

## **SEARCHING FOR THE MORAL BOUNDARIES OF *NUDGE***

Daniela Goya Tocchetto \*

He who lets the world (...) choose his plan of life for him, has no need of any other faculty than the ape-like one of imitation. He who chooses his plan for himself, employs all his faculties. He must use observation to see, reasoning and judgment to foresee, activity to gather materials for decision, discrimination to decide, and when he has decided, firmness and self-control to hold to his deliberate decision.

It is possible that he might be guided in some good path, and kept out of harm's way, without any of these things. But what will be his comparative worth as a human being?

J. S. Mill, On Liberty

**ABSTRACT:** The aim of this paper is to improve the comprehension of the moral implications of *Nudge*. Firstly we will discuss the potential moral concerns elicited by this policy alternative. In the following, we will propose an alternative definition of *Nudge* and a broader categorization of the types of agencies that it deals with. Both these proposals are intended to clarify what a *Nudge* is in order to make it easier to appreciate its moral boundaries. In the light of this refinement of the nature of *Nudge* we will be able to provide a more meticulous examination of its ethical implications

**KEYWORDS:** moral boundaries, nudge, paternalism

---

\* MSc in Philosophy and Public Policy (London School of Economics), PhD Student in Philosophy (Federal University of Rio Grande do Sul). Email: daniela.gt@terra.com.br.

## INTRODUCTION

If we were all *homo economicus*, we would live in the best possible world. Yet we are more like *homer economicus* (Thaler & Sunstein 2009). Thus we can improve the world and make it a better place for us, flawed humans. This is the appeal of *Nudge*, or of what Thaler & Sunstein call the real third way of making policies: Libertarian Paternalism.

*Nudge* advocates the design of choice structures with the purpose of “gently pushing” individual decisions in the direction that is either in their own or in society’s best interest. There are two main examples of *Nudge* provided by Thaler & Sunstein (2009). The first example, *Cafeteria*, illustrates the use of the empirical finding that we are more inclined to choose what appears first in a list of items. If we put the fruits on the top of the menu and the chocolate bar on the bottom, we can nudge people in the direction of a healthier diet – without limiting their freedom to still choose the chocolate bar.

The second example, *Save more Tomorrow*, underlines the influence of both endowment and discounting effects over our choices. People are more inclined to commit to a saving plan involving the money earned in future raises than in raises included in already received pay checks. *Save more Tomorrow* plans nudge people to increase their pension funds.

The impact of this new policy option has been, to say the least, impressive. *Nudge* has been able to bring together politicians with different and even conflicting views, on both sides of the Atlantic (Chakraborty 2008). However matters are not as simple as they are presented when it comes to the nature of *Nudge* – it is much more complex than it appears at a first glance. Given this complexity and the shadowy use *Nudge* makes of the common errors in which we incur, it can be associated with deceptive methods of influencing human behavior. Additionally, it elicits a variety of moral concerns due to its manipulative appearance.

The aim of this paper is to improve the comprehension of the moral implications of *Nudge*. Firstly we will discuss the potential moral concerns elicited by this policy alternative<sup>1</sup>. In the following, we will propose an alternative definition of *Nudge*<sup>2</sup> and

---

<sup>1</sup> In Section 2.

<sup>2</sup> In Section 3.

a broader categorization<sup>3</sup> of the types of agencies that it deals with. Both these proposals are intended to clarify what a *Nudge* is in order to make it easier to appreciate its moral boundaries. In the light of this refinement of the nature of *Nudge* we will be able to provide a more meticulous examination of its ethical implications<sup>4</sup>.

#### **NUDGE AND ITS POTENTIAL MORAL THREATS**

In this *Section* we will examine the ethical concerns elicited by *Nudge*, discussing on a case by case basis the potential moral threats related to this public policy alternative. We identified four potential threats, which will be elucidated in the remaining of this *Section*: (1) Neutrality; (2) Autonomy; (3) Virtue; and (4) Use of Reason.

#### **LIBERTARIAN PATERNALISM: THREAT TO STATE'S NEUTRALITY**

*Nudging* people's choices in ways that will improve the chooser's own welfare, without limiting the choice sets looks like a very seducing and appropriate manner of conducting public policies. Yet the extents to which this supposed "real third way" is able to maintain the libertarian commitment with the principle of neutrality in public policy is still to be analyzed.

Firstly, there is the issue about the results of experimental economics pointing to the inexistence of well-defined individual preferences and its implications for the specification of what Sunstein & Thaler mean by "being better off". Secondly, there are critiques against paternalism that extend its force also against its softer version – libertarian paternalism.

in *Libertarian Paternalism is not an Oxymoron* (2003), Sunstein & Thaler conclude, from the results showing that people lack well-ordered preferences, that the very meaning of the term "preferences" is unclear. In *Nudge: Improving Decisions about Health, Wealth and Happiness*, the authors argue that preferences exist, but are inconsistent. Humans are dichotomous beings: we are partly *Planner* (analogous to *homo economicus*) and partly *Doer* (analogous to *Homer economicus*). "In the morning the inner Planner resolves to forgo dessert at dinner tonight, and then, after dinner, the inner Doer proceeds to devour a bowl of ice cream" (Leonard 2008, 4).

---

<sup>3</sup> In Section 4.

<sup>4</sup> End of Sections 3 and 4.

If well defined preferences do not formally exist, how do we decide in which direction to steer people's choices – taking into consideration their own best interest? How do we solve the problem of identifying preferences empirically? This problem used to be solved by the revealed preference principle. But if choices no longer reveal individuals preferences, because they do not formally exist, in what way can we have access to these preferences in order to know what are the individuals own self-interest?

Conversely, if the meaning of the term preferences is clear, but inconsistent between the preferences of the Planner and the preferences of the Doer inside us, we still have a problem at hand. How do we decide whose preferences, the Planner's or the Doer's, should be satisfied in order to make people better off? Leonard (2008) provides an interesting insight in this respect, pointing out that the libertarian paternalist takes the side of the Planner, and identifies his preferences with the supposed best interest of the individual. The goal of *Nudge* then turns out to be protecting the Planner's preferences from the Doer's preferences.

The worry is that Thaler & Sunstein provide no sufficient reason for why one kind of preference is better than the other – why should we satisfy the Planner's and not the Doer's desires? Wilkinson (2009) also envisions the undetermined point left by Thaler and Sunstein about what is meant by “being better off”, when he emphasizes that “pointing to errors of in practical reason doesn't show someone would be better off for being made healthier” (Mill 1859 *apud* 2009, 6).

“...a man's mode of laying out his own existence is best not because it is the best in itself, but because it is his own mode...”  
Mill, 1859, Chapter III.

