neglecting considerations. We decided to use their basic step-by-step plan to select and analyse the literature. In particular, the coding system that we used (as described in more detail below) was inspired by the recommended method for extracting and synthesising data in their framework (Strech & Sofaer, 2012).

Selection of articles

To find relevant articles for the review, a search was conducted in three scientific databases: Scopus, PubMed and Google Scholar. In Scopus, we used the search words [Nudg* AND Ethic*] in the title, abstract or keywords. In PubMed, we used the search words [Nudg* AND Ethic*] in the title or abstract. In Google Scholar, we used variations of the search words

Nudge and Ethics in the title. A total of 101 articles were identified in the period until spring 2015. To check whether certain relevant articles (that did not mention or discuss nudging in their title, abstract or keywords) were missing, an additional search on a different set of search words ([Choice Architecture AND Ethic* NOT Nudg*] OR [Libertarian Paternalism AND Ethic* AND NOT Nudg*]) was conducted in Scopus. This yielded no new articles. Two researchers selected the articles based on an analysis of the abstract or type of document that was found (AV and MvdH). We excluded conference papers, theses/student papers and (most importantly) articles that did not discuss the ethics of nudging or libertarian paternalism. The latter could be: articles that only used the verb ‘nudging’ in a literal sense (‘to push’); articles that were published before the publication of Thaler and Sunstein’s book *Nudge* in 2008; and articles that only briefly mentioned that there could be ethical issues, but in general focused on another subject, such as the effectiveness or working mechanisms of nudges. This resulted in the selection of 33 articles.¹ We did an additional search in Scopus on a different set of keywords to check whether more articles could have been included in the review. We used the keywords [Nudg* AND Autonom* AND NOT Ethic*] to check whether we missed certain articles that did discuss autonomy in relation to nudging without specifically mentioning ethical theory. After reading the abstracts from the 29 articles that were identified after this extra search, we decided to focus on the 33 articles already included in the review (mainly because most of the ‘new’ articles did not seem to lead to new or different conceptualisations of autonomy that were not yet covered in the included articles).

Analysis

The 33 selected articles were analysed in Atlas.ti following three steps:

- (1) In each article, we identified phrases that explicitly or implicitly referred to a consideration of autonomy and labelled these with autonomy codes, employing terms used in the text (or terms closely mimicking the words in the text).

In more detail, each passage that discussed autonomy was marked as an autonomy consideration and labelled with autonomy codes (AV). Because we used a bottom-up approach, we did not follow strict rules for what was considered a relevant autonomy consideration or autonomy code. Our aim was to identify a broad range of autonomy

¹ The complete list of included articles can be retrieved from the corresponding author.

considerations and codes that, at a later stage, could be narrowed down. The labelling was checked by one of the other researchers (MvdH or MV) and changed, if necessary, after discussion. Some examples of autonomy-codes are: *'autonomy as the capacity to choose a conception of the good life for oneself'*, *'deliberation'*, *'freedom of choice'*, *'manipulation of individual decision-making'*, *'respecting autonomy as letting people decide for themselves free from coercion'* and *'you were never free in the first place'*.

- (2) Subsequently, a first step in systematisation was made by grouping the extensive set of initial codes, employing a more limited (yet still relatively large) set of super-codes.

In a deliberation round, three researchers (AV, MvdH and MV) separately looked at a selection of the longlist of autonomy codes and created super-codes. These super-codes were then checked by at least one other researcher and changed if necessary. When it was unclear which super-code would fit best with a certain autonomy code, the researchers went back to the original article(s) and the passage of text where autonomy was being discussed ('the consideration') in order to see which super-codes would fit best with certain autonomy codes. Some examples: the autonomy codes *'awareness'*, *'reflection'* and *'unconscious influences'* were assigned the super-code *'consciousness'*; *'degree of control'*, *'fully noncontrolled by others'* and *'substantial noncontrol'* were assigned the super-code *'control'*; *'easily resist the influence'*, *'easy to opt out of'* and *'fairly easy for people to choose a different direction'* were assigned the super-code *'easy resistibility'*; *'falls short of our own ideals'*, *'focus on people's own ends'* and *'what I would want were my desires properly informed and internally consistent'* were assigned the super-code *'own goal'*; *'able to keep my cravings in check'*, *'temptation'* and *'to constrain my own behaviour'* were assigned the super-code: *'self-control'*; and *'leading a self-determining life'*, *'respect for self-determination'* and *'self-determination'* were assigned the super-code *'self-determination'*.

