COLOGNE GRADUATE SCHOOL
in Management, Economics and Social Sciences

Vol 5 | No 2

CGS WORKING PAPER

Is Soft Paternalism Ethically Legitimate? – The Relevance of Psychological Processes for the Assessment of Nudge-Based Policies

Mira Fischer (University of Cologne)
Sebastian Lotz (Stanford University)

Cologne Graduate School
in Management, Economics
and Social Sciences
Albertus-Magnus-Platz
50923 Köln
www.cgs.uni-koeln.de

University of Cologne

Is Soft Paternalism Ethically Legitimate? – The Relevance of Psychological Processes for the Assessment of Nudge-Based Policies[1]

Mira Fischer and Sebastian Lotz[2]

First version: March 23, 2012

This version: May 9, 2014

Comments welcome!

Abstract

In this article we develop a taxonomy of behavioral policy measures proposed by Thaler and Sunstein (2008). Based on this taxonomy, we discuss the ethical legitimacy of these measures. First, we explain two common reservations against nudges (choice architecture) rooted in utilitarian and Kantian ethics. In addition to wellbeing, we identify freedom of action and freedom of will (autonomy) as relevant ethical criteria. Then, using practical examples, we develop a taxonomy that classifies nudges according to the psychological mechanisms they use and separately discuss the legitimacy of several types of behavioral policy measures. We hope to thereby make a valuable contribution to the debate on the ethical legitimacy of behavioral policy making.

1. Introduction

Behavioral economics is the new paradigm in consumer protection, health policy, and social policy, which allows it to have substantial influence on the legislative process (*Amir* et al. 2005; *Camerer* et al. 2003). This influence stems from a series of empirical findings on regularities in human behavior (*Benartzi/Thaler* 2013; *Camerer* 2003; *Fehr/Gächter* 2002; *Tversky/Kahneman* 1981; *Smith* 1965). Key results from behavior research have helped to better understand phenomena such as tax evasion, tax compliance, or willingness to engage in environmental protection and, on that basis, build more effective institutions.

The British government, for example, maintains the "Behavioral Insights Team," also known as the "nudge unit," which is responsible for a number of policies that are intended to make citizens behave more rational. The most prominent advisor to the Behavioral Insights Team is *Richard Thaler*, one of the two authors of the book *Nudge – Improving Decisions*

[1] An earlier version of this article was published in German in Sozialer Fortschritt/German Review of Social Policy, 3/2014

[2] Contact: mira.fischer@uni-koeln.de, sebastian.lotz@stanford.edu

We thank participants at the 11th TIBER Symposium on Psychology and Economics and the ESA European Conference for helpful comments.

about Health, Wealth, and Happiness (2008). His co-author, *Cass Sunstein*, advised the Obama Administration from 2009 to 2012 and has been described by mainstream media as "Obama's superego"[3] or "an intellectual mentor to President Obama."[4] There has been lively debate among experts about policy rooted in behavioral economics; however, skepticism seems to run deeper in Germany than in Great Britain or the United States, where *nudges* are currently widely employed. This debate features numerous arguments for and against the "soft paternalism" presented by *Thaler* and *Sunstein* (2008).

The goal of this paper is to develop a taxonomy that classifies policies proposed by Thaler and Sunstein (2008) according to the psychological mechanisms they exploit, thereby offering political decision makers and other parties with an interest in the subject a framework for ethical assessment of these policies. Two main observations drive our motivation for this paper. The first observation is that different authors either accept or reject nudges, while too little attention is paid to achieving a nuanced view. This gave us the impression that the term nudge, which is used to represent a collection of diverse policy proposals, is conceptually vague and possibly not fully coherent. We therefore propose that at least a few different types of nudges should be individually analyzed as part of a normative discourse – possibly with varying outcomes.

The second observation is that in the public debate on the desirability of policy measures critics' ethical assumptions have remained largely implicit. At times political and media actors assume that consensus exists on the conception of man, in particular on how people make decisions, as well as the relative value to be placed on various (sometimes conflicting) positive and negative liberties. It is often overlooked that "freedom" and "well-being" are not descriptive but rather heavily contested normative terms. In our view, the ambiguity surrounding these terms and the political contest about their interpretation contribute to the misunderstandings and lack of clarity found in the present debate on nudges.

Scientists and political decision makers have certainly understood that they *can* change behavior through the use of nudges (choice architecture). Next comes the question of whether they *should* do so. Welfare and freedom, two core values of our society that are often pitted against each other occupy key roles in this assessment. With the introduction of a

[3] http://features.blogs.fortune.cnn.com/2013/02/22/cass-sunstein-simpler/
[4] http://articles.washingtonpost.com/2012-08-03/news/35492512_1_top-obama-adviser-president-obama-white-house

taxonomy of nudges we hope to make the ethical discourse surrounding them more prolific and clear.