There are two main critiques against paternalism: one consequentialist and another deontological. The consequentialist critique reflects the skepticism about the possibility of a person making better decisions for and in the interest of another individual. As repeatedly stressed by Friedman, the individual himself is the best judge of his own interests and therefore the one who should always be in charge of his decisions.

The deontological critique is related to the intrinsic value associated with leading an autonomous life. Even if it were possible for an individual to precisely

determine the best interest of another and to decide for him so as to further this best interest, still this intervention would be morally undesirable. As Kant (1997) would put it, to deny individuals the control over their own lives would be to treat them as merely means to some good, and not as ends in themselves.

*Nudge* claims to be a solution to the consequentialist problem of paternalism. Behavioral Economics has documented a number of human inconsistencies in decision making. As a result, either well defined preferences do not exist or preferences are conflicting between our Doer and our Planner. Thus it is now possible to determine a person's best interest and to improve her well-being through the use of libertarian paternalism. Trout (2005) makes precisely this case. He argues that in particular circumstances we are able to detect that a person's considered judgments or long-term goals are not being pursued given some cognitive bias, and we can now use these very same bias to guide them in the direction of those interests.

Nonetheless, as discussed in the precious *Subsection*, a person's best interest still demands normative judgments to be defined, even with the help of the results of behavioral economics. Especially now with the results of his new science, we can no longer rely on the principle of revealed preference to empirically discover an individual's best interest.

The deontological critique to paternalism also applies to *Nudge*. When individuals are influenced by a *Nudge* and guided in some direction unthinkingly, they are less accountable for their choice and less autonomous. In this sense, *Nudge* is, like paternalism, demeaning of the intrinsic value connected with human autonomy.

Sugden (2006) emphasizes precisely this deontological point when he argues that the new findings of behavioral economics do not justify paternalism. In spite of the results pointing to the existence of incoherent preferences, Sugden claims that any form of paternalism would still threaten an important form of autonomy, represented by the opportunity to act based on unconsidered preferences. Even if we are choosing something that we under reflection would not choose, preserving the liberty to do so is an important form of freedom that cannot be prevented.

When people know that their choosing process is flawed, they bind themselves in ways to achieve what they think is best for them. For example, if we are on a diet and we do not want to eat big amounts of chocolate, many times instead of buying the

large cheaper pack, we buy the small and more expensive one. We are intentionally restricting our future set of choices in order to resist later temptation and facilitate the exercise of self-control. This sounds perfectly acceptable on moral grounds.

Yet there is an important issue about the autonomy of the choices made under nudging policies. If the State designs the choice structure and we just choose under it, we are deprived of the autonomous part of the decision. People will be unautonomously choosing under the illusion that they are engaging in a free choice.

At least part of the discomfort we feel about nudge is comparable with our resistance against all kinds of manipulation of the human mind. In order to better understand what lies underneath our common sentiments towards mind manipulative techniques<sup>5</sup>, we shall first imagine a brief and rough categorization of the ways in which we can influence human behavior; from the most perverse to the most socially acceptable.

- *Brainwashing*: could be mainly characterized as the unconscious and involuntary change of convictions and beliefs (Taylor 2006). This technique produces a real change in the preference structure of the individual. One example appears in the film *Clockwork Orange*, where an extremely violent boy undergoes a process in which, through a high number of sessions, he is induced to always and involuntarily associate bad feelings with violence.

- *Subliminal images*: when we change our behavior because of the influence of subliminal messages we undergo a momentary change of preference but not a change in our beliefs. If we choose soda instead of water because we were influenced by a subliminal image of a happy person drinking soda, we will not be able to give reasons for the option we made.

- *Advertising*: can be defined as persuasion involving rational and emotional tools employed on a conscious level. It is not part of the purpose of advertising to make us deliberate about the use of the product in question, but given that this technique influences us in a conscious level we remain with the possibility of deliberating about it if we so wish.

---

<sup>5</sup> Researchers disagree about the real effects of these techniques, but this is not relevant to the argument being made here. We are only concerned about our reactions to commonly known ways of manipulation.

- *Provision of information*: the disclosure of information by private institutions or by the government, with the goal of instructing individuals about some subject.

In Brainwashing, it is essential to underline that the change which the individual undergoes is complete: the person brainwashed has a new set of beliefs and preferences. Given these new preferences, the person is able to give reasons and to identify with the choices she makes after being brainwashed. We assume this to be the most deceptive technique mainly for two reasons.

Firstly, under brainwashing we change our set of beliefs and preferences not because we were persuaded by rational arguments, but through a process completely beyond our control. Secondly, we do not realize that our change of beliefs was due to the application of a deceptive technique. Consequently, it becomes more difficult to reverse the process and to change our beliefs again. And even if we had been through brainwashing by a voluntary choice, given the difficulty in reversing the process we still believe it is dangerous and morally condemnable.

In the case of Subliminal Images, the change is partial: we only change our behavior in specific ways, without changing our overall set of preferences. When acting under the influence of a subliminal image, we end up with a fragmented self (Bovens, 2009): we do not identify with the action we performed and we are not able to recognize the reasons why we acted in that manner.

Marc Blitz (2009) points out that the majority of American judges and scholars assume subliminal messages to be excluded from the protection of Free Speech given by the First Amendment. They argue that instead of being meant to offer information for our assessment, subliminal messages are meant to influence us in some determinate direction without giving the opportunity to deliberately question this direction. Yet if we undergo the influence of subliminal images by voluntary choice, we would not judge the practice as morally condemnable. This is so because we would still be in control of the situation if we wanted to stop being influenced, unlike the case of brainwashing.

The provision of information is not by any means deceptive. In this case, we are fully aware of the reasons that influenced our actions and we are able to deliberate about which is the best course of action for us. We are entirely under control of our behavior.

After analyzing these different ways of influencing human behavior, we can better evaluate what aspect of their nature elicits moral uneasiness. The common feature present in all of the above forms of manipulation is the idea of control that we have over our decisions. The problem does not seem to lie in the emergence of a fragmented self (Bovens 2009), like under the influence of subliminal images, since under brainwashing we end up with a coherent self but the technique is generally considered even more deceptive.

The problem seems to lie in one of the moral concerns about *Nudge* identified by Bovens (2009) – the incapacity to be in control of our own lives. In this sense, transparency is important in an instrumental way, given that it is needed for us to be able to deliberate about our choices and, as a consequence, take control of our decisions. In view of that, truthfulness is crucial, so that we can recognize how these things are affecting us.

