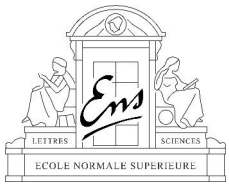

Cahiers de Recherche
Série « Décision, Rationalité, Interaction »

Cahier DRI-2014-01

Rethinking Nudges

Philippe Mongin et
Mikaël Cozic

INSTITUT D'HISTOIRE ET DE PHILOSOPHIE DES SCIENCES ET DES TECHNIQUES

Cahiers de recherche de l'IHPST
Série « Décision, Rationalité, Interaction »
Sous la responsabilité scientifique de Mikaël Cozic et Philippe Mongin

Rethinking Nudges[1]

Philippe Mongin[2] & Mikaël Cozic[3]

Résumé : Le concept de *nudge*, qui trouve son origine dans le livre éponyme de Thaler et Sunstein (2008), a plusieurs significations. En un premier sens, il s'agit d'une intervention qui cherche à réorienter les choix d'un agent en modifiant légèrement les conditions de son choix, en un second d'une intervention qui s'appuie sur sa rationalité limitée et en un autre encore, d'une intervention qui, quand elle est bienveillante, cherche à éliminer les obstacles constitués par sa rationalité limitée. L'article s'intéresse aux relations, sémantiques et factuelles, que peuvent entretenir ces trois concepts de *nudge*. Il soutient que les deux premiers sont fondamentalement découplés dans les principaux exemples de *nudge* invoqués par Thaler et Sunstein. Ce découplage n'a pas été aperçu par les auteurs en raison de l'identification, fautive, entre le second et le troisème concept de *nudge*, mais également parce qu'ils surestiment le pouvoir qu'a l'économie comportementale, relativement à la théorie du choix rationnel classique, de rendre compte des interventions qui atteignent leur but. A la suite de cette analyse, l'article considère certaines des questions normatives soulevées par Thaler et Sunstein. Leur affirmation audacieuse selon laquelle le libéralisme et le paternalisme peuvent être réconciliés en une unique éthique sociale – le *paternalisme libertarien* – a fait l'objet d'une critique philosophique minutieuse. Plutôt que de suivre cette approche abstraite, l'article prend une voie plus directe et fait valoir que Thaler et Sunstein perdent leur meilleur argument en faveur du paternalisme libertarien une fois que les différents concepts de *nudge* ont été distingués. Leur argument est en effet fondé sur l'idée que des interventions *peu intrusives* peuvent avoir des effets *puissants* grâce à un usage astucieux de la rationalité limitée, mais nous avons justement montré que celle-ci n'est pas réellement à l'œuvre dans les interventions considérées. L'article conclut qu'il vaut la peine d'explorer les trois concepts de *nudge*, mais indépendamment les uns des autres. Et en particulier que le dernier, qui vise à corriger les effets négatifs de la rationalité limitée, devrait recevoir une attention soutenue de la part des analystes de l'intervention.

Mots-Clés : nudge, paternalisme libéral, analyse des politiques, économie comportementale, théorie du choix rationnel.

Abstract: *Nudge* is a semantically multifarious concept that originates in Thaler and Sunstein's (2008) popular eponymous book. In one of its senses, it is a policy for redirecting

---

[1] Earlier versions were presented in 2013 and 2014 at seminars or conferences in Trento, Paris, Nashville, Lyon, Saint-Denis and Rotterdam. Each time the authors received helpful comments for which they wish to thank the participants.
[2] Centre National de la Recherche Scientifique a& HEC Paris. Email address:mongin@greg-hec.com.
[3] Université Paris-Est Créteil, Institut Universitaire de France & IHPST. E-mail : mikael.cozic@ens.fr

an agent's choices by only slightly altering his choice conditions, in another sense, it is concerned with bounded rationality as a means of the policy, and in still another sense, it is concerned with bounded rationality as an obstacle to be removed by the policy, when the latter has a benevolent aim. The paper centres on the interrelations, both semantic and factual, of these three nudge concepts. It argues that the first and second are basically disconnected on Thaler and Sunstein's major examples of nudges, and that this has gone unnoticed to them because they wrongly equate the second with the third concept, and also because they overestimate the explanatory power of behavioural economics, compared with that of classical rational choice theory, to account for successful interventions. After completing this analysis, the paper moves to some of the normative issues raised by Thaler and Sunstein. Their thought-provoking claim that liberalism and paternalism can be reconciled within one and the same doctrine of social ethics - *libertarian paternalism* – has been subjected to thorough philosophical criticism. Rather than following this abstract line, the paper takes the shortcut of arguing that Thaler and Sunstein lose their best defence of libertarian paternalism after the nudge concepts are disentangled. They had effectively based their case on the view that *slight* interventions could have *powerful* effects through a clever use of bounded rationality, and it has been shown that the latter is not really at work in the interventions they consider. The paper finally concludes that the three nudge concepts are worth pursuing, though independently of each other, and in particular that the third one, which involves correcting the pitfalls of bounded rationality, should receive sustained attention from policy analysts.

Keywords : nudge, liberal paternalism, policy analysis, behavioural economics, rational choice theory
Classification JEL: D03, D18, D70, K32, K39, M38.

.

1. Introduction

Nudge is a semantically multifarious concept that originates in Thaler and Sunstein's (2008) eponymous book and has disseminated from there to current law and economics, economics and philosophy, and behavioural economics. [1] According to one of its senses, it is a policy for redirecting an agent's choices by very slightly altering his choice conditions, so that the interference is kept to a bare minimum. Thaler and Sunstein have introduced nudge as a general category of interventions, by either public or private parties, and specifically argued for *benevolent* nudging - i.e., nudging intended to promote the agent's welfare - by any of these parties (the State as a particular case). They view benevolent nudging as either a substitute for, or a complement to, more traditional benevolent interventions, which typically rely on bans, commands, or heavy manipulations of choice incentives. Following another sense, nudge has to do with bounded rationality. Thaler and Sunstein repeatedly stress the agents' cognitive and practical limitations in decision making, referring to behavioural economics at large, and they conceive of nudges as being interventions that make strategic use of these limitations. Sometimes they view bounded rationality mechanisms as being universal tools of intervention, but they more often view them as a specialized tools, which, like homeopathic medicines, serve the objective of countering the unfavourable consequences of bounded rationality itself.

In sum, there are three distinctive senses available for nudge: (1) an intervention that interferes with the choice conditions minimally; (2) an intervention that uses bounded rationality strategically; and as a prominent case of benevolent nudging, (3) an intervention that tries to remove the negative effects of bounded rationality. We will call them *nudge 1*, *nudge 2* and *nudge 3* for brevity. That nudge means more than one thing matters a great deal to Thaler and Sunstein's social ethics position, which they capture by another special expression, *libertarian paternalism*. In their work, the terminology of this position has actually antedated the terminology, if not the idea, of nudge. [2] Their long standing interest is

[1] *Nudge* (2008) is intended for a wide audience and can be supplemented usefully by more academic work by Sunstein and Thaler (2003), Thaler and Sunstein (2003), and Sunstein (2013). However, most of our references will be to the first text, here cited in the paperback slightly expanded edition (*Nudge*, 2009).