2. The Normative Dimension of Nudge

Thaler and *Sunstein* define nudges as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid." (*Thaler/Sunstein* 2008, p. 6). Nudges promise to make people make more sensible decisions without restricting their freedom. They are therefore often equated with the terms "soft paternalism" or "libertarian paternalism."

Although nudges have been the subject of public debate internationally and some political leaders have adopted policies based on behavioral insights, normative problems associated with them have yet to be thoroughly discussed. Interdisciplinary discourse on the legitimacy of shaping of behavior by exploiting constrained rationality is quite recent (*Bovens* 2009; *Hausmann/Welsh* 2010; *Frerichs* 2011; *Axtell-Thomson* 2012; *Blumenthal-Barby/Burroughs* 2012; *Selinger/Whyte* 2012; *Schnellenbach* 2012; *Fischer/Lotz* 2013). Problems concerning people's autonomous decisions when acting as voters (Schumpeter 1942), patients (Cohen 2013), and consumers (Schwan 2009) have long been discussed within various specialized fields. However, the political sphere's strong interest in and targeted use of findings on the limitations of rationality is new, which lends the topic renewed controversy. In German-speaking countries, it is especially *Held* et al. (2013) who are contributing to this debate (see also articles from *Kirchgässner*, *Güth/Kliemt*, *Ott* and *Witt/Schubert* in this volume).

2.1 The Two Primary Objections to Nudges: Costs and Manipulation

The recent debate has brought together experts from a number of different areas of expertise, sometimes with widely varying perspectives on the ethical concerns surrounding nudges. For example, *Schnellenbach* (2012), an economist, regards nudges as problematic because psychological barriers are nothing more than a different form of costs. He argues that, although prohibitions and gentle nudges vary in the amount and type of costs incurred, both lead to behavior desired by the state. Therefore Schnellenbach sees nudges as being similar to other economic policy instruments, such as taxes, which impact an individual's cost-benefit analysis and thereby distort decision-making.

Philosophers *Hausman* and *Welsh* (2010) argued that nudges are harmful for other reasons. They share *Thaler* and *Sunstein's* (2008) opinion that, for example, the establishment of default settings, does not restrict freedom of action or reduce individual utility. They argue, however, that autonomy, meaning the deliberative process that brings about an individual's own preferences, is infringed. In their view it is problematic that decisions thus reflect the tactics employed by the "decision architects" because the degree of control over one's own mental processes would be reduced, which leads to preferences that in a certain sense are not one's own. Here the problem is not regarded as restricting a person's freedom to act, as argued by *Schnellenbach* (2012), but rather as restricting freedom of will.

On the basis of these examples, the criticism of nudges can be roughly divided into two groups. On the one hand, nudges have been criticized for attempting to increase people's long-term utility at the expense of their short-term utility by encouraging behavior that either (supposedly) benefits them over the long run or promotes a goal that is seen as positive and supported by the state but not the affected (present) self. Second, nudges have been criticized for manipulating people's preferences by interfering with their autonomy, i.e. the ability to form behavior-shaping preferences through independent deliberation of reasons. The first objection is thus rooted in utilitarianism and directed at the goals that are pursued with the help of nudges. It assumes that an individual always behaves such that, given her preferences and her discount factor, she maximizes her utility and nudges reduce this utility. The second objection is related to the principles of Kantian ethics and directed at the mechanisms employed by nudges. It holds that an individual's autonomy, which is also the basis of her dignity as a person, is worthy of protection and should not be interfered with by a nudge.

2.2. Conception of Man: Implicit Assumptions about Freedom, Intentionality, and Utility Maximization

The criticism that nudges curtail individuals' freedom of action by adjusting the non-monetary costs of alternatives, as well as the criticism that they interfere with the deliberative process, rest on the assumption that prior to being nudged an individual enjoys freedom of action and freedom of will, and that intervention by the state curtails both of these. Both points of criticism are therefore based on a negative understanding of freedom, meaning the right to non-interference by others.[5] As Benjamin Barber (2011), former advisor to Bill Clinton and Roman Herzog, writes, the meaning of freedom is based on "theories about our

[5]This conception of freedom has been criticized by, among others, *Sen* (1989) and *Nussbaum* (2000) who stress the importance of "capabilities" that enable individuals to actually apply theoretical possibilities. See also *Nussbaum* and *Sen* (1993).

world, and what we want from and for that world" and the people in it. Our understanding of freedom also rests upon whether we view people "as "social beings", embedded in relationships, or "natural solitaries" born alone. Freedom can be seen as something to be preserved *against* social and political relationships, or something to be achieved *through* them." The debate on the legitimacy of nudges is therefore always also a debate on the conception of man, which encompasses beliefs about freedom, decision-making processes, and preferences.

