POLICY AND SOCIETY
2019, VOL. 38, NO. 1, 14–40
https://doi.org/10.1080/14494035.2018.1511188

Routledge
Taylor & Francis Group

ARTICLE

OPEN ACCESS Check for updates

# The challenges of behavioural insights for effective policy design

Colin R. Kuehnhanss

Department of Applied Economics, Vrije Universiteit Brussel, Brussels, Belgium

**ABSTRACT**

Behavioural insights are becoming increasingly popular with policy practitioners. Findings and methods originally provided by cognitive psychology and later behavioural economics have found use in the formulation of public policies. Their most popularised application has emerged under the auspices of libertarian paternalism in the form of 'nudging'. Its proponents claim to provide a new instrument to facilitate the formulation of effective and evidence-based policy, taking people's actual behaviour into account from the outset, while preserving their liberty to choose. This article reviews the origins of libertarian paternalism and the behavioural insights it builds on and takes a critical look at the foundations nudging relies on as a policy tool. It also discusses the ongoing efforts to build policy capacity to integrate behavioural insights and experimental methods in the creation of public policy. Behavioural insights offer a powerful tool to reshape and design new evidence-based policy. However, designers ought to be aware of the underlying assumptions on individual behaviour, the broadness of the mandate the nudging approach claims and the challenges they pose for design effectiveness.

**KEYWORDS**

Libertarian paternalism; nudging; behavioural economics; heuristics and biases; policy effectiveness

## 1. Introduction

In designing policy, the behaviour of citizens has commonly been modelled under the assumption of rationality, providing a coherent framework for analysing the effect of policy interventions and to predict their social outcomes. However, this assumption and its role as the foundation for the analysis and prediction of behaviour in the social sciences are being challenged. People tend not to act like *Homines oeconomici* but face limitations in their mental resources (such as willpower, computational capacity and memory) leading to predictable biases. A large literature empirically explores these cognitive biases and the underlying mechanisms (see, for example, Kahneman & Tversky, 2000, for a collection of seminal works). An early example of the integration of its findings into policy analysis can be found in the establishment of behavioural economics and law, expanding the traditional economic analysis of policy interventions to include the positive findings of behavioural economics (see DellaVigna, 2009, for a review of the latter). In the currently most popularised application of behavioural

**CONTACT** Colin R. Kuehnhanss colin.kuehnhanss@vub.be Department of Applied Economics, Vrije Universiteit Brussel, Pleinlaan 2, Brussels 1050, Belgium

research to policy-making, Richard Thaler and Cass Sunstein promote 'libertarian paternalism'. As one of the tools available in this approach, they argue for the use of so-called nudges to correct what they perceive as 'errors' in people's choices through alterations to their choice environment and without restricting their (nominal) freedom to choose (Thaler & Sunstein, 2008). Nudging, and behavioural research more generally, has resonated well with policy-makers (see Schubert, 2017, on the political economy of nudges). The suggested use of behavioural insights, while not undisputed or free of ethical and technical concerns, promises to provide new and to improve on old instruments for policy-makers to design targeted and effective policies.

Parallel to the emergence of behavioural insights, policy design studies have been shifting from a focus on rational instrument decisions in abstract or idealised policy-making circumstances towards an appreciation of the contextual and real-life factors surrounding policy-making (Capano & Howlett, 2015). As Bali et al. (Forthcoming) discuss, to achieve design effectiveness, the designers need to be able to anticipate what will happen during policy implementation and the effects of the available and often-competing policy instruments to reach the intended policy outcome (see Howlett, 2000, 2005; May, 2012; Öberg, Lundin, & Thelander, 2015). In doing so, the instrumentality of policy interventions may be improved by applying behavioural insights throughout the different stages of the policy formulation process. However, the emergence of concepts such as libertarian paternalism, and the use of tools such as nudging, raises questions on the normative basis of policy interventions and the legitimacy of necessary adaptations to existing policy capacity for designing and implementing behaviourally grounded policy.