In this context, *Nudge* is obviously less deceptive than brainwashing, which changes our beliefs without providing any opportunity for rational deliberation. It is also less deceptive than subliminal messages for two distinct reasons. Firstly, if completely alert we are able to detect its influence; thus it is possible to control and regulate its use (Bovens 2009). Secondly, it operates in a different level of consciousness; but we cannot say that *Nudge* works below the threshold of consciousness.

Therefore *Nudge* may be problematic in contexts where we do not detect its presence and we are influenced by it. In this case, the problem is precisely that we are not in control of our choices; we are not fully autonomous.

Mere respectability, the decent conformism of law-abiding citizens, is a facade that will collapse under pressure of temptation; because its possessor will not have the knowledge and judgment to do what is right no matter what the circumstances. (Gibbs & Pole, *Virtue and Reason*)

When choosing under a choice structure like *Save More Tomorrow* or *Cafeteria*, we are incurring in what could be called *mindless choosing*. We may be making the best decisions for ourselves or not, this will depend on our preference structure. Yet even if we are choosing the best course of action we are doing so without rational deliberation and without effort to restrain our impulses and educate our will.



An analogous experience can be thought of to elucidate the importance we assign to the effort we place in doing the right thing: the placebo effect. When a patient receives a placebo, the doctor has to lie to him in order for the “fake medication” to have an effect. Still we are ok with this practice. But when we are led to choose in some direction through a non-transparent process, we are less ok. We can understand this difference exactly in terms of the importance of effort. The healing is not something that we consider due to our own self worth, to our effort in becoming someone better; while choosing the right action is something that we directly connect to someone’s worth as a human being.

The importance of effort is promptly related to the assignment of responsibility for our choices in life and, consequently, to moral character building. One thing is to mindlessly choose which shoes to wear, and yet another to mindlessly opt for one way or another of conducting life.

At this point, one could point out Aristotle’s habituation argument to develop a virtue: through habit and practice of virtuous actions we can acquire virtue. However, if we are being nudged, we do not learn about the subject under consideration, we do not exercise the virtue in a conscious way and especially we do not exercise this virtue because we recognize its importance and aspire it in the first instance. There is no feature of what Aristotle describes as self-management in *Nudge*. So the question is: how can we build genuine character moral change in a blind way?

Bovens (2009) also calls attention to another kind of worry: that *Nudge* may change people’s behavior only in the short term, without long term effects. The only way to guarantee long term effects would be by assuring that people actually learn to make the best choices for themselves.

Within a view of virtue ethics, *Nudge* would also not be consistent with the requirements we expect from a virtuous person. A virtuous person is one whose life we admire. As Aristotle describes, a virtuous person is the one who acts “knowing that he is acting virtuously, choosing what he does and choosing it for its own sake, and acting from a condition that is firm and unalterable” (1105 a30-34). When choosing under a *Nudge*, we are not reasoning about the subject of our decision. Yet as emphasized by Annas (1993), “virtue is a habit of acting on reasons. However stable, the virtues bring it about that the agent acts rationally, not unthinkingly or

mechanically”, the purpose of acting virtuously “is to make the way I am one which I endorse, rather than a mindless pattern of acting” (Annas 1993, 51-57).

The last source of moral uneasiness elicited by Nudge could be understood in terms of a teleological view of human nature, drawing on Aristotle. He described nature in a goal-oriented way and consequently his ethics was guided by this teleological view – including human nature. He characterized a *good X* as one that performs its function really well. Thus a good man is the one that performs his function really well.

But what is a man’s function? It is his ultimate goal, i.e., a happy life – the activity of the soul in accordance with virtue. In Aristotle’s writings, soul means reason: the activity of the soul is the exercise of reason, of our deliberative skills. Thus every man’s function is to live a life in which he exercises reason; and the morally good man is the one who does so with excellence. Within this view, leaving a life without deliberation is immoral.

Kant in the *Groundwork of the Metaphysics of Morals* discusses the case of a man who “prefers to give himself up to pleasure than to trouble himself with enlarging and improving his fortunate natural predispositions” (4:423) as one who is immoral. Kant believes that the enhancement of our natural talents is our moral duty; the duty to the furtherance of our own humanity: “as a rational being, he necessarily wills that all the capacities in him be developed” (4:430).

Given a teleological stance on morality, whenever we make choices without reflection we are corrupting our end as humans. In this sense, when guided by a *Nudge* in a mindless way, choosing without thinking about the reasons for doing so, we are incurring in a perversion of our reasoning skills. We are corrupting our end as humans, in an Aristotelian gist; and we are not furthering our rational capacities, in a kantian gist.

Yet moral codes based on teleological interpretations are seen as greatly controversial (Perlman 2004). One clear example is the defense by the Christian church of the performance of sex only when we are doing it for its function: procreation and mutual bond (marriage). We disagree with this view in modern society. However it does not entirely eliminate our tendency to equate the idea of vice with the improper performance of a function.

## WHAT IS A NUDGE?

The original definition of *Nudge* is that it “is any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives” (Thaler & Sunstein 2009, 6). This original definition can be refined in order to provide a better basis to deliberate about the moral permissibility of *Nudge*.

One of the possible paths to comprehend what is a *Nudge* is to go back to the theory that originated it in the first place, and try to understand it in terms of that theory. This is the way Thaler & Sunstein structured their book *Nudge*, and I believe this is a good starting point in order to try to clarify some ideas about the subject.

Thus, the aim of this section is to come up with a definition for *Nudge* in terms of this cognitive model – what we will call a *Cognitive Definition of Nudge*; and to explore the characteristics and moral permissibility of a *Nudge* in the light of this definition.

As we now know, our brain has an amazing ability to process and solve problems; nonetheless, this ability is limited. There is a limited amount of information that we can handle – we are not equipped to scrutinize every single decision we make in our lives (should I use this mug or the other one? Should I brush my teeth for 3 or 5 minutes?). Thankfully, our brain developed “shortcuts” that enable us to quickly decide on a variety of issues. These so-called *heuristics* usually yield very good judgments that allow us to survive through thousands of small decisions that we have to make every day and that we are not even aware of.

However, these same helpful heuristics may sometimes give rise to systematic errors – the so-called *biases*. And here is where all this business gets interesting. The point is not that we as humans occasionally make mistakes when making decisions. The point is that many of these mistakes are based on the shortcuts that our brains developed to help us, and this feature adds a systematic character to the mistakes that we usually make. This systematic character presents us with the astounding new possibility of predicting human fallibility.

Thaler & Sunstein (2009) rely on the dual process approach to explain why we make systematic mistakes in a wide variety of choice situations, and therefore base the idea of *Nudge* on this theory. Accordingly, we will try to understand what a *Nudge* is in terms of the concepts developed within dual process theories. This attempt will result in what could be called a cognitive definition of *Nudge*.