[2] The word "nudge" does not seem to appear in Thaler and Sunstein before their eponymous book. In a previous use, Kahan (2000) meant by "nudge" a mildening of criminal penalities that would make social norms more conformable with the law.

to reconcile liberalism [3] (in the sense of respecting the individual's freedom of choice) and paternalism (in the sense of giving priority to the welfare improvement over the individual's spontaneous will), and they believe to have found a key to this problematic reconciliation in the idea of benevolent nudging. They are committed by their moral position to showing that strong interferences are not always necessary to improve the agents' welfare, and their basic argument is that nudges, in the bounded rationality sense, strike the right kind of balance between efficacy and nonintrusiveness, hence can also be viewed as good nudges, in the slight interference sense. Although Thaler and Sunstein are primarily interested in pushing a moral claim, they need their conception of nudge, in its several senses, to ground it properly.

Especially in *Nudge*, but even in their more academic work, Thaler and Sunstein primarily argue by way of examples. Their preferred ones are changing the display of food in public canteens, proposing plans for arranging one's future savings, allowing for withdrawal periods in consumers' choices, translating complex data into pragmatically usable information, enlarging the choice among complicated items with suitable default options. From what they say or rather suggest, these interventions aim at circumventing the obstacles to a good decision that cognitive biases and related features involve, exploit these very weaknesses for their success, and are so devised that they do not meddle severely with the initial decision processes. That is, each example in the list supposedly satisfies all senses of nudge together – nudge 3, 2 and 1 in that order. We may concentrate on the suggestion that all examples satisfy nudge 1 and nudge 2, since the target claim that strong interferences are unnecessary to improve the agents' welfare does not depend on whether nudge 3 holds.

In the philosophical literature that Thaler and Sunstein have inspired, the normative assessment occupies centre-stage. Some have evaluated nudges *per se*, i.e., independently of their alleged connection with liberal paternalism, [4] but most have scrutinized the moral soundness of this doctrine while also evaluating nudges. Nearly unexceptionally, conclusions are negative. According to the first group, to employ nudges lacks respect for the agent and is more invasive that first seems, and the second group makes this dismissal part of a larger claim, to the effect that libertarian paternalism is an unsound doctrine. The attention that

[3] We will often say "liberalism" instead of "libertarianism", Thaler and Sunstein's preferred expression, because t the liberal tradition over and beyond its specific, mostly U.S. libertarian form.
[4] See especially Bovens (2009).

writers in either group pay to the multifarious sense of nudge varies, but even those who notice it do not explore it in detail.[5] This is what we propose to do in this paper. We are first of all concerned with a better understanding of the new concept, and it is only after having tried to separate its constituent ideas that we will attempt to draw some normative conclusions. We will end up being critical of libertarian paternalism, but on the strength of a semantic and factual analysis, not of a metaethical argument. This is where we differ from the second group of writers, and perhaps take a methodological advantage, given the contentious nature of the metaethical claims involved. We also distance ourselves from the first group as a consequence of our analyses. They lead us to downplay its moral objections, and on a more positive line, to endorse corrective interventions that are neither straightforwardly classical nor nudge-like in Thaler and Sunstein's sense.

The next section analyzes nudge 1, 2 and 3 semantically. It points out that the three concepts cannot be connected at the level of meaning alone, whence it follows that any argument for libertarian paternalism that aims at connecting nudge 1 and 2 must be factual. As section 3 shows in detail, Thaler and Sunstein's own list of examples defeats this factual connection, whence the inductive presumption that it generally does not hold. This is the core argument of the paper. Section 4 relies on it to reconsider the issues of liberal paternalism along the lines sketched in the last paragraph.

2. Nudge 1, 2 and 3

Compare the following two sentences from Thaler and Sunstein:

"A nudge, as will use the term, is any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any option or significantly changing their economic incentives" (2009, p. 6).

"In accordance with our definition, a nudge is any factor that significantly alters the behavior of Humans although it would be ignored by Econs" (2009, p. 8).

The two sentences enunciate meanings for what we call nudge 1 and nudge 2 respectively. When taken together, they can be understood in two ways, depending on how "in accordance to our definition" is read. If this phrase refers to what comes next in the second sentence, it

[5] Hausman and Welch (2010), and Selinger and Whyte (2011) after them, have a clear sense of the extensive range of meaning of nudge. It does not come out so transparently from Mitchell (2005), Grüne-Yanoff (2012) and Qizilbash (2012).

introduces a new definition of "nudge", to be made consistent with the definition already introduced by the first sentence. But the phrase may also refer to that very definition, in which case the second sentence would state a consequence of adopting it instead of a competing definition. In either interpretation, the two sentences need to be connected *analytically*, i.e., by resorting only to the meaning of the words and the laws of logic. [6] Either they must be analytically equivalent (in case they provide two definitions) or the second must analytically derive from the first (in the other case).

There seems to be only one argument to connect the two sentences analytically, and here it is in outline. Suppose an intervention alters an agent's behaviour while leaving the set of options and the economic incentives exactly unchanged; whether it does that in a predictable or unpredictable way is immaterial to the supposition. Then it apparently follows that the agent exhibits some form of bounded rationality. Indeed, a fully rational agent - an "Econ" - would have previously reached his optimum position and the intervention would give him no reason whatsoever for departing from it. Thus, the agent fitting the situation must be imperfectly rational – he must be a "Human". Specifically, he may still be in the process of searching for a satisfactory decision, and thus oscillate from one temporary decision to another; or if he has finalized a decision, he may revise it as a consequence of the intervention because he wrongly believes that the environment was altered. [7] This argument points towards an analytical implication from (the sentence stating) a certain kind of nudge 1 to (the sentence stating) nudge 2.

Importantly, the implication does not follow from nudge 1 *in general*. Consider Thaler and Sunstein's twin clauses "without forbidding any option or significantly changing economic incentives". They are compatible with the set of options being increased and the economic incentives undergoing minor changes, in which case the test of bounded rationality cannot be probing anymore. If the objective environment is altered, however slightly, the agent who revises his decision can very well *rationally* do so. Notice in passing that the dual test - the agent does not revise his decision despite the fact that the objective environment varies – is probing under no circumstances. The agent may be static not out of inertia, but rather because he is on the alert and does not find the signal powerful enough (formally, he would maximize a constant preference on the interval of variation of the intervention parameter).