If one follows the standard assumptions of neoclassical economics that an individual's utility function is defined for all possible states of the world, allowing him to rank states with respect to their individual desirableness, then nudges must be viewed as a manipulation of the costs associated with various choices. However if one accepts that the utility function has gaps – that is, there are certain things to which we have not (yet) assigned a preference rank, possibly because we do not know that they exist, because we cannot imagine them, or because we do not regard them as important enough to justify spending our limited mental capacity on them, then a nudge can shape behavior without diminishing utility. In this case, a nudge would fundamentally differ from a prohibition or tax, both of which result in a reduction of utility when affecting an individual. Nevertheless, there are many situations where we view taxes and prohibitions as justified. What then has caused the debate around nudges to become so controversial? This heated debate appears to be fueled by considerations that are not solely rooted in utilitarianism, but instead draw upon the question of how individual preferences are formed, or should be formed. Neoclassical economics fails to provide an answer to either question, as it assumes that preferences are given. The economic literature on endogenous preferences, however, has dealt with these issues for many years (cf. *von Weizsäcker*, 2002).

Assumptions about the mental processes at work in a given situation and how a nudge alters them are decisive for whether one regards specific interventions propagated by Thaler and Sunstein as ethically justified (also see *Crusius* et al. 2012 for a discussion of the importance of information processing procedures in economic behavior). A behavioral description of change in behavior brought about by a nudge, as can be found widely in the discussion, is thus not suitable for assessing its ethical legitimacy as this description does not allow for any conclusive answer to the question of whether the nudged person has no preference, is indifferent, has a preference for one of two alternatives and the nudge changes one's calculus, or whether the nudge induces one to unconsciously behave in a different manner.

Criticisms rooted in utilitarianism and Kantian ethics first of all rest on the assumption that behavior shaped by a nudge is intentional. Philosopher and former German culture minister Julian Nida-Rümelin (2005) explains in his book "Über menschliche Freiheit" (On Human Freedom) that in order for behavior to count as an action it must to be accompanied by intentionality, and that this intentionality must to be guided by "adequately" balanced reasons in order for the action to count as rational. The fact that people fail to save for the future or do things that have a negative impact on their future health does not in itself indicate the prevalence of irrationality. People can intentionally act in a "negative" way and have no regrets when later confronted with the consequences. These people may have a strong preference for present consumption over future consumption and perhaps are willing to lead a Spartan life in their old age as long as this allows them to enjoy life in the present. Therefore we cannot determine only by observation that behavior represents an action and that this action is irrational, because an action is irrational when it contradicts individual goals and values, both of which we learn nothing about by mere observation. A description of behavioral change brought about by a nudge can indeed shed light on the effectiveness of nudges, but not on the psychological mechanisms at work. A normative assessment of policy proposals is only possible when making assumptions about mental processes. We illustrate this using examples from *Thaler* and *Sunstein* (2008).

2.3 Examples from the Nudge Book

The "default rule" for retirement schemes is one of the most famous examples of nudge-based economic policy (*Beshears* et al. 2009). It rests on the observation that not only active decisions have consequences, but the failure to make decisions does as well. If by default no part of one's income is put in a retirement savings program, and a person does not actively decide in favor of private retirement savings, she will not have savings in old age (just like a person who actively decided against it), even if she has never given thought to the issue. Nudge-based economic policy attempts to, among other things, arrange the "default option" such that taking no action leads to "optimal" results. In this case, optimal is understood with respect to the long-term benefit of the affected person or the government's goal of avoiding poverty among pensioners. For example, a default rule could be arranged such that taking no action would cause 5% of a worker's salary to be diverted into a pension fund. Thus the "default" would be participation in the retirement scheme. As long as a person does not actively decide against a private retirement scheme, she will participate in the program. It has been repeatedly shown that selecting this type of "default option" increases

the likelihood of participation in such schemes (*Beshears* et al. 2009; *Bernartzi/Thaler* 2013). The effectiveness of such default options is also evident in other areas, such as willingness to become an organ donor (*Johnson/Goldstein* 2003).