Psychological research has led to the development of a two system approach to the way we make choices, the way our minds work when we act (Kahneman 2003). Despite divergences in terminology, these two systems can be called the Automatic System (AS) and the Reflective System (RS). The AS is intuitive and automatic, does not involve what we usually understand as “thinking”, is associated with the oldest parts of the brain (parts we share with other animals), is uncontrolled, effortless, associative, fast, unconscious and skilled. On the other hand, the RS is reflective and rational, deliberate and self-conscious, controlled, effortful, deductive, slow, self-aware, and rule-following.

Kahneman (2003) associates the AS with intuition and the RS with reasoning. He understands both systems in similar lines as explained above, regarding intuition as spontaneous and effortless, and reasoning as rational complex thinking and therefore effortful. Kahneman explains the relations between these two systems: the RS can be said to teach the AS to perform its tasks and also to monitor its performance. The AS, in turn, is responsible for the majority of our thoughts and actions – even if we may not want to admit that.

To better understand this “learning and teaching relation” between the AS and the RS, let us think about how we perform our daily tasks. As already discussed, we have to do numerous activities; like brushing our teeth, driving, choosing what and where to eat; not to mention the several things we have to do at work. When we learn to drive, for example, it is usually a slow process: in the beginning we have to pay attention to every little detail, but with time we are able to go to work without even realizing that we did that. This is so because at first we have to use our RS to learn how to drive. Once the RS learned it well and performed it many times, it can delegate the repeated task to the AS, so that we can use our RS to think about other important issues. It does not mean that the RS can teach the AS to perform all kinds of tasks – we still have to use the RS to solve different problems and when we want to deliberately reason about any subject. Thus, the idea is that when we are exposed to repetitive tasks the RS can first learn to execute it and, with practice, “teach” the AS to do it by itself.

Psychologists and neuroscientists argue that we have this dual system because the most developed part of our brain, the neocortex, is not able to carry out all the activities demanded from us on a daily basis. Therefore, we tend to use our RS only

when confronted with problems that require active reasoning. Still, Kahneman claims that the RS monitors the AS, correcting its decisions whenever possible.

Kahneman and Frederick (2002) describe this monitoring process in the following way: “System 1 quickly proposes intuitive answers to judgment problems as they arise, and System 2 monitors the quality of these proposals, which it may endorse, correct or override”(2002, 51); emphasizing that “errors and biases only occur when both systems fail”<sup>6</sup>(2002, 52).

However the monitoring that the RS executes is usually loose, consequently allowing many erroneous judgments reached by the AS to be expressed in human action – Ellen J. Langer (1992) refers to this erroneous judgments as “mindless behavior”. When these “mistakes” are made, Kahneman highlights that intuition is associated with poor performance.

We currently know both how these biases work and that they are systematic. Consequently, we can predict under which circumstances they are most likely to arise. Thus we are able to influence people’s behavior by changing those circumstances – the idea of *choice architecture*. And this is what the *Nudge* business is about: changing the circumstances in which people find themselves making choices, with the purpose of influencing their behavior in some desired direction.

For example, when we choose to eat the first item displayed in *Cafeteria*, be it chocolate or fruit, we are not reflectively thinking about what we should eat, we are not making use of our RS in this choice. The logic underneath this *Nudge* is that we knowingly have a heuristic of choosing what is displayed firstly in a menu. To this heuristic, there is an associated bias of sometimes surrendering to temptation. Therefore, when we place some food firstly in the menu we are inducing the occurrence of this framing bias. Consequently, we are relying on intuitive judgments that people make and to which they are not deeply committed (Kahneman and Frederick 2002, 59).

Nevertheless, Sloman (2002) reveals that some manipulations in the choice structure are able to reduce some biases, i.e., “a manipulation that reduces a bias by making an extensional probabilistic or logical relation transparent is in essence providing a representation that affords rule-based inference, allowing people to go beyond

---

<sup>6</sup> They refer to the RS as System 2 and to the AS as System 1.

associative reasoning” (Sloman 2002, 380). Thus *Nudges* of this kind would induce the use of the RS, instead of simply aiming at surpassing it.

We are now able to understand what is a *Nudge* in terms of this cognitive approach. **A *Nudge* is any aspect of the choice architecture deliberately designed to either: (1) trigger the use of the AS by inducing it to incur in some specific known bias, surpassing the monitoring of the RS, thus choosing in a direction that would supposedly be endorsed by the RS; or (2) reduce the occurrence of some specific known bias by making a particular information apparent in the choice structure, thus inducing the use of the RS<sup>7</sup>.** Given Thaler & Sunstein’s definition, the characterization just provided is restricted by what we could call a *libertarian rule*: “without forbidding any options or significantly changing their economic incentives”.

This definition, restricted by the so-called libertarian rule, applies for the majority of types of *Nudge* discussed in the literature. Still, there are some *Nudges* that do not seem compatible with the definition here provided – the *Emotional Nudges*. A clear example is the happy or unhappy emoticon displayed in the energy bills of about three hundred households (Thaler & Sunstein 2009, 75) – individuals who received the unhappy emoticon significantly decreased their energy consumption.

This *Nudge* is not clearly related with the application of any cognitive heuristic by the AS; neither it displays information unknown or forgotten by the individual that would induce the use of the RS. What is at work here is some kind of emotional mechanism that influences our actions by inducing an emotional response. If we take this emotional response to be part of the works of the AS, then this *Nudge* conforms to the first part of the definition. However this is not a consensus in the literature, and therefore I will treat these *Nudges* in a separate way, more similar to the mechanisms used by the advertising industry – mechanisms that completely surpass the use of our cognitive means in order to influence our behavior.

Summarizing, we will understand a *Nudge* under three different forms: one that aims at surpassing the use of the Reflective System, one that induces its use, and still another that triggers an emotional response.

---

<sup>7</sup> Examples of both kinds of *Nudges* will be discussed in Section 4.

Hitherto we have seen, in *Section 2*, the potential moral difficulties involved with *Nudge*. In order to better sort out these difficulties we have presented in this *Section* a cognitive way of defining it. In the light of this cognitive definition we are able to better assess which of the moral threats portrayed in *Section 2* are indeed implicated by *Nudge*.

The first important remark about this definition of *Nudge* is what is meant by the introduction of the word *supposedly*. Thaler & Sunstein throughout their work assume as an objective fact that some options are in people's own best interest and, were they to choose under ideal circumstances (complete information, unlimited cognitive abilities and no lack of will power); those options would be endorsed by their RS. Notwithstanding, as pointed out by Wilkinson (2009) through the example of *nudging* in the direction of healthier options, "pointing to errors in practical reason doesn't show someone would be better off for being made healthier" (2009, 6).