[6] We gloss over the classic Quinean difficulties of this analyticity concept.
[7] For instance, the agent may be subject to a framing effect; see below.

Hence Thaler and Sunstein cannot claim any relevant analytical connection between the two concepts of nudge 1 and 2. To tidy up their semantics, we could take one of the above sentences to state a definition and construe the other as expressing the factual claim that nudges so defined satisfy the other concept (at least broadly or typically). If one is primarily interested in the normative consequences, it seems appropriate to take nudge 1 as being the defining feature, since it is closer to the normative than nudge 2. It describes a kind of interventions that libertarian paternalism would approve of, and thus provides an intermediate-level criterion to evaluate interventions from this perspective. Then, nudge 2 would describe factual properties that are instrumentally relevant to fulfilling the criterion. However, this reconstruction can be reserved for the normative discussion, and in the next section, we will rather use a symmetric and non-committal semantics. We eschew the definition problem, just talk of *interventions*, assuming that this has a sufficiently clear meaning, and interpret Thaler and Sunstein as making the factual claim that there is a significantly wide range of interventions that jointly satisfy nudge 1 and nudge 2.

To check the claim properly, we need to analyze further the nudge 1 and nudge 2 concepts. Regarding the former, the conditions are (i) *not forbidding any option* and (ii) *not significantly changing the economic incentives*. [8] In order to have some bite, these conditions should be referred to the objective environment, and not to subjective particulars, as decision theorists would be prone to conclude. That is to say, options should be externally and empirically defined - a plate of meat or an insurance contract – and not identified with bundles of perceived characteristics, or with prospects stating the uncertain consequences of getting the object, following two classic constructs of decision theory. Similarly, economic incentives should only mean the physical quantities of options and the money costs and benefits attached to them, disregarding the immaterial costs and benefits that decision theory would typically also consider.

Suppose to the contrary that options are construed in one of the subjective ways just said. Then adding options, as (i) permits, could very well bring out characteristics that the agent had not taken into account when comparing the initial options – alternatively, it could very

[8] The authors have initially contented themselves with the first condition. "Choices are not blocked or fenced off" was all they required in Sunstein and Thaler (2003a, p. 1162). Only at the stage of *Nudge* do they introduce the second condition.

well bring about a revision of the subjective probabilities that the agent previously attached to the consequences of these options. In brief, the addition would lead the agent to reconceptualize *all* options, and the new set would not be set-theoretically comparable with the old one, as (i) precisely requires. Similarly, if the construal for economic incentives includes immaterial costs and benefits, the corresponding condition (ii) may not be easily met. Part of the problem here is that the extra factors would be hypothesized rather than observed, and thus likely to be measured in terms of the behavioural changes they supposedly produce. Such a circularity may not be too much of a problem for some theoretical purposes, but it sounds disastrous in the present epistemic context. An efficient intervention is one with large behavioural effects, and by the circular measurement procedure, it would automatically be associated with large changes in the incentives, thus defeating (ii). The only way to salvage this condition appears to interpret economic incentives literally, in terms of physical quantities and money values, *pace* Thaler and Sunstein themselves, who believed they could afford to be more sophisticated. [9]

Thaler and Sunstein require (i) and (ii) to be met together for an intervention to qualify as nudge 1. This may be asking too much, even with the low profile semantics recommended by the last paragraphs. An example will bring this point home. It is meant to show that even small changes in economic incentives - such that (ii) holds - can be accompanied with a complete reshuffle of the option set - so that (i), which requires set-theoretic comparisons, does not hold.

Moral hazard is a familiar problem faced by insurance companies. Once they are insured, customers often do not always make the prevention efforts that their contracts formally requires, and the company can oppose this drift only by either *ex ante* altering its menu of insurance contracts or by *ex post* instructing more inspections, which will of course have an *ex ante* effect through the customers' increased expectation of being caught. Thaler and Sunstein would presumably favour the former move as being less intrusive, i.e., closer to nudge 1, than the latter. For it seems intuitive that the *ex post* solution changes the economic incentives more drastically than the *ex ante* one; moreover, to keep the *ex post* solution credible, the company would have sometimes to deny indemnities and even cancel contracts,

[9] "Some of our nudges do, in a sense, impose cognitive (rather than material) costs, and in that sense alter incentives. Nudges count as such, and qualify as libertarian paternalism, only if costs are low" (2009, footnote of p. 8). However, this could be an incidental comment.

and this would amount to restricting the option sets of at least some customers. However, consider the hypothetical case in which none of the contracts in the initial menu had a deductible. The company could readily conclude that this contributed to the poor incentives and a deductible should be introduced in all and every contract. (Standard insurance economics rigorously deduces the conclusion, but sheer commonsense already warrants it.) This involves the company in redesigning the menu - the option set - entirely. This is still the case if, for some reason, say because the head accountant has enjoyed reading *Nudge*, the company decides to produce only a *small* change in the customers' economic incentives. There will be a small deductible in every new contract, and the resulting menu will be disjoint from the previous one. Thus, (ii) will be realized but not (i), and the apparently more gentle *ex ante* solution fares better than the *ex post* solution only on one of the two dimensions.

Examples like this point to an unnoticed tension between Thaler and Sunstein's two conditions for nudge 1. By allowing for (ii), they often nullify (i). A condition put on the option set appears to convey the lightness of an intervention more transparently than a condition put on incentives, especially if the next step is to connect light interventions with freedom of choice. So why not simply restrict attention to the former? Interestingly, Thaler and Sunstein were satisfied with this sole condition in their writings before *Nudge*. However, the insurance example suggests that it is right to add something on incentives. Otherwise, it would be more difficult to account for the intuition that the company's *ex ante* solution is less intrusive than the *ex post* one. The former enjoys over the latter the advantage of carrot over stick, and the option set condition does not properly account for this commonsical feature. All and all, Thaler and Sunstein are justified in having two conditions, and what they need is a scheme of compatibility. We submit that (i) should be lexicographically prior to (ii). That is, if the option set weakly increases, the intervention counts as nudge 1, and it weakly decreases, it cannot be nudge 1; and the two option sets cannot be compared, the intervention counts as nudge 1 if and only if it changes the economic incentives only by a small margin.

The nudge 2 concept also suggests second thoughts. Thaler and Sunstein express it by contrasting "Humans" with "Econs". This reflects their adherence to the programme of behavioural economics, which is to replace the rationality assumptions of decision theory and neoclassical economics by descriptively more realistic – if possible, experimentally based - assumptions. The general public having itself long entertained doubts on the "rational economic man", Thaler and Sunstein's formulation is a walkover to rhetorical success.