In the field of health policy, campaigns designed to reduce risky behavior by subliminally addressing people's fears and insecurities have been termed "nudges" as well. A clear example of this can be found in smoking cessation campaigns, which sometimes use billboards showing photos of smokers' lungs to make possible negative consequences of smoking salient. In contrast to tobacco taxes or smoking bans, such campaigns make smoking less appealing without adjusting economic incentives.

The selective positioning of food in lunch cafeterias to prevent obesity represents another idea based on nudges. Research has demonstrated that people change their behavior based on the arrangement of individual dishes offered in a buffet (*Rozin* et al. 2011). Marginally less accessible placement of calorie-rich foods, for example, in the second row of a buffet, was sufficient for significantly reducing overall calorie intake. The form of the serving utensils similarly impacted calorie intake. If the "productivity" of a serving utensil was lower, the size of the portion chosen by the individual, and thus calorie intake, was also reduced.

A further example can be found in transportation policy. Lake Shore Drive in Chicago was infamous for a dangerous curve that had a high demonstrated accident risk. To create a nudge mechanism, lane markings on the road surface were painted such that the gaps between them became constantly narrower, creating the optical illusion of increasing speed. This led drivers to unconsciously reduce their speed, lowering the number of accidents by 36%.

3. A Taxonomy of Nudges according to Psychological Mechanisms

We propose a taxonomy of nudges based on several distinctions: intentional and unintentional behavior, utility and probability components of expected utility, and monetary and non-monetary costs. We distinguish between nudges that address people's intentional behavior as utility maximizers and nudges that address unintentional (automatic) behavior. The choice of a utility maximizer can be influenced by changing either utilities derived from different possible outcomes of a choice option or by changing the subjective probability of this outcome occurring. Furthermore, when a policy changes the utility a person derives from an outcome, in order for it to count as a nudge the change must not be in monetary terms, as

stipulated by Thaler and Sunstein (2008). What remains is the possibility of altering of non-monetary utility derived from an outcome.

We assume that an individual prefers action A to action B ($A \succ B$) if the expected utility from A is greater than the expected utility from B ($U_A > U_B$). In its most basic form, when assuming the additivity of utility, expected utility equals the sum of the weighted utilities of all possible outcomes of an action, whereby the respective weighting assigned to the outcomes corresponds to the probability that they will indeed result from the action. The utility derived from action A, for which for simplicity two probable outcomes are assumed, can be rendered schematically by $U_A = \pi_{A1}(u_{A1M} + u_{A1N}) + \pi_{A2}(u_{A2M} + u_{A2N})$, where π_{A1} represents the probability that outcome 1 occurs, while u_{A1M} represents the monetary utility of outcome 1 and u_{A1N} the non-monetary utility of outcome 1. This applies analogously for outcome 2 resulting from action A, and for the utility derived from action B U_B.

It is thus apparent that within this framework a person's behavior can be changed by influencing utility calculations without changing monetary utility. There are many different ways of accomplishing this. We term nudge Type 1 (discomfort nudge) policies that impact expected utility by changing the non-monetary (psychological or social) utility of outcomes. Strictly speaking, the differences between Type 1 nudges and traditional economic incentives can be seen in the fact that the nudge manipulates the non-monetary utility of a choice consequence, rather than its monetary utility, and does so only "marginally". Default settings on electronic devices, communication of social norms, and the activation of social norms through framing all fall into this category. All other forms of nudges generally function in a different manner from economic incentives.

We term Type 2 (probability nudge) interventions that impact expected utility by changing the subjective probability that certain outcomes are realized. This type of nudge works for example by reminding individuals of worst-case or best-case scenarios to raise the salience of certain possible outcomes, or by providing feedback that makes apparent the hidden consequences of an action. Informational campaigns represent another application of a Type 2 nudge, in which individuals are informed about the possible consequences of an action and the probability that they will suffer these consequences, which are intended to bring the subjective probabilities closer to the objective probabilities.

These two types, however, do not cover all policies that Thaler and Sunstein designate as nudges. What happens, for example, in the cafeteria case, in which a change in the order

that dishes are presented in leads to healthier eating behavior? It can be plausibly assumed that many people have a preference to eat *something* when they go to the cafeteria at lunchtime, but have not formed a detailed preference about what *exactly* they want to eat. It is possible that they simply have more important things to think about, or maybe they are aware of the options but are indifferent. In both cases, people see something and realize in that moment that they want it. If they were not to see the object, or to see something else first, then they would not have a preference for it. One can think of many examples where this appropriately describes human behavior. Therefore, we term Type 3 (indifference nudges) measures that exploit gaps in the utility function or individuals' indifference when choosing between alternatives to induce people to form specific and predictable ad-hoc preferences.