A complete list of super-codes is presented in Online Appendix 1.

- (3) Finally, we aimed to formulate several general conceptualisations of autonomy that could cover the whole range of super-codes in a sensible way. For steps 2 and 3 of the analysis, we used an iterative approach, meaning that we went back and forth between the codes to come to a final list of autonomy codes linked to super-codes and grouped under the three conceptualisations of autonomy.

Having identified the super-codes, we needed to figure out what underlying assumptions of autonomy appeared most applicable to these categories of considerations. At this interpretative stage, we also referred to philosophical discussions and theories of autonomy. The aim was to find a small set of different conceptualisations that best matched the data found in the literature, and simultaneously could be grounded in the philosophical literature on autonomy. Finding this match involved going back and forth between philosophical concepts and the data. Three researchers (AV, MvdH and MV) each used the super-codes to suggest and describe several general conceptualisations, and in a joint meeting it was discussed which of their proposals was most convincing in terms of (a) the smallest possible set of conceptualisations, which (b) still had sufficient explanatory power (i.e. covering the variety of the data in a philosophically meaningful way).

Results

From considerations to conceptualisations

In the 33 articles of our review, we identified 280 autonomy considerations (i.e. pieces of text in which terms were used such as ‘autonomy’, ‘freedom’ and ‘self’). Each of the considerations was next labelled with one or more codes. The codes used, as much as possible, literal words from the marked considerations themselves. A total of 1282 unique codes were ascribed to the considerations, of which 790 were unique autonomy codes. Non-autonomy codes were, for example, codes that indicated the context in which nudging was discussed (e.g. ‘health’, ‘medical’ or ‘safety’) and the types of nudges that were being discussed (e.g. ‘warnings’, ‘defaults’ or ‘nudging in general’).

The 790 autonomy codes were then grouped under 61 unique super-codes. At this stage, some codes were excluded from the analysis because they were actually not indicating an aspect of autonomy. These codes we labelled ‘no super-code’. Some examples of no super-codes were ‘framing’, ‘non-maleficence’, ‘power’ and ‘responsibility’.

Finally, in our last interpretative stage, we grouped the 55 remaining super-codes under three general conceptualisations of autonomy, which we labelled ‘freedom of choice’, ‘agency’ and ‘self-constitution’. The conceptualisations are explained in detail below. In this last step of the analysis, we excluded super-codes such as ‘autonomy’ and ‘paternalism’, because these super-codes are too general to be linked to one particular conceptualisation. [Table 1](#) presents the total number of codes and super-codes linked to each of the three conceptualisations. ‘Agency’ covered the largest group of autonomy

Table 1. Conceptualisations, super-codes and codes.

Conceptualisations	Super-codes (<i>n</i>)	Codes (<i>n</i>)
Freedom of choice	11	188
Agency	29	326
Self-constitution	15	95
Total	55	609

codes (326 codes), ‘*freedom of choice*’ covered 188 autonomy codes and ‘*self-constitution*’ covered a smaller group of 95 autonomy codes.