However, as the record shows, standard theories have the resources to account for some of the alleged rationality failures, and the alternative assumptions currently floated in behavioural economics are not powerful enough to displace rationality assumptions entirely. The theory of risky decisions exemplifies either point well. [10] In the present state of the art, "Humans" and "Econs" are overlapping populations, and to contrast them is not a good way of introducing nudge 2.

Moreover, is it is the agents' characteristics, not the clash between different schools of thinking about them, which is at stake, and even if behavioural economics were clearly superior to the standard theorizing, there would be little reason to stress this epistemological contrast. Let us then translate nudge 2 into the language of objective properties. [11] The literal translation is that the intervention succeeds only with boundedly rational agents, not with perfectly rational ones. This is still too theoretical a distinction, since real agents are not so segregated, and what matters in actual fact are not the individuals, but their decision processes. Hence we suggest, nonliterally, that an intervention counts as nudge 2 *if it strategically exploits bounded rationality within an agent's decision process*. This can take the form of either bringing about a new decision process or just modifying an existing one, but in the latter case, not to the point of inducing perfect rationality. Otherwise, nudge 2 would cover educative strategies, which would take us too far away from the initial idea.

How should one conceive of bounded rationality in the present context? We will take up the list of psychological factors that *Nudge* relates to bounded rationality and criticize it. The list has three broad groups: "biases and blunders", "temptation" (they say "resisting temptation", quoting the remedy before the ill), and "following the herd", which we briefly consider in turn.

Thaler and Sunstein are aware that the third group may be out of scope. Apparently gregarious behaviour can often be justified in terms of strategic equilibria, which may even be

[10] For example, today's well-regarded *cumulative prospect theory* is an admixture of relatively standard assumptions with some experimental economics findings, and it has become common ground between the two camps. See Wakker (2010) for a recent account.

[11] Thaler and Sunstein (2009, p. 19) also distinguish between "system 1" and "system 2", two terms that refer to objective properties, not to schools of thinking. However, it is not clear how this distinction from recent psychology - see, e.g., Kahneman (2011) - relates to Thaler and Sunstein's intented contrast of "Humans" with "Econs". Heilmann (2014) uses the former directly as a classificatory tool to discuss nudge and libertarian paternalism.

based on explicit strategic thinking on the agents' part. [12] Thaler and Sunstein do not express much doubt about the second group, although they should. Standard decision theory allows for *ex post* revisions of *ex ante* plans when new information occurs, and some of the behaviour that is informally described under the temptation heading is precisely of this revision type. The theory has also considered the possibility that information remains the same, and even in this case, some scholars have questioned the view that rationality entails one's sticking to one's *ex ante* plan. There is an ongoing debate on time-consistency in relation to expected utility theory, with some claiming that it is a rationality condition and others disagreeing. [13] Furthermore, the temptation group also raises the issue of how to discount the future if one should, and this is another unsettled area. On a widespread interpretation, giving in to temptation simply means applying a high discount rate to the future, which is viewed a psychological feature unconnected with either rationality or irrationality.

By contrast, the first group is indisputably relevant to bounded rationality, with its classic list of biases: anchoring and adjustment, availability, representativeness, overconfidence, loss aversion, status quo bias, framing. [14] With some effort, decision theorists can accommodate overconfidence, loss aversion and status quo bias into standard preference models. Anchoring and adjustment, availability and representativeness are not so easy to fit in, and framing is absolutely recalcitrant, which suggests separating this item from the others. Another internal distinction is that some factors can be viewed as heuristics leading to biases, while others appear to be straightforward biases. Tversky and Kahneman (1974) introduced anchoring and adjustment, availability, and representativeness in the context of judgment under uncertainty, and argued that they were common heuristics to replace or simplify probability calculations; they also described them as biases since they lead to errors in many cases. By contrast, there is no heuristic sense to be made of overconfidence, loss aversion, status quo bias. Framing is an outlier again because it is not a heuristics and perhaps not even a bias; indeed, a bias goes always in the same direction, whereas framing errors have no general pattern.

[12] See Chamley's *Rational Herds* (2003). Thaler and Sunstein (2009, p. 54) express doubts but do not use the strategic language.
[13] Machina (1989) and McClennen (1990) survey the issues. A major problem is that in the presence other conditions, which may seem cogent, time-consistency entails the expected utility criterion, which some, like Allais, have complained might lead to violations of rationality.
[14] See Kahneman, Slovic and Tversky (1989) and Kahneman and Tversky (2000).

The secondary literature on nudges has expressed concerns about the reliability of bounded rationality phenomena for the purpose of an intervention. [15] The internal distinctions just drawn suggest that these concerns should be hierarchized. Biases that result from using a heuristics seem to have less predictable effects than do straightforward biases. The reason is that, unlike the latter, the former are partly under the individual's control and their application is to that extent optional. They can be given up, whereas it is more difficult to correct one's pure biases. Framing is at the bottom of the predictability ladder.

In sum, nudge 2 goes in the opposite direction to nudge 1, i.e., underdetermination. Its conceptual meaning is vague, and its empirical extension is unclear beyond the first group and even within it. This is easily explained by the mixed record of today's behavioural economics, which offers more hints than genuine results. This has at least the advantage of not precluding the possibility that nudge 2 and nudge 1 factually overlap.

Unlike some discussants, we follow Thaler and Sunstein by considering interventions more generally than *benevolent* ones. Benevolence here means that the intervention aims at increasing the agent's welfare, as the intervening party views it; welfare may not be that party's sole objective, but it must be more than a side-effect; and it must bear *some* relationship to the agent's subjective sense of welfare, even if discrepancies are precisely a motive for the intervention. There are so many problems surrounding these ideas that it is best not to examine them here.[16] For simplicity, we will only suppose that the intervening party has some interest in reorienting the agent's choices. This may be a benevolent interest, but possibly also a mischievous or indifferent one. In this way, we can do justice to Thaler and Sunstein's recurring claim that agents are often nudged by others without any consideration for their good, and quite possibly at their expense.

