Review of Philosophy and Psychology (2020) 11:137–156
https://doi.org/10.1007/s13164-019-00450-z

Nudging and Autonomy: Analyzing and Alleviating the Worries

Bart Engelen[1] · Thomas Nys[2]

Published online: 16 December 2019

Abstract
One of the most pervasive criticisms of nudges has been the claim that they violate, undermine or decrease people's (personal) autonomy. This claim, however, is seldom backed up by an explicit and detailed conception of autonomy. In this paper, we aim to do three things. First, we want to clear up some conceptual confusion by distinguishing the different conceptions used by Cass Sunstein and his critics in order to get clear on how they conceive of autonomy. Second, we want to add to the existing discussion by distinguishing between 'autonomy' as the ability to set your own ends and 'autocracy' as the ability to actually realize those ends (which is what most of the current discussion is actually focusing on). This will allow for a more careful ethical evaluation of specific nudge interventions. Third, we will introduce the idea of 'perimeters of autonomy' in an attempt to provide a realistic account of personal autonomy and we will argue that it can alleviate most of the worries about nudging being autonomy-undermining.

Keywords Nudging · Sunstein · Manipulation · Autonomy · Autocracy · Perimeters of autonomy

1 Introduction

One of the most pervasive criticisms of nudges has been the claim that they violate, undermine or decrease people's (personal) autonomy. In this paper, we analyze what

✉ Bart Engelen
B.Engelen@tilburguniversity.edu

Thomas Nys
T.R.V.Nys@uva.nl

[1] Tilburg Center for Moral Philosophy, Epistemology and Philosophy of Science, Department of Philosophy, Tilburg University, PO Box 90153, Tilburg, The Netherlands

[2] Philosophy and Public Affairs, Department of Philosophy, University of Amsterdam, Oude Turfmarkt 143, 1012 GCAmsterdam, The Netherlands

exactly this supposed violation of autonomy entails and whether (or better: when) the criticism holds.

In what follows, we take nudges to be deliberate changes in people's choice architectures with the intention of predictably influencing their behavior by tapping into a-rational psychological mechanisms – often labelled 'heuristics and biases' – and thus without merely informing, rationally persuading, incentivizing or coercing them (see also: Thaler & Sunstein 2008: 6).[1] Typical examples are the use of defaults (for example in organ donation), the verbal framing of different options (for example presenting the survival or the mortality rates of medical treatments) and changing the physical choice architecture in ways that make specific aspects salient (for example putting apples at eye-level in cafeterias, painting flies on urinals and white lines on roads). Some nudges have both rational (informative, reason-giving) and a-rational (nudging) aspects to them. Think of pictures of cancerous lungs on cigarette packages, which provide information in salient and emotion-inducing ways and (at least partly) rely on less rational mechanisms and thus do not *merely* inform people.

The fact that nudges involve the intentional exploitation of people's a-rational psychological mechanisms such as the 'status quo bias', 'conformity bias' and 'availability bias' but also people's emotions, laziness and akrasia,[2] explains why they have been criticized for violating people's autonomy. In their influential criticism of nudging, Daniel Hausman and Brynn Welch (2010: 128) forcefully argue that nudges "may 'push' individuals to make one choice rather than another. (…) when this 'pushing' does not take the form of rational persuasion, their autonomy – the extent to which they have control over their own evaluations and deliberation – is diminished". Other authors join them and claim that (at least some kinds of) nudges violate people's autonomy. Hansen and Jespersen (2013: 27) argue that non-transparent nudges such as framing techniques render "autonomous choice a mere fiction". Bovens (2009), who

[1] Note that we label the cognitive heuristics involved in nudging 'a-rational', and not 'rational' or 'irrational', because we believe that rationality requirements do not apply to them. Take the distinction between calling something 'a-moral' or 'immoral': only agents capable of making moral choices can be called 'moral' or 'immoral' (when they succeed or fail to live up to morality's requirements). Other things, like a piece of rock or a fish (to which morality's requirements do not apply) can only be 'a-moral' and not 'immoral'. In that respect, we call heuristics – such as loss aversion or the status quo bias – 'a-rational', since rationality requirements do not apply to them (but only to the agent and her beliefs, preferences and actions).

Like autonomy, rationality is a complex and heavily debated term. Conventionally, beliefs, preferences or actions are 'rational' if they involve some kind of uptake of information or reasons. On a minimalist understanding, only consistency is required. An ecological understanding of rationality asks whether decision-making processes are in some sense successful or not. Here, heuristics are typically taken to be 'fast and frugal', producing 'good' decisions in most circumstances. Only in some circumstances do heuristics lead to 'reasoning failures' or 'errors', in which case there is a trade-off between speed and accuracy. For more on this, see Andreas Schmidt (2019), who stresses that "decision-making procedures are not rational or irrational per se but only relative to particular environments and agents" (Schmidt 2019: 527). Schmidt uses an ecological understanding of rationality to argue that nudges are compatible with and may even support rationality, understood in this sense. As we want to focus on autonomy here, not rationality, we do not go into this discussion any further, which is the topic of Engelen (2019). We thank an anonymous reviewer for clarifying this point.

[2] Sunstein often employs a broader definition of nudges, which includes purely rational processes such as information provision (Sunstein 2015a: 512). We take the narrower approach, which is in line with most of the literature on nudging and with the spirit of Sunstein's work, which starts from the observation that ordinary people (Humans) are not always as rational as they can or want to be (Econs) because a-rational heuristics and biases influence their decisions.

was the first to express this worry, claims that nudges, much like subliminal messages, violate or undermine people's autonomy because they influence people's judgements and actions 'in the dark' and 'behind their backs'.[3] Or take Frank Furedi (2011), who deems it necessary to defend "autonomy against an army of nudgers" as those aim "to deprive people of the capacity for making wrong choices", which implies that "they cease to be choice-makers." According to Furedi, a world in which nudgers reign freely erodes people's responsibility, their capacity to make value judgments and their "moral independence".

Another well-known criticism of nudges, namely that they are manipulative, often involves worries about nudges undermining autonomy. As Robert Noggle (2018) puts it, "the assumption that manipulation undermines autonomy is so common in discussions of manipulation and consent that it would be difficult to cite a paper on that topic that does not at least implicitly treat manipulation as undermining autonomous choice." Martin Wilkinson (2013: 345), who analyzes whether nudges are manipulative and whether that should worry us, argues that autonomy is key to understanding both what manipulation is and why it is wrong: "manipulation is a form of influence that subverts and insults a person's autonomous decision making (…). What is primarily wrong about manipulation is that it violates autonomy". Regardless whether it is framed in terms of manipulation or not, the main worry remains the same.[4] Nudges involve two elements that arguably reduce people's control over their own actions: 1) the a-rational psychological mechanisms that influence people's behavior without them being fully aware of that influence and 2) another person's willful exploitation of those "less than fully autonomous (…) patterns of decision-making" (Bovens 2009: 209).

In fact, one can distinguish between two different concerns about autonomy. First, a short-term concern that autonomy is undermined whenever a nudge shapes one's preferences or influences one's choices behind one's back (cf. Bovens 2009; Hausman & Welch 2010). Second, a long-term concern that a world full of nudges erodes people's capacities for autonomous decision-making over time (cf. Furedi 2011). In what follows, we focus primarily on the first concern.[5]

In response to these criticisms, the best-known proponent of nudging, Cass Sunstein, has maintained that most, if not all, nudges respect people's autonomy and can even be said to promote it (Sunstein 2015a; 2015b; 2015c; 2016). In light of this, it becomes crucially important to clarify what exactly is meant by 'autonomy'. Is there agreement on some conception of autonomy that specific nudge strategies are plausibly claimed to thwart or support? Or are we witnessing a confounding debate at cross purposes?