Freedom of choice

The first conceptualisation of autonomy, freedom of choice, involves at least the availability of options that are accessible (in the choice environment), thus creating the possibility that people can make a choice. It therefore refers in important respects to external circumstances that influence choice. Freedom of choice is most clearly violated when the most relevant alternative options are taken away and persons are forced or even coerced to act in specific ways, thereby not having a real choice at all. The idea of freedom of choice connects most clearly to John Stuart Mill’s work on individual liberty as something that in principle should not be constrained by others (unless such restrictions are necessary to protect others) and to Isaiah Berlin’s concept of negative freedom (Mill, 1859; Berlin, 1958). Endorsing autonomy in this sense implies that nudges should not make options impossible or costly, and incentives are ideally positive and not negative. Freedom of choice is clearly at stake in Yashar Saghai’s work on nudging, where he presents the idea of easy resistibility. Saghai explains the concept of easy resistibility and in doing so he differentiates between full noncontrol and substantial noncontrol, the latter being, according to him, “more difficult to delineate ... and more relevant to the analysis of nudges.” He defines easy resistibility as “A’s influence is easily resistible if B is able to effortlessly oppose the pressure to get her to ϕ if she does not want to ϕ ” (2013b, p. 489). If nudges have at least the opportunity to choose differently, then the nudge could be considered easy resistible. If the design of the environment allows people to navigate away from the nudge, then the preferred option in the choice architecture is easy resistible. One could argue that freedom of choice is respected as long as a nudge is easy resistible. Moreover, the idea of freedom of choice is contained in Thaler and Sunstein’s definition of a nudge.

Freedom of choice is not only at stake as a value that can be infringed, but nudges may also strengthen freedom. For example, nudges may promote freedom of choice if they clarify the differences between types of options and offer some meaningful structure in a context where choice options are abundant (see Table 2).

Agency

Apart from a context that allows choice, autonomy also requires a *capacity* to choose and decide, and this refers to *agency*. Agency involves being able to lead one's life and act on the basis of reasons and intentions – hence going beyond mere stimulus–response behaviour. An agent is – at least to some extent – in control and capable to navigate in the world on a voluntary basis. This presupposes that the person has relatively stable ultimate goals, can reason about what options are preferable given those goals, and can reflect on the choices he or she makes and has made. Practical reasoning is a necessary capability in agency. Such a capacity to set goals, engage in practical reasoning and govern one's life (rather than just following instincts or immediate desires) are indispensable elements of any understanding of autonomy, although they are arguably most explicitly and most influentially elaborated upon in the work of Immanuel Kant (1785) and in neo-Kantian moral philosophy (e.g. Scanlon, 1998). The ultimate threat is manipulation, as in lying or deception (Kant, 1797), where someone's reasoning and decision-making processes are hijacked by others. Indeed, nudges that *exploit* weaknesses in people's reasoning capabilities in such a way that they might lose control could be considered a form of manipulation in that sense. It is often assumed that nudges are potentially manipulative. However, some discussions also refer to possible autonomy-enhancing aspects of nudges. Wilkinson comments, “not every influence that helps people do what they want is manipulative; some will be what Saghai [2013b, 2013c] calls ‘facilitative’”. The problem is in distinguishing manipulative and non-manipulative influences” (2013, p. 486). Indeed, nudges may also strengthen a person's ability to be in control, such as by helping them to avoid irrational behaviour or steering them away from instinctive choices they would not have made if they would have had the opportunity to make a deliberate decision. Moreover, nudges that help persons to overcome reasoning gaps and support learning processes are empowering and may strengthen their agency (see Table 2).

Self-constitution

Although freedom of choice and agency cover a large part of the autonomy considerations that we found in the selected articles, it was less easy to

Table 2. Freedom of choice, agency and self-constitution: illustrative quotes.