This review takes a critical look at the use of behavioural insights as a basis for the use of policy tools such as nudging and discusses caveats to their suitability for design effectiveness. Given the diverse streams of the literature involved in the debate, it must be stressed that it is selective in nature. It introduces some of the most discussed concepts and shows their relevance for policy design. It also highlights current efforts to increase the policy capacity of governments and bureaucracies to use the empirical methods of behavioural research to reshape existing policy tools. With the increasing interest of policy-makers, the use of behavioural insights in the design of policy will undoubtedly grow further. It is, therefore, important to discuss the difficulties that come with the implementation of behavioural insights and methods. In particular, policy-makers need to be aware of the distinct and often far-reaching assumptions on the origins and consequences of limitations in people's mental resources that 'value-free' and 'pragmatic' nudging (e.g. Chetty, 2015; Madrian, 2014) relies on.

## 2. The traditional view of behaviour in policy design

In the broadest definition, public policies are a political actor's main tool to prescribe or alter the behaviour of agents and citizens. They are devised in the context of existing laws, institutional and constitutional boundaries and other guiding factors such as the encouragement of active citizenship.[1] Policy design thereby refers to the centrally planned efficient and effective development of policies to reach the goals specified by

[1]See Page (2006) for a more differentiating discussion of the term 'policy' and its different levels of meaning.

governments.[2] In those instances when an active design process is followed, governments have various instruments at their command to implement policies. Their common denominator is a notion of control of collective action to reach specific targets and goals (Bobrow & Dryzek, 1987; Salamon, 2002; Schneider & Ingram, 1990). They are 'concrete and specified operational forms of intervention by public authorities' (Bemelmans-Videc, Rist, & Vedung, 1998, p. 4), targeted at the public at large, parts or subgroups of society, specific organisations or foreign targets. The consideration of certain instruments as a valid means thereby depends on the form of governance, ideology, cultural norms, socio-economic differences and public opinion. The same policy problem can be met with a range of possible instruments, and their perceived legitimacy will vary with circumstances (Howlett, 1991, 2005; Klein & Marmor, 2008; Salamon, 2002). At the same time, when trying to anticipate the effectiveness of a policy design, its designers need to be able to understand and predict the functioning of the available instruments in achieving the desired political and policy outcomes (Bali et al., Forthcoming). In particular for estimating the effects on the latter, the expected reactions and behaviour of the policy targets are crucial.

Importantly, the underlying assumption in using policy instruments has traditionally been the rationality of their targets. Rationality may be seen as substantive or procedural (Simon, 1986). Substantive rationality is only concerned with the goal a decision-maker may have and the external environment in which a relevant choice needs to be made. The void between the goal and the external environment is filled with a set of normative rules of choice (see, for example, Jones, 1999). In this line of thinking, rational choice theory, which can be traced back to Bernoulli (1738/1954) and has been formally axiomatised as Expected Utility Theory (EUT) by Von Neumann and Morgenstern (1947), has been the bedrock of decades of decision modelling in economics and political science.[3] EUT and its descendents take axioms such as complete, reflexive and transitive preferences as basis for a deductive analysis of the decision-making of self-interested agents.[4] In a critique of this approach, Herbert A. Simon (1955, 1957), among others, pointed to the empirical failings of EUT in describing and predicting people's actual behaviour. He introduced the concept of bounded rationality, which takes a procedural view on decision-making. People intend to act rationally but are constrained by internal limitations (such as capacity constraints to deal with information, limited predictive ability and inability to ascribe pay-offs to outcomes) leading them not to globally optimise their choices as under EUT but to reasonably approximate (or as Simon termed it to 'satisfice') their goals by meeting preset aspiration levels. The criticism of rational choice theories as being at odds with actual human decision-making is at heart a negative statement about the former. While the lack of a unifying theory of decision-making without rational choice has remained an important

[2]Of course, not all policies are the result of conscious design. Many are simply the outcome of ad hoc decision-making, bargaining or other 'non-design' processes (Capano & Howlett, 2015; Howlett, 2014). This article abstracts from these instances to the benefit of a deliberate design process.

[3]Views on the emergence of rational choice theory and its implications for the development of economics and political science differ between the disciplines. For a political science focused summary of common rational choice theories see Jones (2001). See Amadae and Bueno de Mesquita (1999) for a review of their development and use.

[4]The concept of the self-interested agent obviously predates EUT and is often traced back to Adam Smith's self-interested butcher, brewer and farmer (Smith, 1776/2010, p. 9-10) and to the works of John Stuart Mill on political economy. The term 'economic man' and its Latin form *Homo oeconomicus* arose in the early nineteenth century as criticism of such an understanding of human interests, see Oxford English Dictionary and Pareto (1906/2014).

caveat, the recognition of the disengagement between rational choice and actual decision-making has sparked interest in positive research on behaviour.