Instead of delving into the multiple conceptions in the complex literature on autonomy (Dworkin 2017), we will focus primarily on the conceptions of autonomy used in the nudge literature itself in order to clean up the conceptual confusion that

[3] In their literature review, Vugts and co-authors (2018) analyzed 33 articles and identified no less than 280 autonomy considerations. While they employ a quite broad approach to autonomy, allowing for conceptions that stress freedom of choice, agency and self-constitution, this suffices to show how widespread autonomy worries are in the literature.

[4] For more detailed analyses about the interplay between nudging, autonomy and manipulation, see: Barton (2013), Blumenthal-Barby & Naik (2015), Mills (2015), Nagatsu (2015), Schubert (2017), Wilkinson (2013) and Nys & Engelen (2017). While all these authors introduce and explain the worries about nudges being manipulative and/or undermining autonomy, they argue that nudges do not necessarily undermine autonomy.

[5] Some of the criticism is based on the long-term effects of living in a world in which most of our choices would be influenced by nudges. We put this criticism aside for the purposes of this paper.

pervades it. In the first section, we will tease out four different conceptions of autonomy used by Sunstein and his critics as a first step to understanding the debates and fleshing out the worries. We show why Sunstein's view on this issue remains unconvincing due to his failure to provide a proper definition of autonomy and his tendency to evade the criticisms rather than responding to them directly.

In the second section, we introduce the distinction between 'autonomy' and 'autocracy', which has been absent from the literature. The distinction reveals that most of debates between Sunstein and opponents of nudges focus on whether the latter enable or inhibit people to *achieve* their goals (autocracy), while the real worry concerns the potential impact of nudges on people's ability to *set* their own goals (autonomy). In the third section, we analyze seven different scenarios to assess how which nudges can possibly impact autocracy and autonomy. We will argue that the real and underlying worry is that nudges might promote the autocracy of some at the expense of the autonomy of others.

In the fourth section, we further explore the notion of autonomy and argue that we should think of autonomy as having 'perimeters': boundaries within which changes in people's preferences and decisions can occur without those necessarily undermining their autonomy. This alleviates most worries since nudges typically do not make people end up beyond those perimeters. As such, we provide a novel justification of nudges that does not assume, as Sunstein does, an inner rational agent that knows exactly what it wants. In the fifth section, we draw some conclusions.

2 Four Approximations of Autonomy

It is interesting and revealing that Thaler and Sunstein (2008) nowhere mention the term 'autonomy' in their seminal book *Nudge*. Although Sunstein refers to it in his later work, he again never properly defines it. Instead, he 'approximates' rather than pinpoints its meaning exactly. We discuss four of those approximations, the responses from nudge critics and stress why these approximations do not succeed in clarifying and addressing the real worry about nudges undermining autonomy.

- **Autonomy as Liberty**

Most prominently, Sunstein addresses concerns about nudges violating autonomy by stressing their *liberty preserving* character. Liberty, which he understands as 'freedom of choice', has two aspects. First, it is about retaining the same set of *options*: nudges do not remove options that were available *before* the choice architecture was altered. Unhealthy food items are still on sale in the cafeteria; people can always opt-out of becoming an organ donor; and framing only reformulates the available options. Second, liberty concerns the freedom *to* choose. According to Thaler and Sunstein (2008: 5), "they do not want to burden those who want to exercise their freedom". Nudges do not impose extra burdens on the choice-making process and should be "easy and cheap to avoid" (Thaler & Sunstein 2008: 6) or "easily resistible" (Saghai 2013: 489).

As such, nudges are no forms of coercion, which Sunstein calls "the antonym of autonomy" (Sunstein 2016: 67). Worries about autonomy are addressed by stressing how nudges differ from more coercive interventions such as mandates, fines, laws and taxes, which clearly change the options at hand and are harder to avoid or resist.

Therefore, "respect for autonomy is adequately accommodated by the libertarian aspect of libertarian paternalism" (Thaler & Sunstein 2003: 1167).

However, this rhetorical strategy evades the real issue. After all, critics can respond by arguing that, even if nudges are liberty-preserving, they may still violate autonomy (Bovens 2009; Hausman & Welch 2010; White 2013; see also: Blumenthal-Barby 2012: 349–352). Even without coercion, they say, influences like nudges can diminish the control people have over their judgements and decisions. *Pace* Sunstein, the antonym of autonomy is actually not coercion, but heteronomy. At least some nudges influence people 'behind their backs', in ways that are sneakier and less obvious to people than laws and taxes. Chris Mills (2015: 497–498) explains why in the following quote.

> Heteronomous behaviour can be caused by any reason for action that motivates an individual contrary to (e.g., by overriding or subverting) their authentic will. Heteronomy specifically threatens the independence of an individual's will by disregarding her decision-making competency, thus bypassing part of what makes her decision her own (…). Critics may suggest that choice architecture is necessarily heteronomous because it seeks to exploit heuristics and cognitive biases in our reasoning. Accordingly, choice architects pursue a programme of manipulation that undermines the independence of an autonomous agent's will by subverting the flaws in her decision-making competency to bring about particular outcomes.

The criticism that nudges decrease autonomy does not concern people's options or ability to resist the nudge (this is why Sunstein evades rather than addresses the issue) but with the formation of their will.

- **Autonomy as Informed Decision-Making**

So autonomy seems to require more than just liberty. Sunstein acknowledges as much when observing that *ignorance* threatens autonomy, which implies that autonomous decisions not only need to be free but also informed.

> Many nudges promote autonomy. A choice should not be considered autonomous if it is based on ignorance; if someone takes a stomach medicine in the belief that it will help with a cold, there is no interference with autonomy in informing him that it is stomach medicine. Autonomy requires informed choices, and many nudges are designed specifically to ensure that choices are informed. (Sunstein 2015c: 437–438)

For people to be autonomous, they should have options but also sufficient information and the cognitive capacities necessary to make well-informed choices between those options. Autonomy as self-rule or self-governance can be obstructed not only by external interferences but also by internal "limitations such as an inadequate understanding" (Beauchamp & Childress 2008: 101).

In this light, nudges promote autonomy to the extent that they improve people's understanding of the options at hand. Nudges can provide the right kind of information *in the right way* so as to increase people's ability to digest the information well. Think of the smart use of visual or audio cues in GPSs, on cigarette packages or on the (green and red) labels for (healthy and unhealthy) food options. The salience of the information provided arguably corrects or avoids poorly informed decisions (Sunstein 2015c: 418; Sunstein 2016: 64–65).

Again, Sunstein fails to take seriously the worries critics have. While some nudges lead to more informed choices, not *all* nudges do. In fact, if informed choice is a requirement for autonomy, then some nudges may not promote but even undermine it.[6] First, putting apples at eye-level or changing organ donation defaults does not inform at all. Second, nudges might improve informed *choice* but not informed *choosing*. Autonomy, some say, means active *choosing* or choosing in light of and *because* of the evidence. People should be rationally persuaded or convinced by the available information. By tapping into a-rational, largely unconscious and automatic heuristics, nudges do not influence behavior via the uptake of information and thus do not secure the kind of decision-making processes needed for genuine autonomy. (We will later argue that this is an inflated account of autonomy that puts the bar for autonomy too high and would label most day-to-day decisions as heteronomous.)