Illustrative quotes	Super-codes
Conceptualisation: freedom of choice	
“Nudging through changing the default choice is a paradigm example where significant – even dramatic – shifts in public behavior can be achieved while leaving the choice set unchanged and all choices readily accessible.” (Cohen, 2013, p. 3)	Choice
“It is a wrong assumption ... that public health policies should always leave people a choice among options, and thus give precedence to freedom of choice.” (Verweij & van den Hoven, 2012, p. 17)	
“The second misconception ... is the idea that paternalism goes hand in hand with coercion.” (Ménard, 2010, p. 232)	Coercion
“Libertarian paternalism does not justify any policies that completely eliminate certain options or that are coercive.” (Ménard, 2010, p. 233)	
“Yet [libertarian paternalism] does not threaten liberty: It neither blocks any choice nor makes it (more than trivially) costlier or more burdensome.” (Cohen, 2013, p. 3)	Liberty
“Autonomy is preserved ... by non-interference ...” (Cohen, 2013, p. 5)	
Conceptualisation: agency	
“The common objection against nudges [is] that they are autonomy-thwarting because they foster irrationality.” (Saghai, 2013a, p. 33)	(Ir)rationality
“Nudge therapy can even enhance patient autonomy. So much of our daily behavior is irrationally and unconsciously determined.” (Gold & Lichtenberg, 2012, p. 19)	
“Removing unimportant options may enhance the influencee’s ‘deliberative capacities and, perhaps, her freedom’.” (Wertheimer, as cited in Saghai, 2013c, p. 499)	Deliberative capacities
“All nudges rely on what I call ‘shallow cognitive processes’. I will use the term to cover both nondeliberative and incompletely deliberative processes ...” (Saghai, 2013b, p. 489)	
“... to ensure that everyone can unmask the manipulation if they wish to do so.” (Bovens, 2009, p. 15)	Manipulation
“The problem is that nudges might appear to manipulate the choices patients make, rather than merely improve the information required to enable informed consent.” (Brooks, 2013, p. 22)	
Conceptualisation: self-constitution	
“In the case of subliminal images, we are in no way the author, or even co-author, of these propensity changes. And this is why it offends a requirement of self-determination.” (Bovens, 2013, p. 495)	Authorship
“... a life endowed with enough material resources and opportunities to put the individual in a position actually to elaborate and execute a ‘plan of life’. The overarching ethical goal is to secure a substantial degree of control over the broad shape of one’s life.” (Saghai, 2013b, p. 493)	

Table 2. (Cont.)

Illustrative quotes	Super-codes
<p>“It may be that working through non-rational mechanisms fails to respect the rational core of the agent, in virtue of which she or he has human dignity – a version of Kantian ethics.” (Ashcroft, 2013, p. 494)</p>	Dignity
<p>“The idea of ‘dignity’ is complex and contested. We might begin by suggesting that ... the antonym of dignity is humiliation. Some nudges might seem to compromise dignity and respect for persons.” (Sunstein, 2015, p. 31)</p>	Own goal
<p>“We shall not refer to it as autonomy enhancing unless it uses its suggestive power to steer people only toward self-proclaimed goals.” (Cohen, 2013, p. 9)</p>	Own goal
<p>“When nudging goes beyond the supply of simple information, or reminders ... what is involved is the nudger acting in a semi-covert or covert manner to further the nudger’s own conception of the target’s welfare – which may involve re-shaping the target’s idea of their own welfare.” (Baldwin, 2014, p. 846)</p>	Own goal

include some remaining super-codes, notably ‘*authorship*’, ‘*own goal*’, ‘*choose not to choose*’ and ‘*trust*’. Considerations that were labelled with these super-codes deal more with a person’s identity than his or her agency and freedom of choice. Individuals constitute, develop and reaffirm their own identity in the choices they make, the goals they endorse and the values they cherish. A person’s practical identity is not so much a factual description of that person, but rather a description under which he or she values herself as a person and considers her actions to be worth undertaking or even obligatory (Korsgaard, 1996, p. 101). Autonomy as self-constitution puts this element central, emphasising a person’s individuality and authenticity: his or her ultimate values are not contingent, they are essentially *his or hers* – the result of a reflective endorsement that constitutes his or her *self*. This does not imply that personal autonomy is necessarily self-orientated: any individual person’s self is constituted by his or her relations with others, and relationships of mutual trust and support are indispensable preconditions for autonomy (Baylis *et al.*, 2008; Wardrope, 2015). One of the deepest threats to autonomy as self-constitution is *indoctrination*, making persons endorse ideas and values that are not truly their own. Marketing strategies or other policies that make people develop preferences and make choices on the basis of the ‘image’ of particular products or options, rather than on what they have reason to value, could be considered as mild ‘threats’ to autonomy as self-constitution. Nudging policies might undermine autonomy as self-constitution if they are effective in adjusting and shaping people’s preferences and values. Baldwin,

for example, argues that “when nudging goes beyond the supply of simple information, or reminders ... what is involved is the nudger acting in a semi-covert or covert manner to further the nudger’s own conception of the target’s welfare – which may involve re-shaping the target’s idea of their own welfare” (2014, p. 846). Autonomy as self-constitution would be threatened in a world in which a large number of choice situations are designed to promote a specific value. If every choice possibility is exploited and designed to nudge people towards healthy choices, it becomes less and less clear how a healthy lifestyle can be authentic. Another way in which nudges could infringe on autonomy in this sense is by diverting persons from the core values they see as constituting their own identity. On the other hand, autonomy as self-constitution can also be *supported* by nudging policies. Well-considered nudges enable people to live their lives without having to consider and deliberate about what they choose all the time, but rather trusting that their (social) environment facilitates better choices and enabling them to pay attention to what they truly care about, like commitments, life plans and special projects (see Table 2).