Our last technical concept, nudge 3, enters stage at this juncture. It covers those benevolent interventions which proceed *by countering the unfavourable effects of the agent's bounded rationality on his decisions*. There are other ways in which the objective of improving the agent's welfare could be reached, even if the agent has limited rationality, and even if the intervention also consists in redirecting his choices, so this is but a very particular case of benevolent interventions. Nudge 3 should carefully be distinguished from nudge 2, the latter

[15] See, e.g., Hausman and Welch (2009) and Grüne-Yanoff (2012).
[16] The problems raised by these ideas are central to Qizilbash's (2012) critique.

being concerned with bounded rationality as a *means*, while the other includes it into the *objective* (though of course negatively, as an obstacle to welfare). Thaler and Sunstein pay little attention to this elementary distinction, perhaps because they lay so much stress on concrete examples, which are often equivocal in this respect. Sometimes a given bias – say, representativeness – can be turned against itself, and this fosters confusion between nudge 2 and nudge 3. A programme that means to inform the agent can pedagogically exploit bounded rationality devices, and this nudge 3 intervention is also nudge 2 if it turns out that the agent would not have absorbed the information without the devices. Public health campaigns offer excellent examples of these complex transitions.[17]

Nudge involves some secondary equivocations that are easy to dispel. First, by the same word, the authors refer both to interventions and the biases and decision faults underlying some of them, as in nudge 2. The glide is explainable, but it is better to avoid it, as there is so much more to the idea of a mechanism used with a purpose than there is to the idea of this mechanism taken in and by itself. [18] Second, Thaler and Sunstein think of nudging sometimes as being successful (in the sense of altering behaviour according to the purpose) and sometimes as being only intentional (it may either fail or succeed). Many words of the intentional language share these two meanings, which are in fact inseparable, because the meaning of an intention comes out more easily if one hypothesizes some form of success in realizing it. [19] By this token, we will often discuss the examples of the next section *as if* they reached their aim of changing the agent's behaviour along the proposed scheme, although this is far from obvious concerning several of them. What we plan to investigate is whether interventions fulfill the semantic conditions put forward for nudge 1, nudge 2 and nudge 3, and not whether they deliver the promised changes. [20]

---

[17] Our threefold division of nudges can be compared with other definitions in the literature. We read Bovens (2009) as defining nudge only like nudge 2. By contrast, Hausman and Welch (2010, p. 126) define nudge by the conjunction of nudge 1, 2 and 3: "To sum up: Nudges are ways of influencing choice without limiting the choice set or making alternatives appreciably more costly in terms of time, trouble, social sanctions, and so forth. They are called for because of flaws in individual decision-making, and they work by making use of these flaws".

[18] This complaint has been made in the literature to rebuke Thaler and Sunstein's claim that nudging is an inevitable component of social life; see footnote 27.

[19] We echoe Ryle's (1949) analysis of "success words".

[20] Commenting on paternalism, Hausman and Welch make a somehow related comment: "What characterizes paternalism are the aims with which one acts and the means one employs, not whether one is successful" (2010, p. 129).

3. Categorizing interventions

Let us start with the didactic introductory example of *Nudge* – the by now famous cafeteria example. The head of a cafeteria decides to improve consumers' health by rearranging the display of foods in such a way that they will be encouraged to take relatively more of the healthier ones; these foods will be made, say, more conspicuous and easier to take away. Thaler and Sunstein do not cite any experimental evidence that straightforward changes of display had any effect under similar circumstances, [21] but we will apply our semantic convention by assuming that they meet with some success. A changed consumption is not unlikely to take place if the customers did not previously notice the healthier food, but then the intervention is best analyzed as a rational adjustment to new information, and this would defeat nudge 2. Although this is less likely, let us say that some consumers who previously knew about the healthier foods will modify their consumption. Since the intervention neither adds nor subtracts any options, nudge 1 holds by condition (i). But it is again dubious that nudge 2 does. If the change in display is perceived to be a well-intended suggestion from the organization, it may be rational to follow suit. Also, Thaler and Sunstein's narrative suggests that the change in display leads to a weaker request of visual or prehensive effort, and if the selection of the healthier food responds to that change in incentives, we cannot see where the bounded rationality feature lies. Paradoxically, a marginalist "Econ", who is always on the alert, is more likely than an inert "Human" to appreciate the slight signal. There are distinctive reasons for doubting nudge 3 here – since a spontaneous choice for unhealthy food may not involve any lack of rationality - but Thaler and Sunstein's example has already collapsed with the failure of nudge 2.

Are the more realistic examples of *Nudge* any more convincing? First, consider Save More Tomorrow (SMT). This is an ingenious saving scheme that commits the participants to increase their savings at future dates, with the added contributions timed to coincide with the participants' later raises in earnings. On the presumption that no other option leaves the set, SMT makes a net addition, so the intervention qualifies as nudge 1 by (i). This of course only holds *ex ante*, since SMT severely reduces the option set *ex post*, but let us concentrate on the former point of view, examining if nudge 2 would not also hold. To give maximum chance to

[21] Marketing evidence exists, but is not directly applicable, because it usually considers the effect of changing price as well as the display (as in "loss leading" strategies).

the example, we assume that the agent who adheres to SMT was already willing to save in the future exactly as SMT procures; in this way, we block an all too easy rationalization in terms of the new information and changing weights of argument brought about by the added option. Then we can think of only two (nonexclusive) reasons why the agent commits to the plan: either he means to save future deliberative and transaction costs, or he is worried that he could make a different decision when the time comes and voluntarily restricts his future freedom of choice. In either case the agent seems to us to be rational. This is clear if he is moved by the first reason, and more questionable if he is moved by the second, but we take the classic view that it is a rational attitude to bind oneself against temptation. It does not help claiming that time inconsistency belongs to bounded rationality because - regardless of whether or not this claim is theoretically justifiable - time inconsistency is *not* the psychological mechanism that prompts the choice of SMT; it is the preventive attitude towards it which does, and this is of course a different concept, both in terms of meaning and extension. Thus, when viewed *ex ante*, the intervention is nudge 1 but not nudge 2, and whether it is nudge 3 depends on the normative analysis of time-inconsistency, which we have left open. [22]

Second, take the imposition of a withdrawal or "cooling-off" period to some consumers or borrowers. Many countries have made this mandatory in cases where the agent may be influenced (as in door-to-door sales) or is about to commit large amounts to money (as with mortgage loans). The obvious rationale is to avoid impulsive and unreflective decisions. This is bounded rationality at least in a loose sense, but the intervention clearly makes it *the object of the intervention, not the employed means*, i.e., it is nudge 3 and not nudge 2. This is one of several cases in which Thaler and Sunstein have fallen into the equivocation mentioned in section 2. The intervention reduces the agent's choice set dramatically, so it is not nudge 1 either. It would not do saying that the agent's decision is only postponed; this is loose wording because possible decisions differ according to the time, and the current decision to buy is clearly blocked. So we have turned another of Thaler and Sunstein's nudge examples into a counterexample, and regrettably of the worst possible kind, because it satisfies neither of the two important senses. Interestingly, the authors seem to have been reluctant to include

[22] This discussion is sufficient as far as the analysis of SMT in *Nudge* is concerned. However, Benartzi and Thaler (2007) and Benartzi, Peleg and Thaler (2012) have a more elaborate treatment of SMT, which may also be considered.

withdrawal periods in their list, recognizing that they include a ban and that this contradicts nudging. [23]

Third, consider Thaler and Sunstein's recurring advice of implementing RECAP (Record, Evaluate, Compare Alternative Prices) into commercial law. Agents taking out insurance, contracting for mortgage loans or even making basic utility subscriptions are likely to be faced with too complex information for them to process it entirely well. What the authors call RECAP is a regulation of disclosure practices; it requires sellers to translate the technical and financial data of their products into information that is pragmatically relevant to consumers. Physical units will be turned into money figures, one-shot payments will be separated from regular ones, renewal clauses and exit penalties will be stated transparently, the delays to get the benefits or other time constraints will be estimated, and so on. To give legal force to RECAP is to impose bans, but on producers, not on consumers, who are the only concern here. Does this intervention satisfy nudge 1? Since the option set remains the same, it does by (i) alone. Observe that the economic incentives may change heavily as a consequence of the new information provided; for example, the agent may find out that the contract has too many exclusions, once they are laid bare to comply with RECAP. But we have given (i) lexicographic superiority over (ii) precisely to settle conflictual cases like this one.