In response, Sunstein (2014a) argues that nudges can and should respond to systematic and predictable *failures* of rational agency that are so pervasive that active choosing does not cut it. White lines on the road and apples at eye-level should be preferred over 'dangerous curve ahead' signs or nutrients labels because we *know* that most people fail to digest such information. But again, critics can stress, Sunstein confounds autonomy with other values here. While nudging may be preferable to informing in such cases, it is not because they promote autonomy but because these are situations in which we value people driving safely and eating healthily more highly than we value the promotion of autonomy.

- **Autonomy as Reflectively Making Choices that Matter**

In a third approximation, Sunstein argues that it is not only coercion and ignorance that can threaten autonomy, but also *hyper-reflexivity*. Autonomy then does *not* require active choosing, reflecting, deliberating and weighing reasons.

> It is also important to see that autonomy does not require choices everywhere. It does not justify an insistence on active choosing in all contexts. If we had to make choices about everything that affects us, we would quickly be overwhelmed (…). If people have to make choices everywhere, their autonomy is reduced, if only because they cannot focus on those activities that seem to them most worthy of their attention. (Sunstein 2015c: 438).

Overly relying on active choosing can actually reduce people's autonomy because it inhibits them from focusing on what matters. While active choosing *itself* is not the enemy of autonomy, it is the fact that our cognitive resources – what Sendhil Mullainathan and Eldar Shafir (2013) call mental 'bandwidth' – are limited. If we want people to make good, well-informed decisions, we should not overburden them, since there are limits to the mental capacities needed for reasoning, remembering and problem solving (cognitive capacity) and for planning and controlling impulses (executive control) (Mullainathan & Shafir 2013: 47). If we want to protect autonomy, we should economize and ensure that sufficient bandwidth is available when making

[6] Perhaps they often do not have any real (positive or negative) impact on people's autonomy as there may not have been much autonomy in the first place (before the nudge was implemented). People often make uninformed or insufficiently informed choices. In such cases, a lot of nudges will not promote or diminish their autonomy but leave them as autonomous or heteronomous as they were before (see also: Nys & Engelen 2017).

choices that really matter. Hyper-reflexivity thus tends be over-demanding and thereby threatens autonomy.

Sunstein thus argues that nudges that facilitate the many small, low-stake and day-to-day decisions actually "promote autonomy, in part because they open up time and resources for more pressing matters" (Sunstein 2014b: 21). Think of well-designed door handles facilitating people entering and exiting rooms. The rationale is to have people rely on their automatic processes for such minor decisions in order to enable them to reflectively make the bigger choices in life.

Again, Sunstein's approach here is partly evasive, as it neglects the following two worries. First, such nudges only concern trivial choices, while many of the proposed (and criticized) nudges are geared at important "decisions about health, wealth, and wellbeing" (Thaler & Sunstein 2008), such as defaults for organ donation and for insurance and pension schemes.[7] Second, 'bandwidth-saving' nudges at best promote autonomy *indirectly*. In a more direct sense, however, they may be at odds with autonomy. Again, the automatic or subconscious mechanisms that nudges tap into arguably preclude autonomy if you believe that autonomy essentially involves reflection, deliberation or active endorsement. Nudges can only free up mental bandwidth for reflective choices (which is what autonomy requires) by inducing unreflective choice processes for other decisions. Remember the quote by Hausman and Welch (2010: 128): "in addition to or apart from rational persuasion, [nudges] may 'push' individuals to make one choice rather than another (…). When this 'pushing' does not take the form of rational persuasion, their autonomy – the extent to which they have control over their own evaluations and deliberation – is diminished".

Sunstein, however, should be able to resist this line of criticism and provide a more robust defense of nudges leaving a person's autonomy intact. Behavioral insights about the ways in which automatic, unconscious and unreflective processes influence people's beliefs, preferences and decisions, show the need for a psychologically realistic conception of autonomy. If autonomy requires reflection and active endorsement throughout, this puts the bar unrealistically high and therefore risks making most judgements and decisions non-autonomous. In fact, the autonomy conceptions implicit in the criticism of nudging by Hausman and Welch (2010) and Bovens (2009) have been said to "demand a level of self-knowledge or self-transparency on the part of the individual that cannot be found in a behavioral world" (Schubert 2017: 337).

This point is increasingly acknowledged in the literature on personal autonomy. Take, for instance, Amy Mullin (2007: 537) who argues that reflection is not necessary for autonomy, and that it only requires a self, governing its own actions in accordance with what it cares about. Lowering the bar in this way opens up a wider range of things that can plausible be called 'autonomous'. The rationality and reflection required for autonomy does not mean that each and every choice is carefully pondered upon and only executed when it has explicitly and consciously received a 'stamp of approval' from our rational faculties. Such autonomy might be suited for Econs, but not Humans. Nor does it require 'active choosing' in the sense of *identifying* with a particular course of action. John Christman (2001), for example, has argued that it is more about *not being alienated* than about actively endorsing a certain option. In this sense, the support

[7] Note that some 'trivial' choices, such as what meal to have, can have big consequences over time, for example when they are made over and over again.

for a certain motive or desire might be quite weak. We agree with Christman and Mullin that people can act autonomously without much thought or reflection.

For now, let us contrast our approach with Sunstein's, who argues that autonomy is threatened by hyper-reflexivity. After all, Sunstein here concedes to his critics that autonomy is about reflective and active choosing by arguing that nudges indirectly serve that goal by clearing away potential distractions for trivial choices. In contrast, if autonomy is not about reflection or active choosing, nudges can respect and promote it as long as they enable people to live according to their own values and goals. Perhaps surprisingly, Sunstein's fourth approximation of autonomy, which we believe best formulates what Sunstein is really after, supports this reading.

- **Autonomy as Successfully Reaching Your Well-Considered Goals**

Sunstein (2014b: 138) argues that autonomy is not undermined as long as people's ends are respected. This shifts the focus from the decision-making process and the psychological mechanisms at play to the outcomes of those processes. If we want to assess their impact on people's autonomy, we should ask whether nudges help people actually achieve their goals. Quite often, Sunstein's *starting point* is that many people, on many occasions, do not manage to reach the ends they themselves have set. Nudges should facilitate that and help people end up where they want to end up (eating more healthily, not crashing in dangerous curves, et cetera). This focus on people's ends is clear in the well-known phrase that nudges should "make the choosers better off, *as judged by themselves*" (Thaler & Sunstein 2008: 5). Sunstein (2014b: 50) argues that nudges should be 'means paternalist' in that they facilitate people choosing the right means for their ends instead of correcting people's ends. When people predictably fail to achieve their ends, "occasions for paternalism" (Sunstein 2014b: 25) occur and nudges are justified.

Autonomy, on this understanding, is about people doing the things they really want (living healthily or saving more for retirement). You are self-determining and in control when you manage to realize what you hold dear. Nudges can help promote such autonomy.