Discussion

The aim of this study was to investigate how autonomy is understood in the ethical debate on nudging. We used a systematic literature review of considerations to identify and differentiate between conceptualisations of autonomy as used in the literature on the ethics of nudging. The resulting conceptualisations – freedom of choice, agency and self-constitution – reflect the different ways in which autonomy is understood in relation to nudges. We showed how each conceptualisation of autonomy can be promoted and threatened (see Table 3). This matches the diversity found in the literature so far, where some claim that autonomy via nudges is protected, encouraged or infringed.

Our study, however, shows that claims do not simply refer to one view on autonomy, but to different conceptualisations. Distinguishing between these different conceptualisations can be helpful in order to determine the legitimacy of nudging in two ways. Firstly, if autonomy can both be infringed and empowered, then ideally nudges should be chosen in policies in such a way that they have a net positive impact on autonomy. Secondly, it needs to be clarified what conceptualisation of autonomy nudges need to equip in order to be (more) legitimate. One way to explore such questions would be to empirically investigate what sense of autonomy is considered most important by people in a specific context and how these are affected by specific nudges. Interestingly, most articles discussed the ethical aspects of nudging in relation to (public)

Table 3. Potential effects on autonomy.

Potential effects on autonomy	Freedom of choice	Agency	Self-constitution
Threaten	Coercion	Manipulation	Indoctrination
Promote	Navigability	Deliberative capacities	Authenticity
In sum, nudges influence:	What to do	The capacity to think and act	Who you are and what you value

health issues, which may reflect the tension between paternalistic approaches and libertarian approaches that is strongly pronounced in public health. Public health may thus provide a suitable context for empirically investigating nudging in relation to autonomy.

With regards to the three conceptualisations, we can observe that the first conceptualisation, freedom of choice, reflects what we expected to find in the literature. It fits well with Thaler and Sunstein's (2008) definitions of nudging and libertarian paternalism. Thaler and Sunstein stress that freedom of choice is preserved as long as choices are not limited or made more expensive. Yet, this conceptualisation was not dominant in the literature. Most codes ($n = 326$) and super-codes ($n = 29$) were covered by the conceptualisation of agency. Agency relates to various sub-concepts such as consciousness, deliberation and reasoning, and an important concern about nudges is that they aim to bypass those aspects of agency. How the two – freedom of choice and agency – intertwine in a nudging debate, whether the protection of the one can infringe the other, needs further elaboration. If that were the case, then the legitimacy of nudging would be less likely.

A possibly interesting finding is that the third conceptualisation, self-constitution, appears to play less of a role compared to freedom of choice or agency (95 autonomy codes versus 188 and 326 autonomy codes, respectively). This may reflect how philosophers and bioethicists often put reasoning, reflection and control central, rather than trust and dependency. Arguably, some nudges might be considered as ethically problematic if one values authenticity. Moreover, the idea that someone's self or identity could be strengthened by nudges, such as by enabling individuals to trust that simple choices preserve their well-being, should receive more attention.

For that matter, the idea that nudges can empower one's autonomy is criticized by Tengland. He argues that "the major ethical problem with the behavior-change approach is that it allows and endorses manipulative strategies" that create inauthentic wants and install false beliefs in agents, thus disregarding the right to autonomy (self-determination) (2012, p. 144). We think that his

critical stance is based on a certain conceptualisation of autonomy, namely that of agency, and that it leaves open the possibility that autonomy can be promoted or empowered in different conceptualisations. This also depends, of course, on the question of whether nudging is a type of manipulation or not, a debate that is also ongoing in the literature. It might be that the legitimacy of nudging will depend on both aspects, namely the core characteristics of a nudge (manipulative or not) and the possibility that it will strengthen the autonomy of individuals. In our literature review, we also found that manipulation (and thus agency) was discussed in more detail and more often than the idea of autonomy empowering nudges.