## 3. The emergence of behavioural insights

While behavioural insights, behavioural economics and nudging have become widely popularised, their potential applications to policy-making depend on the interpretations given to their respective findings and the assumptions used in translating empirical evidence on human decision-making to policy interventions. To appreciate the opportunities behavioural insights may offer, it is necessary to consider the development of the respective research and policy agendas and highlight the normative and positive underpinnings, respectively.

Initially, work in (cognitive) psychology identified anomalies in the decision-making of individuals (see Gilovich, Griffin, & Kahneman, 2002; Kahneman & Tversky, 2000; Loewenstein, Read, & Baumeister, 2003, for extensive collections and reviews). Early studies uncovered a range of heuristics, or mental shortcuts, people use to arrive at decisions when faced with complex situations or an overload of information. While it is important to note that such heuristics provide an important tool in complex environments and intuition can be 'powerful and accurate' (Kahneman, 2003, p. 1450), they can lead to bias. Three classes of biases are particularly relevant. First, people do not handle probabilities very well as demonstrated by, for instance, *anchoring, overconfidence* and *availability* biases (see, for instance, Kahneman, 1992; Tversky & Kahneman, 1974).[5] Second, they respond differently to the same information depending on how it is presented. A textbook example in the context of decision-making under risk is the presentation of outcomes of a heart surgery. Imagine your doctor tells you either of the following:

Five years after surgery, 90% of patients are alive.

Five years after surgery, 10% of patients are dead.

Multiple studies have found substantially more favourable responses to surgery in the survival frame than in the death frame (see Moxey, O'Connell, McGettigan, & Henry, 2003, for a review). Such *framing* can significantly alter choice behaviour and lead to preference reversals (Tversky & Kahneman, 1981, 1986, 1992). Third, people tend to make *time-inconsistent choices* by giving too much weight to the present and neglecting future events. They procrastinate, have a tendency towards inertia (to accept the environment in which they are as a fixed given), and show a status quo bias in making decisions (Kahneman, Knetsch, & Thaler, 1991; Samuelson & Zeckhauser, 1988). Kahneman and Tversky (1979) show that choices are dependent on a (according to Kőszegi and Rabin (2006, 2007) endogenous) reference point against which they are evaluated and that people feel losses more than gains. The status quo bias, also independent of time inconsistency, leads people to take the current state as reference

[5]The anchoring bias describes people's tendency of overweighting initial (and potentially irrelevant) pieces of information regardless of their probability of occurrence in making (numeric) judgements. Such behaviour has been shown to influence housing prices (Northcraft & Neale, 1987), consumer's willingness to pay for products (Ariely, Loewenstein, & Prelec, 2003) and salary negotiations (Thorsteinson, 2011). Overconfidence is mostly related with the overestimation of the accuracy of one's own estimates or of one's own performance. It is rooted in a bias in the estimation of subjective probabilities. The availability bias describes people's tendency to overestimate the probability of events for which they can easily recall an example. For instance, the rate of airplane crashes.

point and overweight the potential losses from switching to another state relative to the potential gains. As a result, the current state seems the preferable option.

Two outcomes of this line of inquiry must be mentioned here.[6] First, research on heuristics led to a better understanding of the processes underlying human decision-making. While intuition and heuristics help us to cope with complex environments in a 'quick and dirty' fashion, and yet allow us to reach good outcomes (Gigerenzer & Gaissmaier, 2011), we are equally capable of targeted and effortful reasoning. This duality has been explored in the psychology literature in two-process theories (Epstein, 1994; Evans & Over, 1996; Sloman, 1996). While differences in classifications exist (see, for example, Evans & Stanovich, 2013), the first process is often described as fast, automatic, effortless and emotional, whereas the second is described as slow, controlled, effortful and rule governed. They have been termed System 1 and System 2 thinking, respectively (Stanovich, 1999; Stanovich & West, 2000). Two-system approaches to reasoning, judgement and decision-making have become popular across disciplines (Kahneman, 2003; Loewenstein & O'Donoghue, 2004; Sanfey, Loewenstein, McClure, & Cohen, 2006). They are supported by empirical findings in neuroscience showing decision-making to be the outcome of an interaction of two (or more – see, for example, Kurzban, 2010) processes within the brain.[7] However, it is important to note that there are important caveats. For instance, System 1 should not necessarily be seen as unitary as it refers to multiple autonomous processes in the brain (see Evans & Stanovich, 2013). It is also the evolutionarily older one, suggesting an ability to work independently of the later add-on of System 2. Finally, and importantly, the discussion of the ability of the two systems to reason 'correctly' and make 'optimal' choices is itself reference point dependent. The assumption of rational decision-making itself is as much in need of a psychological explanation as deviations from it (see Infante, Lecouteux, & Sugden, 2016, for a discussion). I will return to the two-system theory of reasoning below.