Again, critics are not convinced and rightly so. They typically point to the many obstacles for nudgers to promote autonomy in this sense of helping nudgees achieve what the latter themselves judge best. Next to the problem that one-size-fits-all nudges in a pluralist world inevitably direct some people towards and others away from their goals, there is the epistemic challenge of figuring out what nudgees (really) want (Rebonato 2014; Rizzo & Whitman 2009; White 2013). Sunstein's strategy of referring to means-paternalist nudges does not address the worry that a lot of nudges fail to be means-paternalist (especially if nudgers have other interests at stake).

In addition, Robert Sugden (2015: 579) argues that Sunstein systematically assumes an "inner rational agent" here, whose idealized, informed and reflective judgements provide the direction in which nudgers should steer people. Sugden criticizes such a rational homunculus as "psychologically ungrounded", since these idealized judgements do not exist and we often have no way of finding out what people 'really' and thus hypothetically think and want.

While there are several ways for proponents to reply (see Engelen 2019), we suggest a middle road, between Sunstein (who justifies nudges based on the assumption that we can discover people's real desires and steer them towards those) and Sugden (who denies the existence of such desires). While Sugden is right in stressing that people often do not know what they (would) want (if they were fully rational), that does not

mean that nudges cannot help them achieve their goals. As will become clear in what follows, we take these goals to be broadly defined by the things people care about, their life plans and commitments. They stipulate constraints on what people want and exclude certain options as clearly undesirable but also leave quite some room for other options. We develop this novel justification of nudges in the fourth section on what we will call the 'perimeters of autonomy'.

3 Autonomy Versus Autocracy

But first, we want to build on Sunstein's last approximation and explain why we think this approach is promising (without, however, fully delivering on that promise). When Sunstein understands autonomy as 'the ability to achieve one's own goals', he shifts the focus from decision-making processes (the psychological mechanisms) to outcomes (what exactly is chosen in the end). The reason for this shift lies in the simple observation that the influence of these mechanisms and choice environments is often inevitable. Psychological weaknesses and shortcomings do not magically disappear when we remove nudges. While we can take away the deliberate 'exploitation' of these a-rational factors by an intentional agent, those a-rational factors typically remain in place. To the extent that autonomy is not questioned in the absence of nudges then, it should not be questioned when nudges tap into the psychological mechanisms that were already and will remain at play (like heuristics, laziness, status quo bias, et cetera). When figuring out whether someone is autonomous or not, an exclusive focus on the psychological mechanisms involved is not really fruitful.

Consider the implausibility of two autonomy conceptions that do focus on decision-making processes. Suppose, first, that autonomy is about *you* determining all of your decisions 'all the way' in the absence of any influences beyond your control. Or, second, suppose that autonomy requires not just achieving your goals (like improving your health) but also doing so for the right reasons (making healthy choices *because* you know they are healthy). In other words, you do what you want because you wanted it (Bovens 2009: 210). Both conceptions of autonomy are obviously more demanding than simply requiring that you reach your goals.

As mentioned before, the first conception turns autonomy into an unattainable mirage, a pristine kind of decision-making that is purely deliberate and reflective. To demand total independence and reflective choosing would be absurd (for we are always influenced, cf. Wilkinson 2013). It would put the bar for autonomy much higher in other situations in which we acknowledge autonomy without questioning it. In addition, it would leave the non-autonomous – like the weak-willed – hanging.[8] What about the people who want to live more healthily, for example, but who cannot follow through? What does respect for their autonomy entail? Should we respect their *non*-autonomy (as they clearly fail to determine their choices all the way through)? It seems that this provides a reason for intervention and that nudges can do exactly that (while informing and rationally persuading them will not help as their problem lies not in making up their mind but in acting on it).

[8] Our discussion below will argue that such weak-willed people are autonomous (they know what they want), but are lacking in autocracy.

The second conception is also overly demanding. While you might have a good reason for getting out of bed in the morning (an important job interview, for instance), what motivates you to do so is the annoying alarm clock and the triggered desire to stop that sound. In our view, this does not interfere with your autonomy (which does not requiring choosing the right things for the right reasons). We outsource so many aspects of our decision-making: to technology, to others, et cetera. There is nothing unique or especially worrying about relying on smartly designed choice architectures in this respect.

In order to elucidate Sunstein's understanding of autonomy and to show why it still is only an approximation and thus fails to provide adequately define autonomy, let us consider Paul Guyer's distinction between autonomy and *autocracy* in relation to Kant's practical philosophy.

> [Kant] urges that we put the idea of autonomy into practice by developing what he calls *autocracy* or 'self-mastery,' 'the authority to compel the mind, despite all the impediments to doing so,' involving 'mastery over oneself, *and not merely the power to direct.* (Guyer, 2003: 91; final emphasis added).

Guyer elaborates:

> The empirical realization of autonomy in the actual circumstances of human existence (…) requires, apparently, both that we directly strengthen the efficacy of the moral law on our conduct, and also that we learn techniques that indirectly support the reign of the moral law, by removing or diminishing impediments to its rule. (Guyer, 2003: 91).

For Kant, autonomy is, as the famous definition in the *Groundwork* has it, "a property of the will by which it can be its own law." Our willing is subject to the moral law. This makes us free and responsible beings who can *set* our own ends (which can be evaluated by the moral law) and either obey or disobey that law. Since we can and often do disobey the moral law, another issue arises: how can we, with all our flaws, come to act upon a good will? This is what Kant calls "mastery over the self" or 'autocracy', the empirical realization of autonomy, i.e. actually being able to act morally. Kant, in his Doctrine of Virtue (1797/1996), for example, suggests that we should visit 'poor houses' so that we vividly realize the plight of others, thereby "removing or diminishing impediments to" the moral law (such as our self-interested nature).

Like Sunstein and his critics, we focus not on moral autonomy (the moral law) but on personal autonomy (personal ends and goals). Self-rule or autonomy then is about the ability to *set* your own ends and autocracy about the ability to *achieve* those ends.[9] The standard of evaluation here is not moral but personal: your own values, the things you care about, your highest order volitions, et cetera. Just like Kant, who correctly drew attention to the problem of actually being motivated by the moral law, we too know that we can fail to achieve our own, personal ends. Determining the ends that are important to you (autonomy) is different from you actually being motivated accordingly and actually realizing those ends (autocracy).

[9] Schubert (2017: 338) asks whether nudges compromise 'self-legislation', which in his view goes beyond 'autonomy', as it relates to "people's *ability to form* (*or learn*) *preferences*" and the ability to cope with the "existential task to 'make something of themselves'", which nudges may compromise over the long haul by discouraging active choice. This is actually close to what we call 'autonomy' here: the ability to determine and set one's own ends.

This shows that the discussion between Sunstein and his critics on nudges facilitating or impeding people *achieving* their goals turns out to be all about autocracy and not about autonomy (which refers to people *setting* those goals). On the one hand, as Harry Frankfurt's example of the 'unwilling addict' (1988) shows, people can have autonomy (she does not want to smoke and is able to *set* her own ends) without autocracy (she fails to realize those ends). On the other hand, people can lack autonomy, for example when they are completely ambivalent and unable to make up their mind and determine what their wholehearted goals are.[10]

If autonomy is about the ability to *set* one's own ends and autocracy about the ability to actually (empirically) *realize* those ends, we can see that proponents of nudging, like Sunstein, base their defense of nudges (as justified means-paternalism) on autocracy. Nudges are for personal autocracy what visits to poor houses are for moral autocracy: helpful pushes in the back to actually do what one deems important. Nudges facilitating healthy choices improve the autocracy of those who want to live healthily but predictably fail to achieve that end 'on their own'. In addition, easily resistible nudges arguably preserve the autocracy of those who decidedly want to lead unhealthy lives as they can still pursue their goals (see also: Nys & Engelen 2017). Note that the critics too focus on autocracy when arguing why nudges often fail to be mean-paternalist. In these debates, autonomy is not at stake. In fact, it is assumed, since people's ends are taken for granted and the discussion revolves around whether nudges facilitate or impede people reaching them.