Furthermore, Thaler and Sunstein describe how people sometimes unconsciously do things that have consequences they would rather avoid. They note that nudges can correct automatisms, such as in the example of the road markings that lead to a lower accident risk because drivers are less likely to underestimate their speed. Type 4 nudges (automatism nudges) are those which use the absence of intentionality to control unintended behavior. In this respect, Types 2 and 4 share the common goal of attempting to prevent unintended consequences. The difference between the two types lies in that fact that with Type 2 there is a (partially) conscious processing of information that contributes to a utility calculation, whereas with Type 4 the processing of information occurs on an unconscious level and directly affects behavior.

4. An Ethical Re-Evaluation of Nudges: Freedom of Action and Autonomy as Relevant Criteria

Type 1 nudges (discomfort nudges), which reduce the non-monetary (psychological or social) utility of certain options in order to push people toward other options, cause a reduction in present utility in favor of future utility. However, individuals presumably behave rationally, given their discount favor, and a nudge favoring future utility would leave them worse off (cf. *Schnellenbach* 2012). From a utilitarian perspective, this type of nudge could only be endorsed if one assumes that hyperbolic utility discounting is a widespread problem, which in turn would lead to the oft-studied behavioral science problems of lack of self-control (weakness of will) and subsequent regret of earlier behavior. From the perspective of Kantian ethics, a Type 1 nudge is not problematic as long as it is understood as a "behavioral economic incentive" that rational actors can consciously react to. However, the relevant

question here concerns to what extent the activation of social norms represents a conscious process, and to what extent it remains unconscious.

Type 2 nudges (probability nudges), which change the individuals' expected utility by manipulating the subjective probability that an action has certain consequences, at first sight appear to not be very problematic. From a utilitarian perspective, it should be noted here that an individual can only maximize his utility given his preference if the subjective probability that certain consequences will occur matches the actual probability. There are many indications that people systematically underestimate certain risks and that, when this is the case, a nudge that increases awareness of certain scenarios can increase an individual's wellbeing and freedom of action. As political scientist Robert Goodin writes in his article "The Ethics of Smoking", people overestimate the risk of dying in spectacular fashion, for example in an automobile accident, while they severely underestimate the risk of dying in a more banal manner, such as from the effects of smoking. We are particularly inclined to intervene in situations in which false beliefs leading to disastrous consequences can be traced back to "well-known forms of cognitive defects" (*Goodin* 1989). With an eye to autonomy, we should assess the extent to which Type 2 nudges interfere in the formation of preferences. It can be concluded that Type 2 nudges, by changing the assessment of causal relationships, leave people's values untouched and merely adjust their instrumental preferences. A person's preference X is an instrumental preference if she only wants X because she wants to use it to acquire Y. If because of a nudge she reaches the conclusion that she is much more likely to get Y, through Z than through X, then she will prefer Z instead of X. If this were the only effect of Type 2 nudges, then they would be harmless. However, in practice the utilitarian objection can be raised that a nudge that leaves one person in a more advantageous situation can also be a disadvantage for another person. For example, perhaps the person estimates risks realistically before being nudged and the nudge leads her to overestimate risks (see Schnellenbach 2012). This problem arises because nudges typically cannot be directed at specific people but rather can only follow the "watering can principle", i.e. address the average behavior of a large group of people.

It is immediately clear that Type 3 nudges (indifference nudges), which exploit gaps in the utility function or indifference between alternatives, cannot be easily criticized from a utilitarian perspective. If an option is made attractive for a person who does not have any existing preferences, then that individual's utility is not thereby reduced. An assessment of a policy's impact on an individual's utility requires that there is an ex ante utility that can be

used as a point of comparison, which does not exist in the absence of ex ante or stable preferences (*Bykvist* 2010). It can be argued, that preferences formed on the basis of a nudge are "manipulated" preferences. Speaking about "manipulated" preferences only makes sense if they can be contrasted with "original" or "authentic" preferences. However, it is not conceivable in what sense people, who have been influenced over their entire lives by other people and are members of society, possess "original" or "authentic" preferences. *Schumpeter* (1942) was one of the first to address this problem. From a utilitarian perspective, the same objection as described above applies. For example, eating a calorie-rich dessert on occasion might benefit an underweight person. The "favorable" positioning of the salad might induce an underweight person into behavior that is personally harmful over the long term. This ultimately occurs because obesity represents, on average, a societal problem that is in turn addressed by nudges.