Second, the insights and methods of this line of research have been adopted and advanced by behavioural economists.[8] Many behavioural economists concern themselves with relaxing assumptions required for strict interpretations of rational choice but do not challenge its function as a useful device in creating testable and refutable predictions about people's choices. The integration of, for instance, considerations of fairness (e.g. Blount, 1995; Charness & Rabin, 2002), reciprocity (e.g. Dufwenberg &

[6]Note that some authors categorise the described and other biases in terms of 'bounded rationality', 'bounded willpower' and 'bounded self-interest' (Jolls et al., 1998). The first two categories roughly mirror the distinctions made above, whereas 'bounded self-interest' captures aspects of altruism, fairness and reciprocity considerations (see below). Given that particularly the term 'bounded rationality' has been used differently across literatures, I will abstain from its use.

[7]In general, advances in clinical methods have provided new avenues for research on decision-making, moving from analysing the expressed choices of people to the underlying cognitive functions of the brain (Camerer, Loewenstein, & Prelec, 2005; De Martino, Kumaran, Seymour, & Dolan, 2006; Sanfey & Chang, 2008; Sanfey et al., 2006). See Felsen and Reiner (2015) for a specific discussion on the potential role of neuroscience in the debate on nudging.

[8]In a turn of tables, some economists argue that behavioural economics is simply a return to the origins of economic theory after venturing off into the neoclassical world of axiomatic rationality. As Camerer, Loewenstein and Rabin (2004, p. 5) put it:

> 'When economics first became identified as a distinct field of study, psychology did not exist as a discipline. Many economists moonlighted as the psychologists of their time. [. . .] For example, Adam Smith commented (1790/2005, p. 192) that "we suffer more ... when we fall from a better to a worse situation, than we ever enjoy when we rise from a worse to a better". Loss aversion!'

Kirchsteiger, 2004; Fehr & Gächter, 2000; Fehr, Kirchsteiger, & Riedl, 1998; Rabin, 1993) or other-regarding preferences (e.g. Sobel, 2005) leads to the expansion of existing theory within a framework of rationality. Behavioural economics uses a positive approach to the behaviour of individuals. While there are concerns of rational choice proponents, that '[…] in theory-making, descriptive accuracy is purchased at a price, the price being the loss of predictive power' (Posner, 1998, p. 1559), exploring the behavioural basis of decision-making leads to better descriptive accounts that can be parametrised, tested (also against the benchmark of rational choice models; see, for example, Rabin, 1998) and eventually used to build new theory making equally stark but more accurate predictions (Rabin, 2013). Importantly, the adaptation of psychological methods in economics does not necessarily imply a normative statement on the validity of the use of rational choice models as such. Rather, as Angner (2015) argues, behavioural economists may rely on the rationality assumptions as an analytical construct and a normative standard against which new advances can be evaluated.

Besides the adoption of methods from other fields, the term *behavioural insights* has been coined to describe an approach taking a broader view integrating findings from other (social) sciences such as neuroscience, sociology and biology to the analysis of human behaviour and policy-relevant phenomena. It thereby often goes further than the definition of behavioural economics ('A method of economic analysis that applies psychological insights into human behaviour to explain economic decision-making', *Oxford Dictionary*),[9] both in its reliance on other fields and its interest in behaviour at least prima facie unrelated to economic decisions.[10]