4 Do Nudges Affect Autonomy and Autocracy?

In order to get a better understanding of how nudges impact people's autocracy but also their autonomy, we distinguish between different possible cases or scenarios. For clarity's sake, we use a single example but the analysis works equally well for other examples. In our example, a nudge aims to steer customers in a cafeteria to vegetarian options (V) for example by making those more salient or making the meat options (M) less attractive. Exclamation marks (!) point to scenarios where the nudge successfully changes people's decisions. Question marks (?) indicate cases that are unclear at first sight. To assess the impact of a nudge on people's autocracy and autonomy, we need to compare the decisions of nudgees when subjected to the nudge to how they would have acted in its absence.

Case	Settled preference	Decision absent nudge	Nudged decision	Autonomy?	Autocracy?
1	V	V	V	Respected	Respected
2	V	M	V (!)	Respected	Promoted
3	M	M	M	Respected	Respected
4	M	V	V	Respected	Not further impaired
5	M	M	V (!)	?	?
6	?	V	V	?	?
7	?	M	V (!)	?	?

[10] On Frankfurt's own analysis (1988), the unwilling addict would be heteronomous, a passive bystander to her own motivating desire. For Frankfurt, a person who is deeply ambivalent is also non-autonomous (because not wholehearted). Our distinction allows us to distinguish these two failures of autonomy.

First, let us assume that some people have a settled preference for vegetarian options (cases 1 and 2). Most of these have made up their mind and stick to it. As they would have chosen vegetarian anyway (case 1), the change in choice architecture does not affect their autonomy (it does not inhibit them from making up their mind) or their autocracy (they end up choosing what they want anyway). However, some vegetarians may be weak-willed and tempted to pick meat if this were made salient. The nudge helps them achieve their ends and thus promotes their autocracy (case 2). These means-paternalist nudges also respect their autonomy, as they do not interfere with their capacity to set their ends.

Second, let us look at decided meat lovers (cases 3 and 4). Most of them will not be tempted to go vegetarian and thus stick to their settled preference for meat (case 3). Because it is easily resistible, the nudge is ineffective. Again, these people's autonomy and autocracy remain unaffected: they are *as able as they were before* in setting and realizing their own ends.[11] Now, some weak-willed meat lovers might, even in the absence of the nudge, fall for vegetarian options (just imagine they have a craving for soy) (case 4). As in case 2, they fail to act on their settled preference and thus have their autocracy impaired. In our view, the nudge towards vegetarian options leaves her autonomy intact (she still has the settled preference for meat) and does not further affect her autocracy (which is impaired anyway).

A critic could say that nudges are not that innocent here (case 4). If someone is 'unhappily addicted' (to soy, or more realistically, to alcohol), nudging her to go against her settled preference is problematic as it presents an additional hurdle for her to do what she really wants (avoid soy or stay sober). There are three responses to make here. First, this is a concern about autocracy (to what extent does the unhappy addict succeed in achieving her ends?), not autonomy (we basically assume she made up her mind). Second, the real question is what would happen if the nudge were removed. Since her addiction would lead her to go against her reflective preference anyway, her autocracy would in no way be restored. In most cases, we believe, opponents of nudges quite naïvely assume that removing them would somehow magically restore people's capacity to achieve their ends. Third, if you (rightly) care about restoring the unhappy addict's autocracy, it seems to us that (means-paternalistically) nudging her accordingly would be the way to go (which would turn it into case 2).[12]

Now, perhaps the most worrying scenario is where a meat lover is successfully nudged away from meat towards the vegetarian option (case 5). Surely, this would impair her autocracy (as it inhibits her capacity to achieve her ends) and perhaps even her autonomy (as it changes her mind). In our view, this scenario is highly unlikely. Nudges are, after all, easily resistible, which means that nudgees can become aware of and effortlessly oppose their influence (Saghai 2013). In fact, this is a conceptual point:

[11] If you think their goals are illegitimate and need to be changed, as ends paternalism would have it, then you need less resistible and more coercive interventions.

[12] Barton (2013) argues that, while salient health warnings on tobacco products decrease (the independence component of) people's autonomy by invoking a-rational heuristics, they also protect against (possibly misleading) messages from tobacco producers. In addition, they lead unhappy smokers to make more authentic choices and thus enable their capacity to rule themselves, increasing (the self-ruling component of) their autonomy. Overall, Barton argues, such nudges can foster smokers' overall autonomy. While we largely agree with Barton, we would like to reformulate his latter argument in terms of autocracy (as the nudge helps unhappy addicts *achieve* their goals).

if an intervention is not easily resistible, it ceases to qualify as a nudge. Given that cafeteria-like interventions on average change only 10–15% of consumption patterns (Arno & Thomas 2016), there is evidence that these are nudges, properly understood.

However, the critic can point to more effective interventions, such as the use of defaults. Consider the amount of people registered as potential organ donors, which ranges from 5 to 30% (opt-in) to 90% or more (opt-out) (Blumenthal-Barby and Burroughs 2012: 3). 'Surely,' critics argue, 'such interventions reduce (at least) some people's autocracy and autonomy'. 'Well, yes,' proponents argue, 'but that only shows that the intervention is not easily resistible and thus not a nudge proper (but a shove)'.

But even gentle pushes can sway our minds, critics can persist.[13] Imagine being a meat lover (or being wary of organ donation) but finding yourself picking the vegetarian dish (or not opting out). Now, cognitive dissonance reduction kicks in and you start rationalizing: 'I never thought of myself as a lover of vegetarian food (or an organ donor), but apparently it suits me. Obviously, this is not me being susceptible to marketing (or just being lazy or status quo biased); so it must be the case that I actually prefer vegetarian food (or organ donation).' While you had a settled preference, the nudge made you prefer something else (than you would have in the non-nudged environment). In this scenario (a refined version of case 5), nudges undermine people's autonomy as they change their (already made up) mind and thus inhibit them from *setting their own ends*. We provide two responses here.

First, we believe that the most plausible story here is simply that nudgees come to accept this change of mind. People can change (and come to acquire a preference for vegetarian options) and this is not necessarily at odds with autonomy. To insist, as critics would, that this is problematic, 'because the nudge made them do it' assumes a rather hefty type of false consciousness and second-guessing (akin to certain types of brainwashing). 'People might believe that they changed their minds and acted accordingly', these critics argue, 'but that is not true: their autonomy is an illusion created or induced by the nudgers.'

Instead of assuming such 'deep manipulation', we believe that this actually happens all the time: the changes that people (and their preferences) undergo seldom occur through processes they fully control. Claiming that this violates their autonomy implies that nearly all of our preferences and decisions are heteronomous. To see why such changes over time do not jeopardize autonomy, take again John Christman's (2001: 202) historical account of autonomy as non-alienation. If a person would have been aware of the origin of her motivating desire, Christman asks, would she feel alienated or not? So, if we would tell you that you were nudged to go vegetarian and then cognitive dissonance reduction kicked in, would you still accept your choices and preferences? We believe that most of you would. Remember all the times you walked into highly nudged environments like supermarkets: even now that you know full well that you have been influenced (and thus are aware of the origin of your preferences), you do not feel alienated from your decisions. (Note that, if you would feel alienated from your newly acquired preference, we agree that this would be problematic. Our claim here is that this hardly ever occurs in the case of nudging.)