If nothing is intended by a certain behavior, then this behavior is not based on preferences or a utility calculation. In this case, any increase in an individual's personal utility resulting from a change in behavior caused by a Type 4 (automatism) nudge does not occur at the expense of short-term utility. In the absence of intentionality, the claim that a nudge infringes on a person's autonomy appears unfounded. For example, if someone unintentionally urinates next to a urinal and after the application of a fly-shaped sticker unconsciously aims more accurately (an example from *Thaler* and *Sunstein*, 2008), it is not apparent how this restricts an individual's utility or autonomy since, in this case, the individual did not make use of his ability to consider relevant factors and behave accordingly. This, of course, does not mean that the individual lacks the ability to behave rationally. Changing unconscious, "automatic" behavior (*Bargh/Chartrand* 1999) is harmless both with respect to freedom of action and autonomy, since it is impossible to restrict something that is not used. It follows that Type 4 nudges, assuming freedom of action and autonomy are the only relevant ethical criteria, should be regarded as harmless.

We would like to note that the four suggested nudge categories represent ideal types. In reality, many policies employ more than one of the mechanisms described in this paper. Our classification of the examples taken from the nudge book into the four categories can be easily challenged given that we merely illustrate possible mechanisms and have not empirically analyzed the mental processes actually at work. Nevertheless, we hope that the proposed taxonomy makes a valuable contribution to the debate about nudges and their ethical evaluation, because this article, for the first time (to our knowledge), discusses the relevance

of different psychological mechanisms when assessing the ethical legitimacy of nudge-based policies.

5. Summary and Outlook

The goal of this article is to make the debate on the ethical legitimacy of behavioral policy making more prolific. At the outset, two common reservations (utilitarian vs. Kantian) about nudges were explained. A case distinction was developed using practical examples. The article argues that nudges differ with respect to the type of psychological mechanism employed and that, accordingly, different types of nudges should be individually subjected to ethical evaluation. In addition to wellbeing, freedom of action and freedom of will (autonomy) are named as relevant criteria. Extending the scope to psychological processes in turn makes a debate about the concept of man necessary.

A fundamental problem pertaining not only to nudges, but any action undertaken by the state, is that the state requires an implicit consensus for the efforts it elects to pursue. Nudges come under particularly heavy criticism from this perspective because they attempt to intervene in an individual's behavior, which primarily affects the individual in question and only to a lesser extent society as a whole. The existence of such a consensus on government activity can most credibly be assumed for primary goods. *Rawls* (1971) defines a primary good as something that one seeks, regardless of what the person otherwise seeks. As *Robert Goodin* (1989) writes, health is an example of such a primary good. It is also precondition for wellbeing and freedom. No one prefers sickness over health, and health is a prerequisite for existing in the future and thus having the capability to fulfill one's idiosyncratic desires. Health as the desired outcome of nudges is therefore more justifiable than other aims. In his work "On Liberty", the utilitarian *Mill* (1789/1975) presented a similar balancing of present and future utility, arguing that it is unacceptable for a person to sell himself into slavery because that single act would result in an individual being deprived of freedom in the future. *Dworkin* (1972, cited from *Sneddon* 2001) uses this example to justify a minimal level of paternalism, provided that it is employed to protect an individual's freedom.

Addressing the question of whether it is acceptable to employ nudges that primarily protect people from themselves, it should be noted that many public policy measures, including those that are binding, for example, seat belt or helmet laws, can hardly be justified by arguing that they mainly serve to avoid harm to others. If freedom is understood purely in

the negative, then nudges can in some cases expand freedom by securing individuals' ability to act freely in the future.

Like freedom of action, autonomy relies on the fulfilment of certain conditions. Autonomy is also something that can be constricted by environmental surroundings. Considering that marketing has for years used findings in psychology and behavioral economics to nudge people toward consumption – with negative consequences for peoples' health and financial situation (*Flegal* et al. 2010; Robert Koch Institut 2013; WHO 2011) – we should now discuss whether and to what extent the state should apply the same methods to mitigate the resulting harm to individuals and society at large. Compared with marketing, action taken by the state must conform to more robust ethical guidelines. While corporations can view people as irrational consumers and design their activities accordingly, the state is required to treat people as rational and responsible citizens or risk losing its democratic legitimacy. This conflict between the state's aims of preserving individual choice and avoiding harm often arises in debates on consumer protection, in which effective policies are not easily reconcilable with the conception of man as a rational actor.