Having been satisfied with nudge 1, we turn to nudge 2, and now difficulties arise. On one reading of RECAP, companies simply provide *new information* to the agents, and it is impossible to infer from an observed change in decision that they are boundedly rational, since they could have applied some rational updating procedure, be it Bayes's rule or something else. The other reading, which is no doubt Thaler and Sunstein's, is that companies will present the same information differently, that ideal consumers could make the translation by themselves, but that real consumers need assistance, because they fall prey to biases or simply intellectual laziness. Which of the two readings is the better?

One argument for the informational reading is that RECAP will typically force companies to adduce extra information. For example, if there exist some obscure legal means to cancel a tacitly renewable subscription beyond the relevant withdrawal period, RECAP could require

[23] See the inconclusive discussion in Thaler and Sunstein (2009, p. 253-254). By contrast, Sunstein and Thaler (2003, p. 1186-89) mentioned mandatory cooling-off periods as if they were an unproblematic example of libertarian paternalism.

the seller to explain it, and this would enrich the average consumer's legal knowledge. However, this argument suggests dividing concrete RECAP arrangements into two groups, one satisfying the former reading, and the other the latter, depending on whether or not extra information is effectively added. The former group of arrangements is presumably the larger, but it is enough for Thaler and Sunstein's case if the latter group is non-empty, and this is also presumably the case.

A more subtle argument for the informational reading is that even though RECAP only requires an exercise in translation, such a process requires technical knowledge that it would be unfair to regard as being part of rationality. Consider the apparently simple task of taking a loan from a bank and agreeing with it on an amortization plan. In most countries, banks are legally required to apportion interest repayment and principal repayment in the amounts due by the customer, and they typically provide various amortization charts that differ in both this apportionment and the duration of the loan. This is an example of a RECAP translation. Now, there seems to be no irrationality involved in the fact that the borrower needs the bank's accountants to make the calculations. We agree that this is a slippery argument. What part of mathematics should be included in the mental equipment of a rational agent? Only simple arithmetic or also the use of the exponential function? Depending on whether one stops early or late along the slippery slope, one will prefer the informational or the bounded rationality reading of RECAP.

If we cannot reach any firm conclusion here, this in part because *framing* is the mechanism invoked by the bounded rationality reading, and it is the most elusive bias of all. We will not use the objection that it is difficult to base an intervention on framing, because we have decided not to object to Thaler and Sunstein on the grounds that some of their interventions are instrumentally dubious. There is a more basic problem that is right in our scope: if the bounded rationality reading applies, framing enters *the objective and not the means* of the intervention; so RECAP is again a case of nudge 3 without nudge 2. Suppose that the customers' idea of the loan was distorted by some framing effect and the banker is led by RECAP to redescribe the loan in such a way that this effect vanishes. It is then the banker's knowledge of framing, not the customers' framing, which shapes the intervention. Things would be different if the banker exploited a framing effect to make the new picture successful, but nothing in Thaler and Sunstein's discussion suggests that RECAP permits maneuvring in

this way. They recommend putting it into commercial law, and they would have a hard time making this case if it were different from a plain corrective device.

As a fourth example, consider the across-the-board recommendation in *Nudge* to include a default option into the choice set. [24] To give maximum chance to nudge 1, let us assume that either a preexisting element is turned into a default, or a new element having this property is added, with all the other options still available. Now, under this assumption, the choice situation is altered as follows: *the agent has the choice between choosing and not choosing*, and if he makes the first (meta) choice, he will be faced with exactly the same (basic) option set as before. The agent is now faced with a decision tree, with the previous option set becoming relevant in only one of the two branches. It is impossible to claim condition (i) of nudge 1, since the structure has been upset. The new options correspond to whole branches, that is to say, (not choosing, default option) and (choosing, chosen option). This is one of those cases of incommensurable option sets, like the insurance example, in which the supplementary condition (ii) must be checked. How do the economic incentives score in the new structure? If we keep to our maxim in section 2 of interpreting them literally, they have not changed. There is of course now an incentive for the agent not to invest as many cognitive resources as he did earlier, but this incentive change is outside scope. We conclude that nudge 1 holds.

Psychological biases play a definite role with default options. The decision failure they are meant to overcome are not so much incorrect deliberate choices as the stubborn adherence to the previous choices, whatever they are. That is, the inclusion of a default serves to counter the biases of availability, representativeness, and anchoring and adjusting, all of which encourage the agent's inertia. This is enough for the intervention to count as nudge 3. Does it also satisfy nudge 2? It would be farfetched to claim that the same biases now push the agent towards the default option. For there is no reason to think that representativeness applies (the default can well be an outlier) and anchoring and adjusting is off the mark (there is no adjustment to be made in the (not choosing, default option) branch). At best availability - a

[24] Normatively, they recommend it across the board for complex financial and they review the organ donation problem in this light; they could have also discussed class action, an area in which the matter of default options looms large. Class action systems are divided into those with opting out and those with opting in, and this distinction corresponds to two ways of setting the default option.

loosely defined bias – can be at work. [25] This would be a shaky defence of nudge 2, whereas a trite and apparently persuasive idea presents itself: default options make it possible for the agent to save an intellectual effort. But on second thoughts, this idea is the Trojan horse of classical rationality. Suppose that the agent regards the default option as embodying useful specialized knowledge that he does not have, and believes that he is unlikely to make a better choice by himself, although this would cost him research costs for sure. Then, to select (not choosing, default option) is a *dominant strategy*. Bounded rationality does play a role in this explanation, but as a consideration in the (meta)decision between choosing and not choosing, and by no means as a factor influencing the way this decision is made. The explanation presumes that the agent trusts the default - or rather the party offering it - to a sufficient extent, and it is easy to think of empirical cases where this does not apply, but one would expect the agent to be more attracted by (choosing, chosen option) in those cases, and our discussion hypothesizes a successful intervention. In sum, nudge 2 is dubious, the appearances notwithstanding, and we rather favour the view that it does not hold. As it seems, Thaler and Sunstein have again mistaken nudge 3 for nudge 2.