[13] We thank an anonymous reviewer for this journal for being one of these critics, pushing us on this and pointing out the role that cognitive dissonance reduction mechanisms can play.

Second, we also have reason to question whether people who are so easily led by nudges actually have clear, settled preferences. Perhaps they qualified more as flexitarians than as meat lovers. This brings us to cases 6 and 7, in which people have not made up their mind yet and are thus undecided, conflicted, or indifferent.[14] These people, we believe, are most susceptible to being nudged successfully and thus having their decisions 'shaped' (Hausman & Welch 2010). Think of every time you walked into a cafeteria not yet sure what to eat. In such instances, you are quite 'nudgeable' because your preferences are still quite 'malleable'. Hence, you are likely to be influenced by the salient presentation of particular dishes. Would that undermine your autonomy and autocracy?

As you have not made up your mind yet, it is difficult to assess your autocracy, since there is no baseline condition or reference point. There is no way of knowing whether you succeed in achieving your ends, as it is unclear what those ends are (hence the question marks in our scheme). Or perhaps such nudges *do* respect your autocracy as they do not interfere with you satisfying whatever preference such nudges help form. Whether intentionally nudged or not, your preferences will be shaped accordingly and you will probably be happy with the end result either way (as you lacked a settled preference for or against whatever outcome).

But that raises worries about your autonomy. Our distinction between autonomy – ability to set your own goals – and autocracy – ability to realize your own goals – clearly reveals the crucial point of critics here. While vegetarian nudges have the benefit of supporting the autocracy of some (weak-willed vegetarians; case 2), they only do so at the expense of the autonomy of others (undecided customers; cases 6 and 7). Nudges shape at least undecided people's preferences, interfering with their process of forming their will and influencing them to set ends that are actually someone else's. Let us analyze the strength of this objection.

While nudge critics would say that nudges diminish these people's autonomy, proponents stress the inevitability of a-rational influences. Given that people lack a settled preference, their decision will largely depend on the choice environment, even if that is not intentionally designed. Absent the intentional nudge, people happened to choose the vegetarian option in case 6 and the meat option in case 7. Although nudges deliberately change the choice architecture, the causal mechanisms remain the same. If nudges are said to 'change people's (undecided) minds', they only change the outcome (switch from meat to vegetarian; case 7), not the choice process. Whenever there is no reflective preference-formation or decision-making process to respect, nudges cannot undermine, bypass or pervert that (even though this is what critics typically blame nudges for).[15] Both parties can be right here. While the factors involved in the decision-making process are the same, what distinguishes cases of deliberate nudging from non-nudged choice environments is that the influence exerted on nudgees is intentional. Two strands in the autonomy literature come apart here: autonomy as independence from others, and autonomy as control (you making your own decisions). Proponents of the latter could claim that these nudgees lack autonomy, with or without the nudge, as

[14] According to Schubert (2017: 337), we live in a "behavioral world" in which "individuals lack consistent and stable preferences". While Schubert overgeneralizes (people often have quite stable preferences), we do agree that nudging works best when people lack such clear preferences.

[15] While the alternative that critics have in mind is not this non-nudged scenario (in which the same 'blind processes' are at work) but rational persuasion, proponents argue that such persuasion is often ineffective.

they are not in control of the process of setting their ends (but instead let those depend on choice environments). Again, we believe that this forms an all too demanding conception of autonomy. Choices influenced by a-rational factors can be autonomous, as long as they remain within what we will call in section 4 'perimeters of autonomy'. And if non-nudged decisions are autonomous in this sense, then so are nudged decisions. After all, it is hard to see why the intentional aspect of nudging would undermine people's *ability* to set their own ends.[16]

Let us recapitulate. To assess the impact of nudging on people's autocracy and autonomy, we typically need a baseline, a counterfactual, non-nudged situation with which to compare the nudge intervention. If people have a settled preference, their autonomy is typically respected and the relevant question is that of *autocracy*, which can be respected or even promoted (cases 1–4). While this is where the debates between Sunstein and critics have focused on, the interesting cases in terms of *autonomy* are those where people have not yet made up their mind (cases 6–7). So far, we argued that these people's preferences are most likely to be shaped by a-rational factors, regardless whether they are nudged or subjected to random choice architectures. As such, we fail to see any compelling reason why nudged environments would undermine people's abilities to set their own ends any more than non-nudged environments.

The problem in this case is that, because people lack settled preferences, we cannot use those as a baseline or point of comparison here. In section 4, we develop a conception of autonomy that moves away from 'settled preferences' (which imply that people know specifically what option they prefer) towards people's care's, commitments and life plans. Even when people lack specific settled preferences, they care about certain things more generally and these cares provide the much needed point of comparison to assess their autonomy and autocracy. This move has the additional advantage that it allows for a-rationally induced but autonomous changes in preferences and decisions.

Now, if critics want to argue that nudges make people choose differently *and* makes them endorse or subscribe to these choices (refined version of case 5), we take this to be such a token of suspicion that it shifts the burden of proof. Such critics should present us with a theory of autonomy (presumably including a historical component) that shows how nudges fail the conditions of autonomous choice, without thereby raising the bar for autonomy to obviously excessive heights.

After all, the more stringent the conditions for autonomy (like requiring that preferences and choices result from reflective processes throughout), the more plausible the claim that nudges violate autonomy. The downside, however, is that this implies calling most day-to-day decisions – also those in the baseline, non-nudged condition – non-autonomous. In cases 1–4, we assumed autonomy in the sense of people having formed settled preferences, which provided a point of comparison to determine the impact of nudges. But if you stick to stringent conditions, the assumption of autonomy no longer holds. Regardless of whether people were or were not autonomous to begin with, nudges typically leave them 'as autonomous as they were before' (which under stringent conditions is not very autonomous at all).

[16] One could argue that by intentionally *and* repeatedly steering a person in one direction, one could gradually erode this ability. But, as indicated, we want to put the long-term effects of nudging aside for the purposes of this paper.

Critics could respond by explaining why we need genuinely autonomous choice in particular situations. High-risk decisions, for example, could require more reflection than low-risk decisions. Here, they would join Sunstein in his third approximation of autonomy as reflectively making choices that matter. Nudging specifically these decisions might then be a bad idea because they will fail to meet the stringent conditions for autonomy. While we are happy to concede that nudging may not be the best strategy in such circumstances, we believe the burden of proof rests on critics (1) to highlight the exceptional nature of those circumstances that require stringent autonomy conditions and (2) show why nudges specifically, in contrast to non-nudged choice environments, violate those conditions.

5 Perimeters of Autonomy

We have already emphasized the need for less stringent autonomy conditions that should enable us to maintain that most day-to-day decisions are autonomous and that instances of non-autonomy are rather exceptional. In general, there is a background assumption of self-determination when assessing the autonomy of people, their preferences and decisions. There are good reasons for any viable theory of autonomy to 'match' the idea that ordinary people in ordinary circumstances are capable of autonomous choice (Roessler 2017). For one thing, if autonomy is *normatively significant* – which it is when one says that autonomy should be *respected* – then we should not put the bar too high. Otherwise, my choice of partner, friend, job or hobby may not be an autonomous choice that deserves respect (or, to refer to a different context, we could no longer ascribe responsibility to evil-doers performing non-autonomous actions).