As suggested before, nudges and other behavioural change interventions could influence the three dimensions of autonomy independently of each other. One could, for example, imagine that limiting options threatens individuals' freedom of choice, while at the same time such an intervention could strengthen the agency of an individual. If we take food products as an example, decreasing product choice from ten to five products or highlighting/strategically placing three out of ten products may hinder freedom of choice (because one would have fewer options to choose from) and promote agency (because one would be able to deliberate better about the choice to be made).

There are some limitations to this study. First, this review was done over a period of two years and the included articles were published before or until the moment we started with the review (spring 2015). It could be that 'newer' articles on the ethics of nudging discuss other conceptualisations of autonomy that were not addressed in the 33 articles included in our systematic review. We are aware of some of the potential 'missed' publications. Second, as with most qualitative research, we tried to limit our own subjective judgements as much as possible. Especially with regard to the numerical results (codes), this means that hard conclusions cannot be drawn based on the quantitative data alone. In contrast, these results should always be seen as *supportive* of the categorisation we offer/introduce in this article. Third, there was some noise in the literature, especially with regards to the definition of a nudge, which was used rather broadly, including behavioural change and persuasion techniques that some would not consider to be 'true' nudges. Ploug and Holm (2013), for example, discussed what they call '*pharmaceutical nudging*'. With this they mean evaluative conditioning in the context of direct-to-consumer advertising of prescription pharmaceuticals. In their article, they argue that if patients are conditioned towards false-positive beliefs about the efficacy and safety of certain prescription drugs, this would compromise autonomy. While this is an interesting discussion, the method described is not a nudge; instead, the techniques should be seen as persuasive advertising (using false information).

In sum, we investigated which explicit or implicit conceptualisations of autonomy are used in ethical discussions on nudging. Through a systematic review of considerations, we explored and identified three conceptualisations of autonomy: freedom of choice, agency and self-constitution. Out of these three conceptualisations, agency was discussed the most in the literature, while self-constitution seems to be underexposed in the current debate. Nudges could not only affect these conceptualisations in a positive as well as a negative way, but also independently of each other. Whether nudges strengthen or threaten autonomy therefore not only depends on the nudge or situation, but also on the often implicit understanding of autonomy.

The distinction between these three conceptualisations of autonomy could be beneficial for various purposes. For instance, it could be used to distinguish between autonomy-threatening and autonomy-empowering nudges or the conceptualisations could be translated to and used in studies to empirically investigate the impact of nudges on autonomy. The differentiation thus provides directions for evaluating the legitimacy of nudges by suggesting how nudges may differentially affect autonomy depending on the assumed notion of autonomy.

Funding

This research was funded by a research grant from the Netherlands Organisation for Scientific Research (NWO) as part of the Welfare Improvement through Nudging Knowledge (WINK) research project.

Supplementary Material

To view supplementary material for this article, please visit <https://doi.org/10.1017/bpp.2018.5>.

References

- Ashcroft, R. E. (2013), 'Doing good by stealth: Comments on "Salvaging the concept of nudge"', *Journal of Medical Ethics*, 39(8): 494.
- Baldwin, R. (2014), 'From Regulation to Behaviour Change: Giving Nudge the Third Degree', *Modern Law Review*, 77(6): 831–857.
- Baylis, F., N. P. Kenny, and S. Sherwin (2008), 'A relational account of public health ethics', *Public Health Ethics*, 1(3): 196–209.
- Berlin, I. (1958), 'Two Concepts of Liberty', In Isaiah Berlin (1969) *Four Essays on Liberty*, Oxford: Oxford University Press.
- Blumenthal-Barby, J. S., and H. Burroughs (2012), 'Seeking better health care outcomes: The ethics of using the "nudge"', *The American Journal of Bioethics*, 12(2): 1–10.
- Bovens, L. (2009), '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*, (pp. 207–219). Springer.