Hence, we cannot say that the person's best interest is to choose the healthy snack (*Cafeteria* example) merely because we are assuming that, were the person to have no *akrasia*, this would be the chosen option. We have no objective basis for this assumption but a belief that this is how people's preferences are. They may be so in average, but we have no logical claim to generalize this kind of preferences.

The fact that the person chooses what is displayed first in the line is not reason enough to say that the unhealthy snack is only chosen as a "mistake", as a result of lack of will power. This is so because the person has the tendency to choose whatever is displayed first; this is just a known framing bias, with no implication whatsoever about what is the preference structure of the individual. Accordingly, we cannot assume that what is chosen under a *Nudge* of the first kind in the definition, the kind that is designed to surpass the use of the RS, would be endorsed by it. We can only *suppose* that some specific option would be chosen were the individual to use its RS.

Recapitulating the definition developed in the previous *Subsection*, a *Nudge* can appear under three forms: (1) an influence that surpasses the use of the Reflective System, triggering some bias incurred by the Automatic System; (2) an influence that makes some particular information apparent and thus triggers the use of the Reflective System; and (3) an influence that surpasses the use of our cognitive system inducing an emotional response.

When our choice is affected by a *Nudge* of the first form, we do not use our Reflective System to deliberate about the options – our decision is completely taken by our Automatic System. We are choosing without reasoning and without effort, in a mindless and mechanical way. Under the influence of the second form of *Nudge* our decision making has a completely different nature: instead of pushing us into incurring in a bias from our AS, the *Nudge* provides us with some information that improves our deliberation about the choice at hand. Finally, under the third form, *Nudge* does not trigger the use of our cognitive system at all, be it automatic or reflective, going straight to the activation of an emotional response. When we choose under the influence of an *Emotional Nudge*, we choose not only without reasoning but also without any use of our cognitive structure.

Given these characteristics and the already raised potential moral concerns about *Nudge*, we perceive that the second form defined in this thesis seems to be morally harmless, given that it is analogous with the provision of information. It does not fall under the category of a paternalistic attempt, since the provision of information does not interfere with liberty or autonomy (Gerald Dworkin 2005). Thus all the moral worries related to paternalism and to a threat to autonomy do not apply in this case.

In addition, under the second form of *Nudge* the individual is in control of his choice and using his reasoning skills. He is able to fully exercise a virtuous life if he so wishes and he is performing his function as a human being in an Aristotelian sense. Thus the concerns rose about moral character building and the corruption of one's natural function do also not apply.

In contrast the remaining two possible forms of *Nudge* are not morally innocuous in the light of the ethical concerns discussed in *Section 3*. Under both forms, we do not have complete control over our choice, since we are making a decision without effort and rational reflection about the nature of the options. Usually when we choose under the influence of a *Nudge* of this sort, we would change our chosen option were we to undergo careful reflection. Thus in these cases *Nudge* is closer to subliminal messages (when we do not realize its existence) and comparable to advertising (when we do realize its influence).

The ethical worries about paternalistic intervention, both consequentialist and deontological, remain in these kinds of *Nudges* since they are able to influence a



person's behavior in some direction – thus interfering with her autonomy. Within a consequentialist critique, there is still no objective way to determine a person's best interest and the only real result of experimental economics is that the revealed preference principle is now unreliable. In the scope of a deontological critique, the intrinsic value of being in charge of one's decisions, even if they are somehow mistaken, is not refuted by *Nudge*.

Issues related to autonomy, virtue ethics and a teleological view on morality also persist in the first and third forms of *Nudge*. The lack of deliberation when making a decision and the automaticity associated with behavior under the influence of these nudges threaten the autonomy of our choosing process and the performance of reasoning – our proper function as humans. Additionally the lack of effort in mindlessly opting for one course of action rather than another jeopardizes the possibility of building moral character, as discussed in the preceding *Section*.

#### WHY DO WE NEED A NUDGE?

In this section, we will provide a categorization of the types of agencies that *Nudge* aims at correcting. Given this categorization, we will associate each source for the need of a *Nudge* with the kinds of biases that it is aiming to reduce or exploit<sup>8</sup> - illustrating these with examples of *Nudges*. The idea is illuminate the nature of *Nudge* and, in doing so, to refine the evaluation of when would a *Nudge* be morally permissible and when it would not.

Thaler & Sunstein (2009) do not address this issue explicitly. However we could say that they distinguish three types of reasons that produce the need for a *Nudge*: (1) lack of complete information, (2) limited cognitive abilities and (3) lack of willpower. Wilkinson (2009) takes this classification and expands it to the following four categories: (1) weak-will, (2) too busy to pay attention to the options, (3) lack of understanding of the options, and (4) indeterminate preferences.

Bovens (2009) presents a careful analysis of the types of agency subject of a *Nudge*, distinguishing between six types. There is *Ignorance*, which he equates with lack of knowledge; *Inertia*, associated with being either lazy or forgetful; *Akrasia*, or weak-will; *Queasiness*, related to avoiding emotional costs; *Exception*, kicks in when

---

<sup>8</sup> As explained in the definition of *Nudge* given in this thesis.

the individual is the exception of a general statistical rule; and *Social Benefits*, when the social interest conflicts with individual interest.

After elucidating the preexisting categorizations of the types of agency that *Nudge* aims at correcting, I will sum them up in what will be an attempt to account for all the reasons why we may be susceptible to a *Nudge*. Within this new categorization, there are four general types of agency that can be the focus of *Nudge*.

**NEW ATTEMPT TO CAPTURE ALL THE TYPES OF AGENCY**

***i. Ignorance***

The first one is *Ignorance*; by what we mean all the biases in which we incur due to lack of knowledge of some nature. As represented in *Table 1*, *Ignorance* can appear under five different forms: Lack of Information, Limited Cognitive Abilities, Low Frequency, No Feedback, and Social Benefits.

*Table 1*

<b>IGNORANCE</b>	<ul style="list-style-type: none"> <li>- Lack of Information</li> <li>- Limited Cognitive Abilities</li> <li>- Low Frequency</li> <li>- No Feedback</li> <li>- Social Benefits</li> </ul>
------------------	---

In the case of *lack of information*, we may be nudged in some direction because we do not possess all the relevant information necessary to be fully able to make the choice. This category is analogous to Thaler & Sunstein’s idea of incomplete information.