On our account of autonomy then, to be a self-determining individual means that you can be motivated to act[17] on the basis of your values, commitments and what you care about. It also means that you can consider and reconsider your values, commitments and what you care about, and thus revise your conception of the good life, in ways that are moderately responsive to reasons (Fischer & Ravizza 1998). On the basis of these pivotal elements of some of the most influential theories of autonomy (Dworkin 1970, 1988, Frankfurt 1988, Watson 1975), anyone would be hard-pressed to argue why and how nudges undermine these abilities.[18]

All of us have values and things we care about, and we try to lead our lives in light of those things. Even when we lack specific, settled preferences (e.g. about which dish to have or whether we want to be an organ donor), we do have things that we care about that constitute 'who we are'. These broader, more general cares are more fundamental (or higher order, if you like) and ground the more specific (first order) preferences that we form in specific circumstances when making everyday choices. Sometimes we form specific (first order) preferences quite deliberatively (when you know which specific dish to order at your favorite restaurant) but often we form those on the fly. While circumstances likely play a role in the latter case, our cares are relevant here too as they set boundaries or perimeters to what we genuinely prefer.

[17] To really act upon that basis would amount to autocracy.

[18] Some of these theories do exclude manipulation, like Dworkin's (1988). However, to argue for the violation of autonomy one would then have to argue that nudges are manipulative, which is not evident at all (Wilkinson 2013).

So imagine (re)considering your values and cares and realizing that you value animal life, oppose animal cruelty and care about the environment. Since it is you who sets your own ends and life plans, your autonomy is realized. You know by what desires and motives you want to be moved. Now, autocracy matters as well: ideally, your specific day-to-day decisions will reflect and align with your broader values and cares. But notice that you still have a plethora of choices available and that your cares set quite broad 'perimeters' or constraints. Supermarkets have plenty of options, a lot of which fall within your perimeters (vegetables, fruits, and in the future perhaps even lab-grown meat) while others obviously run contrary to your values and cares (T-bone steaks and lamb chops). Your autocracy is reduced whenever you fail to lead your life according to your cares.

Within the broad framework set by your cares, you also have other considerations, like personal taste and the preferences of those joining you for dinner. When you enter the supermarket, you want to navigate the space offered by these constraints. You want to remain within the perimeters of autonomy. You want to buy products that are in line with your values and cares, which typically do not single out one option, but provide you with some leeway. In sum, your autonomy sets perimeters and is thus perfectly consistent with multiple specific preferences and decisions.

Of course, supermarkets are heavily nudged choice environments, with companies using all sorts of marketing techniques to sell specific products. However, most of us, most of the time, manage to stay within the perimeters of our autonomy, and manage to avoid alienation (that is, being motivated by impulses and desires at odds with what we value and care about). Vegetarians may be successfully nudged during shopping but it is unlikely that they will be induced to buy lamp chops. Nudges, we argue, seldom lead people to end up outside of their perimeters of autonomy.

Our claim that what happens in supermarkets is actually 'okay' reflects the background assumption of autonomy and gives us a realistic idea of what it means to be a self-determining agent. Multiple conceptions in the autonomy literature are compatible with this, such as the abovementioned accounts of Amy Mullin (2007) and Harry Frankfurt (1988). Or take Michael Bratman's (2005) understanding of autonomy as planning agency. On this account, we position, and sometimes reposition the perimeters of our autonomy. Once the markers are set (and plans are made), we do not (have to) reflect upon that basic structure each and every waking moment and can simply remain within these perimeters in a more or less automatic way.

Now, what does this crude sketch of autonomy imply for the people lacking settled preferences prior to being nudged (cases 6 and 7)? Still undecided, they have no clear thoughts on any particular outcome. But surely, they do have some perimeters as caring and planning agents, even if those are not explicitly on their minds. They might only want to avoid eating meat. As they will make a specific choice eventually, the question whether they really wanted to have fruits or vegetables, becomes moot because both options are compatible with their autonomy. So even without a clear baseline in terms of fully settled and avowed preferences, we can still assess people's autonomy. Even when undecided, we all have our values and cares, which provide the foundation of our agency and set constraints on the things we autonomously can strive for.

When it comes to autonomy as the *ability* to set one's own ends, it is crucial to see that making up one's mind up (forming specific preferences) always happens in relation to a mind that is already made up (having broader cares and values). The only way to

assess the autonomy of someone without settled preferences, is by investigating that deeper system of cares or values. The point of reference when assessing people's autonomy and autocracy then is not one of their specific judgements or preferences (whether these be actual or hypothetical), but a broader compound of values and cares.[19] Although nudges can sometimes *change* our behavior (as we would have acted differently absent the nudge), this does not necessarily preclude autonomy, because we can act autonomously across a *range* of options, as long as we remain within our perimeters of autonomy. This understanding of autonomy then alleviates most of the worries concerning nudging techniques undermining autonomy.

6 Conclusion

By way of conclusion, let us stress three important points. First, one might object that these cares and values are not themselves autonomously chosen and thus provide a poor point of reference. We still do not know what people's real, autonomously chosen values are.[20] As this results in undiluted and generalized skepticism about any kind of autonomy, we will put this worry aside here. Here, we assume that we sometimes are, and sometimes are not, autonomous (and that autonomy conceptions should be able to account for this difference).

Secondly, and more importantly, our notion of 'perimeters of autonomy' does not deny that we can be more or less autonomous. One's motives for action can be more or less 'in line' with one's general system of values and cares (see also: Ekstrom 2005; Arpaly & Schroeder 1999). A particular decision can be backed-up by mutually supporting and reinforcing values and cares, or it could be supported by some but not other elements within that system. Some nudges play into specific vulnerabilities – for example that we care about our physical appearance – even if we believe that we should care less about those things. It would lead us too far off topic to further develop this, but nothing prevents us from understanding autonomy and heteronomy on a gradual scale.

This brings us to our third and perhaps most relevant point. On our account, nudges imply no straightforward violation of or impediment to people's autonomy. Or more carefully, we have argued that the burden of proof should shift to those who think that they do. They need to argue for autonomy being violated *without* merely appealing to the sheer mechanisms at play, and *without* raising the bar for autonomy to unrealistic standards, since that would make almost all decisions heteronomous. Now, defending people's autonomy under nudging does nothing to answer the manipulation objection. It might still be the case that nudges manipulate and that such manipulation is wrong (Wilkinson 2013). In our view, what happens in such cases is that manipulation *operates through* and *abuses* the autonomy of the victim. Manipulation then consists in the *exploitation* of the leeway offered by one's perimeters of autonomy. Someone might intentionally make use of the room offered by what you value or care about. In

[19] It also avoids Sugden's criticism to Sunstein as it does not require finding out what people's 'inner rational agents' hypothetically prefer but only what actual people's broadly conceived values and cares are.

[20] This relates to the problem of infinite regress and the so-called ab initio problem that haunts many theories of autonomy. For a discussion and solution, see Noggle (2005).

fact, this is what supermarkets typically do: they steer you to buy those things within your perimeters of autonomy that also promote their interests. While one may still worry about nudgers exploiting the leeway offered by the perimeters of our autonomy to steer us in whatever direction they like, this should not generally be understood in terms of autonomy.