It should be considered whether in many cases the desired outcome can also be achieved through government rule setting or other interventions, such as the introduction of economic incentives to shape corporations' behavior and informational campaigns that aim at strengthening citizens' capabilities of rational decision making instead of exploiting their irrationality. Nudge policies have been generally criticized for tinkering with symptoms, such as potentially harmful behavior regarding health or money, rather than addressing their economic and social causes (*Frerichs* 2011). For example, developmental psychology has shown that high-risk behavior by disadvantaged youths can be understood as an adaption to a strained and hopeless environment. On an individual level, this behavior is not necessarily dysfunctional or irrational, but rather can improve an individual's image among peers (*Nell* 2002; *Ellis* et al. 2012), even if the behavior is problematic for the wider society.

As stated by *Immanuel Kant* (1785/2007), an individual's dignity is respected when he is perceived as a moral and rational actor who behaves according to rules and obligations that he has set for himself. However, we are not always rational actors, which means that we could occasionally use a nudge. Here it is always important that the nudge operates within the individual's theory of the good (*Goodin* 1989), in other words, that the individual is respected as a normative instance. Nudges should only influence peoples' instrumental preferences, not their values. A careful approach to nudges, adhering strictly to the level of transparency

required in a democratic society, is needed to ensure that the outcomes pursued through nudges truly represent a societal consensus, and that nudges are not simply introduced as a technocratic short cut. Every effective policy measure can be used for harmful purposes; however, our intuition tells us that nudges are particularly prone to abuse because they are very effective and represent, because of the subtle manner in which they affect behavior, a political tool that is more difficult to control.

References

Amir, O./Ariely, D./Cooke, A./Dunning, D./Epley, N. et al. (2005): Psychology, behavioral economics, and public policy, Marketing Letters, 16(3-4), 443-454.

Axtell-Thompson, L. (2012): Nudge ethics for health plans, The American Journal of Bioethics, 12(2), 24-25.

Balz, J. (2010): Measuring the LSD effect: 36 percent improvement, Nudge Blog, http://nudges.org/2010/01/11/measuring-the-lsd-effect-36-percent-improvement/. Letzter Zugriff: 28.03.1013.

Barber, B. (2011): Diskussionsbeitrag zur Reihe Debating first principles: "Freedom from" vs. "Freedom to" auf den Seiten des Radiosenders WNYC vom 01.04.2011, http://www.wnyc.org/articles/its-free-country/2011/apr/01/debating-first-principles-freedom-vs-freedom/. Letzter Zugriff: 30.03.2013.

Bargh, J./*Chartrand, T. L.* (1999): The unbearable automaticity of being, American Psychologist, 54(7), 462-479.

Benartzi, S./*Thaler, R. H.* (2013): Behavioral economics and the retirement savings crisis, Science, 339(6124), 1152-1153.

Beshears, J./Choi, J. J./Laibson, D./*Madrian, B. C.* (2009): The importance of default options for retirement saving outcomes. Evidence from the United States, in: Brown J. R., Liebman J. B., & Wise D. A. (eds.), Social security policy in a changing environment, 167-195, Chicago: University of Chicago Press.

Blumenthal-Barby, J. S./Burroughs, H. (2012): Seeking better health care outcomes. The ethics of using the "Nudge", The American Journal of Bioethics, 12(2), 1-10.

Bolderdijk, J. W./Steg, L./Geller, E. S./Lehman, P. K./Postmes, T. (2013): Comparing the effectiveness of moral versus monetary motives in environmental campaigning, Nature Climate Change, 3, 413-416.

Bovens, L. (2009): The ethics of Nudge, in: Grüne-Yanoff, T./Hansson, S.-O. (eds.), Preference Change. Approaches from Philosophy, Economics and Psychology, The Decision Library, 42, Dordrecht: Springer.

Bykvist, K. (2010): Can unstable preferences provide a stable standard of well-being?, Economics and Philosophy, 26(1), 1-26.

Camerer, C. (2003): Behavioral game theory. Experiments in strategic interaction, Princeton: Princeton University Press.

Camerer, C./Issacharoff, S./Loewenstein, G./O'Donoghue, T./Rabin, M. (2003): Regulation for conservatives. Behavioral economics and the case for "asymmetric paternalism", University of Pennsylvania Law Review, 151(3), 1211-1254.

Cohen, S. (2013): Nudging and informed consent, The American Journal of Bioethics, 13(6), 3-11.

Crusius, J./van Horen, F./Mussweiler, T. (2012): Why process matters. A social cognition perspective on economic behavior, Journal of Economic Psychology, 33(3), 677-685.

Dworkin, G. (1972): Paternalism, The Monist, 56, 64-84.