In the case of *limited cognitive abilities*, we may be nudged because we do not understand the options with which we are faced. This happens when we are faced with complex decision-making situations, where technical expertise is needed. This category embraces the notion of lack of understanding of the options captured both by Wilkinson and by Bovens.

In the case of *low frequency* and *no feedback*, we may be nudged because we are being faced with choices that we either only make very few times in life or that we get no feedback after choosing. In both cases, we have no opportunities to learn about our

tastes and about the consequences of the options with which we are being faced. This category contains the idea of *Exception* defined by Bovens (2009). In this case, due to the nature of the lack of knowledge we may not realize some statistical claims about the options with which we are faced. It may also be the case that we are, as Bovens described, the exception to the statistical rule; and this is why the *Nudge* should be subtle in order to allow individuals to easily opt for their own best option.

In the case of *social benefits*, we may be nudged because we are ignorant about the beneficial consequences some options have for society as a whole. As emphasized by Bovens (2009), it may be the case where this social benefits conflict with the pursuit of our own interests. Again, *Nudge* should be “libertarian enough” to allow people to make their own choices.

**ii. Defects of the Will**

The second type of agency is categorized under *Defects of the Will*; by what we mean all the biases in which we incur due to our inability to act in accordance with our own deliberative conclusions. This is when under rational reflection we reveal one set of preferences, but when we act we reveal a different set of preferences. As represented in *Table 2*, *Defects of the Will* can appear under three different forms: *Akrasia*, *Inertia due to laziness*, and *Selfishness*.

*Table 2*

<b>DEFECTS OF THE WILL</b>	<ul style="list-style-type: none"> <li>- <i>Akrasia</i></li> <li>- <i>Inertia due to Laziness</i></li> <li>- <i>Selfishness</i></li> </ul>
----------------------------	--

In the case of *Akrasia*, we may be subject to *Nudge* because of our lack of self-control. We tend to succumb to temptation, opting for alternatives that satisfy our short-term preferences but there are not in our long-term interest. This category was captured by all authors – Bovens, Wilkinson and Thaler & Sunstein.

In the case of *Inertia due to Laziness*, we may be nudged due to our tendency to procrastination when we are faced with options that require effort to be performed

or understood. We are many times lazy to read forms about different kinds of saving plans, for example. This category is analogous to part of the one of the same name already described by Bovens (2009).

In the case of *Selfishness*, we may be nudged because even when we know about the social benefits of some course of action, we may be still too selfish to change our conduct to achieve a more socially beneficial outcome. Again this is related with *Social Benefits*, portrayed by Bovens.

**iii. Limited Processing Capacity**

The third source of our susceptibility to a *Nudge* is *Limited Processing Capacity*; by what we mean the biases in which we incur given that we are not able to process all the information available for us. This category captures the same idea developed by Bovens under part of the category called *Inertia*; and by Wilkinson’s notion that humans are sometimes too busy to pay attention to the options. As represented in *Table 3*, *Limited Processing Capacity* can appear under only one form: Inertia due to Forgetfulness.

*Table 3*

<b>LIMITED PROCESSING CAPACITY</b>	-	Inertia due to Forgetfulness
------------------------------------	---	------------------------------

In *Inertia due to Forgetfulness* we may be nudged because we are too busy to pay attention to everything, and we consequently forget about some of the choices we need to make or about some relevant information we may need to make those choices. This is strongly related to the conception that humans have limited cognitive abilities, thus not possessing brain capacity to deal with an infinite number of demands – as elucidated by Thaler & Sunstein.

**iv. Avoidance of Emotional Costs**

The fourth and final type of agency is *Avoidance of Emotional Costs*; by what we mean all the biases in which we incur given that under many circumstances we

want to avoid choices that have a significant emotional cost to us. As represented in *Table 4*, *Avoidance of Emotional Costs* can appear under only one form: Queasiness<sup>9</sup>.

*Table 4*

<b>AVOIDANCE OF EMOTIONAL COSTS</b>	- Queasiness
-------------------------------------	--------------

In the case of *Queasiness*, we may be nudged because we associate a high emotional cost with some kinds of choices and, in order to avoid this high cost, we “choose not to choose”.

After categorizing the types of agency that make us susceptible to a *Nudge*, we shall analyze their relation with the kinds of errors in which we systematically incur and the kinds of nudges that can be associated with them. Firstly, I will briefly define the biases that will be part of this attempt to map agency, cognitive patterns and examples of *Nudges*.

*Table 5*

<i>Illusion of validity</i>	People are overconfident in their own judgment, even in the light of evidence that their judgment is wrong
<i>Anchoring bias</i> (reference-dependence)	Whenever people are exposed to a number or reference-point, their judgment is influenced by that number whether they intend to be influenced or not
<i>Status quo bias</i> (default)	People usually do not bother to opt out the default rules
<i>Endowment effect</i>	People tend to overvalue things already in their possession
<i>Framing effect</i>	The option people choose depends on the set of options that they are part of

---

<sup>9</sup> From Bovens (2009).

<i>Projection bias</i>	People tend to project their current emotional state into the future
<i>Representativeness</i>	People tend to build stereotypes that may not be valid when similarity and frequency diverge
<i>Availability</i>	People tend to be more aware about risks readily available, leading to a biased evaluation of risks
<i>Benefits Now, Costs Later</i>	People tend to avoid present costs and to seek present benefits
<i>Follow the Herd</i>	People tend to behave in accordance with others behavior

At this point we are able to relate these biases with the types of agencies defined in the preceding *Subsection*, while discussing some of them and providing related examples of *Nudges*.

**RELATING TYPES OF AGENCY WITH BIASES**

***I. Ignorance***

The first type of agency is *Ignorance* and the cognitive biases to which it relates are listed in *Table 6*.

*Table 6*

<b>IGNORANCE</b>	<ul style="list-style-type: none"> <li>- Anchoring</li> <li>- Status Quo</li> <li>- Framing</li> <li>- Follow the Herd</li> <li>- Availability</li> <li>- Representativeness</li> <li>- Illusion of Validity</li> <li>- Projection Bias</li> </ul>
------------------	--

As already discussed, our ignorance may appear under five different forms. Imagine we are faced with different health care or saving plans and we have to pick one. We may either lack the necessary information to decide or the choice may be too difficult and require some expertise which we do not have. In this case, we may be nudged, for example, in the direction of the default option.

As another example, imagine we have to decide whether or not to undergo a preventive exam for prostate cancer. Even though the probability of having prostate cancer is 3 in 10<sup>10</sup>, since we don't know anyone who has had this kind of cancer we may underestimate this probability. In this case, we lack information and making this information available can nudge us in the direction of the real probability and, consequently, we may decide to undergo the exam. The choice structure in this case could also make use of framing effect in the way of presenting the probabilities as a kind of nudge.