The table sums up the negative results of this section.

	Nudge 1	Nudge 2	Nudge 3
Cafeteria example	yes	no	unclear
Save More Tomorrow (SMT) - *ex ante*: - *ex post*:	 yes no	 no no	 unclear unclear
Withdrawal periods	no	no	yes
Record, Evaluate, Compare Alternative Prices (RECAP)	yes	no	yes in one interpretation

[25] Gregariousness may be invoked in some cases.

Default options	yes (choice set: not applicable; incentives: yes)	no	yes

The typical pattern is yes, no, and an often qualified yes. A good deal of the discussion hinged on the question of whether one accounts for successful nudging better in terms of classical or bounded rationality, and we have inclined towards the former way in every contentious case. The literature has sometimes noted the duality of interpretation but never investigated it in detail. [26] Perhaps the presumption was that it automatically led to the large unresolved problem of whether bounded rationality stood by itself theoretically or could be reduced to classical rationality, with cognitive and practical limitations playing the role of added constraints. [27] We have not illuminated this problem, but did not have to do so, because it was enough to argue for classical rationality explanations *directly*, without the need for reducing bounded rationality explanations to them; in fact, we view the latter as being too shaky for the reduction strategy even to be considered. Notice that we have abiden by our methodological decision of hypothesizing successful interventions. This is a concession to Thaler and Sunstein in one way, but it has facilitated the present critique in another way, because classical rationality often explains success more commonsensically than bounded rationality.

4. Back to the normative

The normative discussion initiated by Thaler and Sunstein is primarily geared at libertarian paternalism, with two overlapping lines of inquiry. The former questions Thaler and Sunstein's understanding of the two doctrines they claim to reconcile, as well as related theoretical issues in moral philosophy. The latter, which is more applied, addresses libertarian paternalism from the angle of nudging interventions. This line is actually closer to Thaler and Sunstein's own approach to moral notions, which basically consists in studying them on particular examples. They argue that their privileged cases of nudges satisfy the requirements of both liberalism and paternalism, granting that, conversely, these examples help specify the

[26] In particular, Hausman and Welsh raise the problem but decide not to pursue it. Here is how they formulate it: "Why should these factors be regarded as *interferences* with rational choice rather than as rational determinants of choice?" (2010, p. 126).
[27] This important discussion is connected with the work of Simon (1982).

ill-defined requirements. In effect, they put the familiar reflective equilibrium method at work, and what can be questioned is not their appeal to the method but the equilibrium it allegedly delivers.

The first line has led to a plethora of objections. One of them concerns Thaler and Sunstein's idiosyncratic view of paternalism, which, unlike the more common one, does not include coercion as a necessary component. [28] What they retain from paternalism once coercion goes is by and large the approval of interventions that aim at improving another party's welfare even if that party does not acquiesce in the change. By reflective equilibrium, the cafeteria example is supposed to confirm this and related intuitions. The objection is that paternalism cannot be defined so widely and that coercion - although not an entirely clear idea - should somehow be included to make the relevant restriction. [29]

Along the same abstract line of inquiry, a symmetric objection is that Thaler and Sunstein conceive of liberalism too laxly. Some criticisms amount to arguing that liberalism, in a genuine sense, requires a deeper notion of freedom than just freedom of choice, which is the only one considered. Other criticisms amount to endorsing freedom of choice as a sufficient criterion of liberalism, while complaining that Thaler and Sunstein handle it inadequately. Both groups of criticisms feed on what may be called *the anti-manipulation claim*: liberalism permits interfering with the choices of another party only if that party has sufficient knowledge and understanding of the interference. Both groups of criticisms take for granted that Thaler and Sunstein fall prey to the anti-manipulation claim. Indeed, it seems to go without saying that bounded rationality can be exploited most effectively when the agents are not fully aware of the intervention ("in the dark"). [30]

---

[28] "The second misconception is that paternalism always involves coercion" (2009, p. 11; already in 2003a and b). The "first misconception" they consider is that "it is possible to avoid influencing people's choices" (2009, p. 10) – the *inevitability argument*, which has raised many objections in the literature.

[29] This objection is strongly made in Hausman and Welch (2009, p.127), who draw upon Dworkin's (1972) classic statement of paternalism, in which the coercive element looms large. Grüne-Yanoff (2012) seems to accept the same view although his argument that nudges are coercive does not logically require it.

[30] See in particular Mitchell (2005) and Grüne-Yanoff (2012). The former critic reinforces the present liberal objection with a redistributive objection that we do not examine here (being intended for the less rational, nudging interventions would impose a cost on the more rational). This added objection connects with the idea of "asymmetric paternalism" in behavioural economics (Camerer et al., 2003).

These are complex objections, and they may involve dubious or even faulty intermediary steps even if they point to well-identified problems. We certainly agree that Thaler and Sunstein stretch the definitions of liberalism and paternalism just in order to make them more easily compatible. However, regarding the issue of coercion, some recent metaethical work rescues them by denying that paternalism entails it, [31] and regarding the issues of freedom and freedom of choice, part of the old liberal tradition supports purposefully restrictive construals not unlike theirs. The pivotal claim that interventions exploit bounded rationality better "in the dark" does not hold generally, as a quick comparison with Thaler and Sunstein's list will readily suggest. It is flatly contradicted by the self-commitment devices of the second group, does not work well for the gregarious devices of the third group, and is not even well supported within the first group. We have distinguished heuristics leading to biases, which can be criticized, and pure biases, which are more inflexible, and thus unlikely to be put out of work just because the intervening party would become more transparent about their strategic use. Whether the pivotal claim is true or false, it ceases to be an objection if it turns out that Thaler and Sunstein's main examples do not satisfy nudge 2, as we have argued. [32]

By and large, the work of the previous sections provides *a shortcut to the conclusion that Thaler and Sunstein have not reconciled liberalism and paternalism*. However they define these doctrines precisely, their more substantial case for reconciliation hinges on the claim that well-designed interventions can be both nudge 1 and 2. The former carries the liberal component, and the latter - which specifies the technology – the potential of reconciling it with the paternalist component (which is often, if not always, captured by nudge 3). Supposedly, by exploiting bounded rationality, an intervention can remain liberal while being paternalist; this is the single most important insight in their work, and their misuse of the moral concepts and words does not matter too much if this insight fails. Admittedly, our shortcut is effective only as an *ad hominem* refutation: it may be that liberalism and paternalism can be reconciled without referring to nudges at all. However, the other critics have not given an absolute proof either that reconciliation was impossible, since their own argument also depended on how Thaler and Sunstein themselves argued for it.

[31] See the works collected by Coons and Weber (2013), and the thorough review made in their introductory essay. Shiffrin (2000) has extended the paternalism concept to the point of denying that it necessarily involves benevolence.