This suggests two promising lines of research: (1) a more fine-grained analysis of the gradations of autonomy within its perimeters and how this is affected by nudging and/or instances of manipulation, and (2) an analysis of manipulation that shifts the focus from the autonomy-denying aspect it supposedly has on victims of manipulation, to the specific role and intentions of the manipulators.

Acknowledgements This paper was previously presented on different occasions and we thank the respective audiences for their helpful comments and feedback. In particular, we thank the organizers of the OZSW Conference at Utrecht University (November 2017), the 'Nudging and Moral Responsibility' workshop at VU Amsterdam (April 2018), the 'Applied Ethics of Nudging' workshop at the University of Stirling (September 2018) and the workshop 'Nudging in Public Health – And Beyond' at Aarhus University (November 2018). In addition, we thank three anonymous reviewers of this journal for their extensive and constructive comments. As always, responsibility for any remaining errors is ours.

References

Arno, A., and S. Thomas. 2016. The efficacy of nudge theory strategies in influencing adult dietary behaviour: A systematic review and meta-analysis. *BCC Public Health* 16: 676–787.

Arpaly, N., and T. Schroeder. 1999. Praise, blame, and the whole self. *Philosophical Studies* 93: 161–188.

Barton, A. 2013. How tobacco health warnings can Foster autonomy. *Public Health Ethics* 6 (2): 207–219.

Blumenthal-Barby, J.S. 2012. Between reason and coercion: Ethically permissible influence in health care and health policy contexts. *Kennedy Institute of Ethics Journal* 22 (4): 345–366.

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.

Blumenthal-Barby, J.S., and A.D. Naik. 2015. In defense of nudge-autonomy compatibility. *The American Journal of Bioethics* 15 (10): 45–47.

Bovens, L. (2009). 'The ethics of nudge', in: Till Grüne-Yanoff & Sven Ove Hansson (eds.), Preference Change: Approaches from Philosophy, Economics and Psychology. Berlin & New York: Springer: 207-219.

Bratman, M. 2005. Planning agency, autonomous agency. In *Personal autonomy: New essays on personal autonomy and its role in contemporary moral philosophy*, ed. J.S. Taylor, 33–57. Cambridge: Cambridge University Press.

Christman, J. 2001. Liberalism, autonomy, and self-transformation. *Social Theory and Practice* 27 (2): 185–206.

Dworkin, G. 1970. Acting freely. *Nous* 3: 367–383.

Dworkin, G. 1988. *The theory and practice of autonomy*. Cambridge: Cambridge University Press.

Dworkin, G. 2017. Autonomy. In *A companion to political philosophy*, ed. R. Goodin, P. Pettit, and Th. Pogge, 443–451. Malden MA: Blackwell Publishing.

Ekstrom, L. 2005. Autonomy and personal integration. In *Personal autonomy: New essays on personal autonomy and its role in contemporary moral philosophy*, ed. J.S. Taylor, 143–161. Cambridge: Cambridge University Press.

Engelen, B. 2019. Nudging and rationality: What is there to worry? *Rationality and Society* 31 (2): 204–232.

Fischer, J.M., and M. Ravizza. 1998. *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.

Frankfurt, H. 1988. *The importance of what we care about*. Cambridge: Cambridge University Press.

Furedi, F. (2011). Defending Moral Autonomy Against an Army of Nudgers. *Spiked*. January 11, 2011. Available online: https://www.spiked-online.com/2011/01/20/defending-moral-autonomy-against-an-army-of-nudgers/.
Hausman, D.M., and B. Welch. 2010. Debate: To nudge or not to nudge. *Journal of Political Philosophy* 18 (1): 123–136.
Kant, I. (1797/1996). The metaphysics of morals. In M.J. Gregor (trans. And Ed.) *Immanuel Kant: Practical Philosophy*, Cambridge: Cambridge University press.
Mills, C. 2015. The heteronomy of choice architecture. *Review of Philosophy and Psychology* 6: 495–509.
Mullainathan, S., and E. Shafir. 2013. *Scarcity: Why having too little matters so much*. New York: Times Books.
Mullin, A. 2007. Children, autonomy, and care. *Journal of Social Philosophy* 38: 536–553.
Nagatsu, M. 2015. Social nudges: Their mechanisms and justification. *Review of Philosophy and Psychology* 6 (3): 481–494.
Noggle, R. 1996. Manipulative actions: A conceptual and moral analysis. *American Philosophical Quarterly* 33 (1): 43–55.
Noggle, R. 2005. Autonomy and the paradox of self-creation: Infinite regresses, finite selves, and the limits of authenticity. In *Personal autonomy: New essays on personal autonomy and its role in contemporary moral philosophy*, ed. J.S. Taylor, 87–108. Cambridge: Cambridge University Press.
Noggle, R. (2018). The ethics of manipulation. *Stanford Encyclopedia of Philosophy*. Available online: https://plato.stanford.edu/entries/ethics-manipulation/.
Nys, T.R.V & Engelen, B. (2017). Judging nudging: Answering the manipulation objection. Political Studies, 65(1), 199–214.
Roessler, B. 2017. *Autonomie: Ein Versuch über das gelungene Leben*. Berlin: Suhrkamp Verlag.
Saghai, Y. 2013. Salvaging the concept of nudge. *Journal of Medical Ethics* 39 (8): 487–493.
Schmidt, A.T. 2019. Getting real on rationality – Behavioral science, nudging, and public policy. *Ethics* 129 (4): 511–543.
Schubert, C. 2017. Green nudges: Do they work? Are they ethical? *Ecological Economics* 132: 329–342.
Sugden, R. 2015. Looking for a psychology for the inner rational agent. *Social Theory and Practice* 41 (4): 579–598.
Sunstein, C.R. 2014a. Choosing not to choose. *Duke Law Journal* 64 (1): 1–51.
Sunstein, C.R. 2014b. *Why nudge: The politics of libertarian paternalism*. New Haven: Yale University Press.
Sunstein, C.R. 2015a. Nudges, agency, and abstraction: A reply to critics. *Review of Philosophy and Psychology* 6 (3): 511–529.
Sunstein, C.R. 2015b. Nudges do not undermine human agency. *Journal of Consumer Policy* 38 (3): 207–210.
Sunstein, C.R. 2015c. The ethics of nudging. *Yale Journal on Regulation* 32 (2): 413–450.
Sunstein, C.R. 2016. *The ethics of influence: Government in the age of behavioral science*. New York, NY: Cambridge University Press.
Sunstein, C.R., and R.H. Thaler. 2003. Libertarian paternalism is not an oxymoron. *University of Chicago Law Review* 70 (4): 1159–1202.
Thaler, R.H., and C.R. Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. New Haven (CT): Yale University Press.
Vugts, A., M. Van Den Hoven, E. De Vet, and M. Verweij. 2018. How autonomy is understood in discussions on the ethics of nudging. *Behavioural Public Policy*: 1–16.
Watson, G. 1975. Free agency. *Journal of Philosophy* 72: 205–220.
White, M.D. 2013. *The manipulation of choice: Ethics and libertarian paternalism*. New York: Palgrave Macmillan.
Wilkinson, T.M. 2013. Nudging and manipulation. *Political Studies* 61 (2): 341–355.