We may also have to decide on issues we only have to decide once in life (low frequency and no feedback), like whether or not to undergo a sex change surgery. In this case, making some information available about the average consequences of those choices can also nudge us into deliberating more accurately about whether or not to do it. Displaying information about the decision of the majority of people or the way in which the probabilities are presented are also examples of nudges that can be used in this case – triggering *follow the herd* and *framing* biases, respectively.

Finally, we may be ignorant about the social benefits of, for example, not throwing used kitchen oil in the sink. Making information about the bad consequences of this habit available can nudge people into changing their behavior.

## *ii. Defects of the will*

The second type of agency is *Defects of the Will* and the cognitive biases to which it relates are listed in *Table 7*.

*Table 7*

---

<sup>10</sup> Hypothetical.

<b>DEFECTS OF THE WILL</b>	<ul style="list-style-type: none"> <li>- Anchoring</li> <li>- Status Quo</li> <li>- Framing</li> <li>- Follow the Herd</li> <li>- Endowment Effect</li> <li>- Benefits Now &amp; Costs Later</li> </ul>
----------------------------	---

As already discussed, these defects may appear under three different forms. Imagine the *Cafeteria* case, where we have to choose between the chocolate and the fruit. We can be nudged to overcome our *akrasia* by framing the menu so as to put the fruit option on the top.

In *Save more Tomorrow* our desire to spend all our money can be overcome with a *Nudge* that triggers both the *endowment effect* and the *benefits now costs later* biases. Another example of a *Nudge* is setting the default option of our annual saving plan review as “same as last year”, helping defeat the bad consequences of our laziness in reading and filling out forms. As a last example, consider cases where we may be fully aware of the best decision in terms of social benefits and where the costs to act accordingly are low. Still we are selfish to worry about the socially beneficial consequences. The case of the urinal fly is an example of a *nudge* that frames our choice structure in a way that steers our decision in the direction of the socially beneficial outcome.

### *iii. Limited Processing Capacity*

The third type of agency is *Limited Processing Capacity* and the cognitive biases to which it relates are listed in *Table 8*.

*Table 8*

<b>LIMITED PROCESSING CAPACITY</b>	<ul style="list-style-type: none"> <li>- Anchoring</li> <li>- Status Quo</li> <li>- Availability</li> <li>- Framing</li> <li>- Follow the Herd</li> </ul>
------------------------------------	---



As already discussed, this agency may appear under only one form. Imagine that we do not read the annual forms about the revision of our saving plans not because we are lazy, but because our brain is so overwhelmed with other activities that we simply forget to do so. In this case, setting an appropriate default option may be a helpful nudge.

Another example of a case in which we make a mistake due merely to forgetfulness is when we travel to London and look to the wrong side to cross the street. In this case, framing our choice with a “Look Right” sign on the road may be a helpful reminder.

***iv. Avoidance of Emotional Costs***

The fourth and last type of agency is *Avoidance of Emotional Costs* and the cognitive biases to which it relates are listed in *Table 9*.

*Table 9*

<b>AVOIDANCE OF EMOTIONAL COSTS</b>	<ul style="list-style-type: none"> <li>- Status Quo</li> <li>- Framing</li> <li>- Follow the Herd</li> </ul>
-------------------------------------	--

As already discussed, this agency may appear under only one form. Imagine that we have to make some choice that involves a high emotional cost, like deciding whether or not to be an organ donor. In this case we may “choose not to choose” in order to avoid this emotional cost, and the setting of, for example, a default option is a strong nudge.

The moral permissibility depends both on the kind of agency that *Nudge* exploits and on the kind of bias that it induces. There are two important features to look into: (i) the nature of the type of agency and, given this nature, the moral relevance of being in control of it; and (ii) if the bias related to a given type of agency is being exploited or corrected.

*Ignorance and Limited Processing Capacity*

Firstly, we have to evaluate the moral relevance of being in charge of this agency. The ethical concerns to which *Ignorance* is most related is the corruption of our natural function (teleological argument), a threat to autonomous action and to moral character building. If we are ignorant about a subject matter and we do not try to improve our knowledge, we do not further our reasoning skills. Additionally, when we are ignorant about the nature of the choices we are faced with we are not capable of deliberating about them, thus not being able to choose autonomously and also to rationally build moral character. For example, how to have the virtue of generosity if we are ignorant about the social benefits of some courses of action?

Yet important to prevent ignorance and to try to gain as much knowledge as possible, we have to keep in mind that knowing everything is beyond our human capacity. Hence *Ignorance* is not a kind of agency that we have to be entirely in control of. So even if nudging in this area of human agency is morally problematic, it may sometimes be acceptable given the impossibility of knowing everything about all sorts of themes. This is exactly the case of the agency defined as *Limited Processing Capacity*. Given that our brain is not equipped to deal with an infinite number of demands, sometimes it happens that we simply forget to choose – for example when we forget to read and fill in a form for our saving plan option. This agency is completely beyond the scope of our control, and nudging in this area seems to be not only morally permissible, but also desirable.

Secondly, the moral permissibility of the nudges related with *Ignorance* is also directly associated with the use made of known biases. As analyzed in *Subsection 4.5*, nudges that induce the incurrence in a cognitive error are morally harmful and nudges that correct the effect of biases by triggering the use of the Reflective System are morally harmless. *Availability* is the heuristic more appropriately tackled with a harmless nudge. Nudges that deal with this heuristic usually make available information that we were not aware of – for example, the probability of contracting a disease. By displaying the information, the *Nudge* enhances our ability to correctly reason about the risks of opting for different courses of action.

*Nudges* based on other biases like *Framing* or *Follow the Herd* tend to fall into the category of those nudges that induce the incurrence in cognitive errors. Nonetheless, when judging the moral permissibility of a *Nudge* it is always important

to keep in mind that each bias will have to be the focus of a case by case analysis, with the purpose of deciding if the bias is being corrected or exploited by the *Nudge* under scrutiny.

*Defects of the Will and Avoidance of Emotional Costs*

Once more we need to look into: (i) the nature of the type of agency and given this nature the moral relevance of being in control of it; and (ii) if the bias related to a given type of agency is being exploited or corrected. *Defects of the Will* is the type of agency more closely related to virtue and moral character building. When we correct our will in order to avoid *akrasia*, laziness and selfishness we are building our self-control. The control over one's own will is an essential feature of a virtuous person. How to fully exercise the virtue of generosity without being able to control selfish impulses? *Avoidance of Emotional Costs* could also be included under a type of agency that, in order to lead a virtuous life, we have to be in control of. We cannot simply avoid making the decision of whether or not to be an organ donor only because it may involve queasiness.