[32] Similarly, the sophisticated discussion of transparency launched by Bovens (2008) becomes less pressing.

Along the second line mentioned above, the literature has assessed nudges morally, regardless of the metaethical problems, and here again, the general opinion is negative. We distance ourselves from this critical line because too much in it depends on invoking the antimanipulation claim, the impact of which our analysis can deflate in two ways. On the one hand, the self-commitment devices rely on the agent's actively collaborating in the intervention, so that the antimanipulation claim can hardly be objected to them. More generally, various devices turn out to rely on choices that are best analyzed by classical rationality under the assumption that agents take the reality of intervention into account; hence what has been argued for self-commitment can be extended to the devices in question. On the other hand, our brief discussion of biases suggests that some of them are so much part of the mental fabric that knowledge and understanding of the interference would not really change their causal effects. Hence *if* the devices really employed these biases, they would also be immune to the antimanipulation claim. Notice the curious fact that opposite assumptions about rationality can make the antimanipulation claim ineffective; notice also that we do not mean to question it normatively, but only to stress that it does not help to build a real case against nudges.

It would be disappointing if the paper contributed only to reinforcing the general rebuttal of Thaler and Sunstein, but the last paragraph has struck a positive note in casting doubt on the common claim that they support an unpleasant kind of manipulation. When it comes to state intervention, their liberal critics seem to be taken in the following dilemma: either accept our conclusion that many nudges are in fact innocuous, or extend the objection to the more standard interventions, like corrective taxes and transfers, which are explicitly planned in terms of rational incentive schemes. By and large, nudges should not be a specific source of worry for liberalism.

Having tried to pay justice to some, not all of the normative objections raised against Thaler and Sunstein, we praise them for having introduced new concepts that can help refine the practice and extend the scope of interventions. These concepts needed reformulating, which was one of our tasks here. With the proposed adjustments, nudge 1 can be used as criterion for a light intervention, if one has any need for such a criterion. The disappointing result that nudge 2 is of little effective use leaves the role of nudge 3 untouched, and we would like to conclude by emphasizing its special relevance. The simplest idea behind nudge 3 is that

rationality failures at the individual level set a task for public interventions, which is conceptually distinct from, and practically complementary with, the more familiar task assigned by collective rationality failures, like tax evasion, insurance market failures, overexploitation of natural ressources, and the like. Education offers a heuristics for the former group of interventions that Thaler and Sunstein were wrong to neglect or downplay. The corrective purposes of nudge 3 can often be realized by enlightening the agents just to the right extent, and by providing them with easy self-control devices, two methods that are very familiar to educators. This sounds like conventional wisdom compared with the teasing idea of the bounded rationality acting both as an ill and a remedy, but the teasing idea has unfortunately not resisted the examination of this paper.

References

Benartzi, S. & Thaler, R. (2007), "Heuristics and Biases in Retirement Savings Behavior", *Journal of Economic Perspectives*, 21, p. 81-104.

Benartzi, S., Peleg, E. & Thaler, R. (2012), "Choice Architecture and Retirement Savings Behavior", in Shafer, E., ed., The Behavioral Foundations of Policy, Princeton: Princeton University Press.

Bovens, L. (2008), "The Ethics of Nudge", in Grüne-Yanoff, T. & Hansson, S.O. (eds) *Preferences Change: Approaches from Philosophy, Economics and Psychology*, Berlin: Springer.

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), p. 1211-1254

Chamley, C. (2004) *Rational Herds. Economic Models of Social Learning*, Cambridge: Cambridge University Press

Coons, C. & Weber, M. (2013), *Paternalism: Theory and Practice*, Cambridge, Cambridge University Press.

Dworkin, G. (1972), "Paternalism", *The Monist*, 56, p. 64–84.

Grüne-Yanoff, T. (2012), "Old Wine in New Casks: Libertarian Paternalism Still Violates Liberal Principles", *Social Choice and Welfare*, 38, p. 635-645.

Hausman, D. & Welch, B. (2010), "To Nudge or Not to Nudge", *The Journal of Political Philosophy*, 18(1), p. 123-136.

Heilmann, C. (2014), "Success Conditions for Nudges: A Methodological Critique of Libertarian Paternalism", *European Journal for Philosophy of Science*, 4, p. 75-94.

Kahan, D.M. (2000), "Gentle Nudges vs. Hard Shoves: Solving the Sticky Norms Problem", *University of Chicago Law Review*, p. 607-645.

Kahneman, D. (2011), *Thinking Fast and Slow*, New York: Farrar, Strauss, Giroux

Kahneman, D., Slovic, P. & Tversky, A. (eds) (1989), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press.

Kahneman, D. & Tversky, A. (eds) (2000), *Choice, Values and Frames*, New York: Cambridge University Press.

Machina, M. (1989), "Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty", *Journal of Economic Literature*, 27, p. 1622-1668.

McClennen, E. (1990), *Rationality and Dynamic Choice*, Cambridge: Cambridge University Press.

Mitchell, G. (2005), "Libertarian Paternalism is an Oxymoron", *Northwestern University Law Review*, 99, p. 1245-1277.

Qizilbash, M. (2012), "Informed Desire and the Ambitions of Libertarian", *Social Choice and Welfare*, 38, p. 647-658.

Ryle, G. (1949), *The Concept of Mind*, Chicago: Chicago University Press.

Selinger, E. & Whyte, K. (2011), "Is There a Right Way to Nudge? The Practice and Ethics of Choice Architecture", *Sociology Compass*, 5, p. 923-935.

Shiffrin, S. (2000), "Paternalism, Unconscionanbility Doctrine, and Accommodation", *Philosophy and Public Affairs*, 20, p. 205-250.

Simon, H.A. (1982), *Models of Bounded Rationality*, Cambridge, Mass., MIT Press.

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

Sunstein, C. & Thaler, R. (2003a), "Libertarian Paternalism Is Not an Oxymoron", *The University of Chicago Law Review*, 70(4), p. 1159-1202.

Sunstein, C. & Thaler, R. (2003b), "Libertarian Paternalism", *American Economic Review*, 93, Papers and Proceedings, p. 175-179.

Thaler, R. & Sunstein, C. (2008) *Nudge*, Yale: Yale University Press ; updated edition New York: Penguin Books, 2009.

Tversky, A. & Kahneman, D. (1974), "Judgment Under Uncertainty: Heuristics and Biases", *Science*, New Series, 185(4157), p. 1124-1131.

Tversky, A. & Kahneman, D. (1981), "The Framing of Decisions and the Psychology of Choice", *Science*, 211(4481), p. 453-458.

Wakker, P. (2010), *Prospect Theory for Risk and Ambiguity*, Cambridge: Cambridge University Press.