- Bovens, L. (2013), 'Why couldn't I be nudged to dislike a Big Mac?', *Journal of Medical Ethics*, 39(8): 495–496.
- Brooks, T. (2013), 'Should We Nudge Informed Consent?', *The American Journal of Bioethics*, 13(6): 22–23.
- Cohen, S. (2013), 'Nudging and informed consent', *The American Journal of Bioethics*, 13(6): 3–11.
- Feinberg, J. (1986), *Harm to self*, New York: Oxford University Press.
- Gold, A., and P. Lichtenberg (2012), 'Don't call me "nudge": The ethical obligation to use effective interventions to promote public health', *The American Journal of Bioethics*, 12(2): 18–20.
- Hansen, P. G., and A. M. Jespersen (2013), 'Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy', *The European Journal of Risk Regulation*, 1, 3–28.
- Kahneman, D. (2003), 'A perspective on judgment and choice: mapping bounded rationality', *American Psychologist*, 58, 697–720.
- Kahneman, D. (2011), *Thinking, fast and slow*, Penguin Books.
- Kant, I. (1785), *Grundlegung zur Metaphysik der Sitten*.
- Kant, I. (1797), 'On a supposed right to lie from altruistic motives', In P. Singer (Ed.) *Ethics. An Anthology*, (pp. 280–281). Oxford University Press.
- Korsgaard, C. M. (1996), *The sources of normativity*, Cambridge University Press.
- Ménard, J-F. (2010), 'A "nudge" for public health ethics: Libertarian paternalism as a framework for ethical analysis of public health interventions?', *Public Health Ethics*, 3(3): 229–238.
- Mill, J. S. (1859), *On Liberty*, London.
- Ploug, T., and S. Holm, (2013), 'Pharmaceutical "Nudging"-reinterpreting the ethics of evaluative conditioning', *The American Journal of Bioethics*, 13(5): 25–27.
- Rozin, P., S. Scott, M. Dingley, J. K. Urbanek, H. Jiang, and M. Kaltenbach (2011), 'Nudge to nobesity I: Minor changes in accessibility decrease food intake', *Judgment and Decision Making*, 6(4): 323–332.
- Saghai, Y. (2013a), 'Nudging in interpersonal contexts', *American Journal of Bioethics*, 13(6): 33–34.
- Saghai, Y. (2013b), 'Salvaging the concept of nudge', *Journal of Medical Ethics*, 39(8): 487–493.
- Saghai, Y. (2013c), 'The concept of nudge and its moral significance: a reply to Ashcroft, Bovens, Dworkin, Welch and Wertheimer', *Journal of Medical Ethics*, 39(8): 499–501.
- Scanlon, T. (1998), *What we owe to each other*, Harvard University Press.
- Selinger, E., and K. Whyte (2011), 'Is there a right way to nudge? The practice and ethics of choice architecture', *Sociology Compass*, 5(10): 923–935.
- Sharot, T. (2011), *The optimism bias: A tour of the irrationally positive brain*, Pantheon Books.
- Steel, P. (2007), 'The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure', *Psychological Bulletin*, 133(1): 65–94.
- Strech, D., and N. Sofaer (2012), 'How to write a systematic review of reasons', *Journal of Medical Ethics*, 38, 121–126.
- Sunstein, C. R. (2015), Nudging and choice architecture: ethical considerations. Forthcoming in *Yale Journal on Regulation*.
- Tengland, P. A. (2012), 'Behavior change or empowerment: on the ethics of health-promotion strategies', *Public Health Ethics*, 5(2): 140–153.
- Thaler, R. H., and C. R. Sunstein (2008), *Nudge: improving decisions about health, wealth, and happiness*, Penguin Books.
- Verweij, M., and M. van den Hoven (2012), 'Nudges in public health: Paternalism is paramount', *The American Journal of Bioethics*, 12(2): 16–17.
- Wardrope, A. (2015), 'Relational autonomy and the ethics of health promotion', *Public Health Ethics*, 8(1): 50–62.
- Wilkinson, T. M. (2013), 'Thinking harder about nudges', *Journal of Medical Ethics*, 39(8): 486.