Consequently these two types of agencies are the most morally problematic to be influenced by *Nudge*. If individuals do not learn by themselves to have the appropriate control over their own will, they will never be able to build moral character in an autonomous and conscious way. Not only that, but when placed under a situation of temptation where they suddenly find themselves without the help of a *Nudge* they will most likely succumb to the temptation – since they did not rationally develop the proper virtue.

Secondly, the analysis related with the morality of the biases associated with these kinds of agencies is analogous to the analysis conceived for the preceding agencies. Biases that are used in a way that triggers the use of the Reflective System are ethically less problematic than biases that are used in order to surpass reflective deliberation. In cases of *Defects of the Will* and *Avoidance of Emotional Costs*, the bias that is most likely to be corrected rather than exploited (*Availability*) does not seem to be the focus of any *Nudge* that deals with these types of agency. This fact points to the morally damaging nature of *Nudges* that aim at influencing these agencies.

## CONCLUSION

The former analysis succeeds in, at least, illustrating the complexity of the ethical features involved in *Nudge*. It raises concerns about the neutrality of the government; the autonomy of its citizens; the building of moral character and the proper use of our reasoning skills. In an attempt to face up to the complexity of an ethical analysis of *Nudge*, the two major strategies were to refine both its definition and the categorization of the agencies with which it deals.

Hence we ended up with three kinds of Nudges: one that aims at surpassing the use of the Reflective System, one that induces its use, and yet another that triggers an emotional response. Our analysis indicated that we should be worried about the use of *Nudges* of the first and third kinds; nonetheless the second kind emerged as morally harmless.

Regarding the agencies, we ended up with four broad categories: Ignorance, Limited Cognitive Abilities, Defects of the Will, and Avoidance of Emotional Costs. The first two revealed themselves less morally demeaning by being the subject of a *Nudge*; while the last two agencies appeared as the most related with moral character – indicating the presence of moral harm in *Nudges* that aim at correcting them.

Lastly, the bias that came forward as being able to be corrected by a *Nudge*, instead of exploited, was *Availability*. Other examples may emerge on a case by case assessment of the influence of the *Nudge* under scrutiny. The criterion is that whenever it triggers the use of the Reflective System it corrects the bias.

So even if nudging is many times morally problematic, we were able to show that it may sometimes be acceptable given the impossibility of knowing everything about all sorts of themes and the possibility of sometimes using a *Nudge* to correct a cognitive mistake.

## REFERENCES

- Annas, J. 1993. *The Morality of Happiness*. New York: Oxford University Press.
- Aristotle. 1999. *Nicomachean Ethics*, Irwin Translation, Hackett.
- Bernheim, B. D. and A. Rangel. 2005. Behavioral Public Economics: Welfare and Policy Analysis with Non-Standard Decision-Makers. *NBER Working Paper No. 11518*.

- Blitz, M. 2009. *Neuromarketing, Subliminal Messages, and Freedom of Speech*. [http://kolber.typepad.com/ethics\\_law\\_blog/2009/05/neuromarketing-subliminal-messages-and-freedom-of-speech-blitz.html](http://kolber.typepad.com/ethics_law_blog/2009/05/neuromarketing-subliminal-messages-and-freedom-of-speech-blitz.html). Accessed on: August 16.
- Bovens. 2009. The Ethics of Nudge. Forthcoming in Till Grüne-Yanoff and S.O. Hansson *Preference Change: Approaches from Philosophy, Economics and Psychology*. Berlin and New York: Springer, Theory and Decision Library.
- Camerer, C.; S. Issacharoff; G. Loewenstein; T. O'Donoghue; M. Rabin. 2003. Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism". *University of Pennsylvania Law Review*, 151, (3) (January): 1211-1254.
- Camerer, C. 2003. Behavioural Economics. London: *World Congress of the Econometric Society*.
- Chakraborty, A. 2008. From Obama to Cameron, why do so many politicians want a piece of Richard Thaler? *Guardian*, July 12.
- Dworkin, G. 2005. "Paternalism". *Stanford Encyclopaedia of Philosophy*. <http://plato.stanford.edu/>
- Friedman, M. 1982. *Capitalism and Freedom*. Chicago: University of Chicago Press.
- Gibbs, B. & D. Pole. 1974. 'Virtue and Reason'. *Proceedings of the Aristotelian Society*, Supplementary Volumes, 48: 23.
- Gilovich, T.; D. Griffin; D. Kahneman. 2002. *Heuristics and Biases*. Cambridge: Cambridge University Press.
- Heilmann, C. 2009. Cognitive Processes and Liberal Paternalism: Defending Nudge with Dual Process. *Working Paper*, London School of Economics.
- Kahneman, D. 2003. Maps of Bounded Rationality: Psychology for Behavioral Economics. *American Economic Review*, 93 (5): 1449-75.
- Kant, I. 1997. *Groundwork of the Metaphysics of Morals*. Gregor, M. J. Ed. Cambridge: Cambridge University Press.
- Kubrick, S. 1971. *Clockwork Orange*.
- Langer, E. J. 1992. Matters of mind: Mindfulness/mindlessness in perspective. *Consciousness and Cognition*. 1 (3): 289-305.
- Leonard, T. C. 2008. Nudge: Book Review. *Constitutional Political Economy*, Springer Netherlands 19 (4) (December): 1043-1062

- Mill, J. S. 2003. *On Liberty*. Edited by Bromwich, D. & G. Kateb. London: Yale University Press.
- Perlman, M. 2004. The Modern Philosophical Resurrection of Teleology. *The Monist*: Jan 04.
- Sloman, S. 2002. Two Systems of Reasoning. In: *Heuristics and Biases*. Massachusetts: Cambridge University Press.
- Spiegel, A. 2009. Using Psychology to Save You from Yourself. *Npr* 8 (June).
- Sugden, R. 2006. Taking unconsidered preferences seriously. *Royal Institute of Philosophy Supplement* 81 (July): 209-232.
- Sunstein, C.R. and R.H. Thaler. 2003. Libertarian Paternalism is not an Oxymoron. *University of Chicago Law Review* 70: 1159-1202.
- Taylor, K. 2006. *Brainwashing: The Science of Thought Control*. Massachusetts: Oxford University Press, 1 edition.
- Thaler, R.H. and C.R Sunstein. 2009. *Nudge*. London: Penguin Books.
- Trout, J. D. 2005. Paternalism and Cognitive Bias. *Law and Philosophy*, 24 (4) (July): 393-434.
- Wilkinson, M. 2009. *Nudging, health, and making people better off*. London: July.