Ellis, B. J./Del Giudice, M./Dishion, T. J./Figueredo, A. J./Gray, P/Griskevicius, V./Hawley, P. H./Jacobs, W. J./James, J./Volk, A. A./ Wilson, D. S. (2012) The evolutionary basis of risky adolescent behavior. Implications for science, policy, and practice, Developmental Psychology, 48(3), 598-623.

Fehr, E./Gächter, S. (2002): Altruistic punishment in humans, Nature, 415(6868), 137-140.

Fischer, M./Lotz, S. (2013): Yes, we can! But should we? An ethical analysis of soft paternalism, Working Paper: University of Cologne.

Flegal, K. M./Carroll, M. D./Ogden, C. L./Curtin, L. R. (2010): Prevalence and trends in obesity among US adults, 1999-2008, JAMA: The Journal of the American Medical Association, 303(3), 235-241.

Frerichs S. (2011): False promises? A sociological critique of the behavioural turn in law and economics, Journal of Consumer Policy, 34(3), 289-314.

Goodin, R. E. (1989): The ethics of smoking, Ethics, 99(3), 574-624.

Hausman, D. M./Welch, B. (2010): Debate: To Nudge or Not to Nudge, Journal of Political Philosophy, 18(1), 123-136.

Held, M./Kubon-Gilke, G./Sturn, R. (Hg.) (2013): Grenzen der Konsumentensouveränität, Jahrbuch Normative und institutionelle Grundlagen der Ökonomik, 12, Marburg: Metropolis.

Johnson, E. J./Goldstein, D. (2003): Do defaults save lives? Science, 302(5649), 1338-1339.

Kant, I. (2007): Grundlegung zur Metaphysik der Sitten. Kommentar von Christoph Horn, Corinna Mieth und Nico Scarano, Frankfurt a. M.: Suhrkamp

Mill, J. S. (1975): On Liberty, in: Wollheim R. (ed.), Three Essays, Oxford: Oxford University Press.

Nell, V. (2002): Why young men drive dangerously: Implications for injury prevention, Current Directions in Psychological Science, 11(75), 75-79.

Nida-Rümelin, J. (2005): Über menschliche Freiheit, Stuttgart: Reclam Verlag.

Nussbaum, M. C./Sen A. (eds.) (1993): The Quality of Life, Oxford: Clarendon Press.

Nussbaum, M. C. (2000): Women and Human Development. The Capabilities Approach, Cambridge: Cambridge University Press.

Rawls, J. (1971): A Theory of Justice, Cambridge: Harvard University Press.

Rozin, P./Scott, S./Dingley, M./Urbanek, J. K./Jiang, H./Kaltenbach, M. (2011): Nudge to nobesity I. Minor changes in accessibility decrease food intake, Judgment and Decision Making, 6(4), 323-332.

RKI (2013): DEGS 2012. Die Gesundheit von Erwachsenen in Deutschland, Berlin: Robert Koch Institut.

Schnellenbach, J. (2012): Nudges and norms. On the political economy of soft paternalism, European Journal of Political Economy, 28(2), 266-277.

Schumpeter, J. A. (1942): Capitalism, Socialism and Democracy, New York/London: Harper.

Schwan, P. (2009): Der informierte Verbraucher? Das verbraucherpolitische Leitbild auf dem Prüfstand. Eine Untersuchung am Beispiel des Lebensmittelsektors, Wiesbaden: VS Verlag für Sozialwissenschaften.

Selinger, E./Whyte, K. P. (2012): What counts as a Nudge? The American Journal of Bioethics, 12(2), 11-12.

Sen, A. (1989): Development as capability expansion, Journal of Development Planning, 19, 41-58.

Shu, L./Mazar, L./Gino, F./Ariely, D./Bazerman, M. (2012): Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end, Proceedings of the National Academy of Sciences (PNAS), 109(38) 15197-15200.

Smith, V. L. (1965). Experimental auction markets and the Walrasian hypothesis, The Journal of Political Economy, 73(4), 387-393.

Sneddon, A. (2001): What's wrong with selling yourself into slavery? Paternalism and deep autonomy, Revista Hispanoamerica de Folosofia, 33, 97-121.

Thaler, R. H./Sunstein, C. R. (2008): Nudge: Improving decisions about health, wealth, and happiness, New Haven: Yale University Press.

Tversky, A./Kahneman, D. (1981): The framing of decisions and the psychology of choice, Science, 211(4481), 453-458.

Von Weizsäcker, C. C. (2002): Welfare Economics bei endogenen Präferenzen. Thünen-Vorlesung 2001, Perspektiven der Wirtschaftspolitik, 3(4), 425-446.

WHO (2011). World Health Statistics 2011, World Health Organization: Paris.