## 4. Integration into policy

Following its emergence and absorption into behavioural economics, the heuristics and biases paradigm transpired into policy-relevant areas such as behavioural law and economics (BLE). In two seminal articles, Jolls, Sunstein and Thaler (1998) and Korobkin and Ulen (2000) argue for the integration of the positive analysis of behaviour in the formulation of law. With the application of psychology traditionally restricted to issues such as jury judgements, witness credibility and cases dealing specifically with people with mental handicaps, they propose that the behavioural approach should be applied to inform the effect and content of law on a broader basis. Notwithstanding strong criticism (often similar to the criticism of behavioural economics) of inadequate empirical bases (see Gigerenzer, 2015, and references therein), lack of predictive accuracy and a failure to provide a coherent theory (Posner, 1998), BLE has brought forth a large body of research. It includes works on interventions into the marketplace, risk regulation (and the institutional set-up responsible for it), the justice system and the interrelation between law and social norms (Rachlinski, 2011). On a normative level, especially Jolls et al. (1998) open the debate on whether the susceptibility of people to cognitive biases offers ground for rethinking the role of paternalistic intervention in policy.

[9]Source: https://en.oxforddictionaries.com/definition/behavioural_economics, accesses 27 February 2018.
[10]In the same vein, and not surprisingly given the origin of the research agenda, Daniel Kahneman (2013) has argued for the use of the term *applied behavioural science*.

Policy-making (most dominantly in the United States) is invariably intertwined with the (partisan) debate between libertarian and paternalistic approaches to government. In an attempt to offer a 'third way' between the classic libertarian and paternalist arguments, Richard Thaler and Sunstein (2003) have brought forward the concept of libertarian paternalism. In their eyes, it combines the preservation of (nominal) freedom of choice valued by libertarians while simultaneously allowing '[…] for self-conscious efforts, by institutions in the private sector and also by government, to steer people's choices in directions that will improve their lives' (Thaler & Sunstein, 2008, p. 5). In their words (Thaler & Sunstein, 2003, p. 175),

> […] a policy counts as 'paternalistic' if it is selected with the goal of influencing the choices of affected parties in a way that will make those parties better off. We intend 'better off' to be measured as objectively as possible, and we clearly do not always equate revealed preference with welfare. That is, we emphasize the possibility that in some cases individuals make inferior choices, choices that they would change if they had complete information, unlimited cognitive abilities, and no lack of willpower.

In the popularisation of the concept in their 2008 book 'Nudge', they abandon the external and objective measurement of welfare and state the paternalistic aspect of libertarian paternalism to be legitimised when an intervention will 'make choosers better off, as judged by themselves' (Thaler & Sunstein, 2008, p. 5). The libertarian aspect of their programme is satisfied as long as the set of choices is not reduced and any intervention is easily reversible.

### 4.1. *Conceptual concerns*

Libertarian paternalism has been met with a wide range of criticisms and concerns, among the most fiercely by libertarians who do not agree with its claim to be choice preserving. In their view, the existence of cognitive biases at most warrants policies aimed at enhancing autonomy and at preventing biased decisions which may lead to the restriction of liberty (e.g. Mitchell, 2005). Interference towards any other means or end, both of which the proposals by Thaler and Sunstein (2008) entail, are in contradiction to the value libertarians place on autonomy, the individual's freedom to make (even suboptimal) choices and the capacity to develop as a result (see Rebonato, 2012, for a brief discussion). While the libertarian aspect of libertarian paternalism seeks not to limit the *nominal* freedom to choose by leaving the choice set unrestricted, the *effective* freedom of individuals under libertarian paternalist interventions is a different question further discussed below. The definition of paternalism Thaler and Sunstein (2008) rely on also differs from most standard definitions. Typically, paternalism requires (i) the interference with the liberty or autonomy of an individual, (ii) the absence the individual's consent and (iii) a (believed) benefit to that particular individual (Dworkin, 2017). As will also become apparent below, many interventions Thaler and Sunstein (2008) propose do not fulfil those criteria – for instance, by failing to benefit the interfered with individual – and therefore cannot be seen as paternalistic. As Dworkin (2017) argues, 'We could replace "paternalism" with "benevolence" and nothing important would be left out since the "libertarian" aspect picks up everything else that is significant'. Besides these definitional issues, conceptual concerns arise with

libertarian paternalists' understanding of welfare and with the reliance on the two-system account of reasoning to justify interventions.

The reliance of Thaler and Sunstein (2008) on their argument that an intervention is acceptable if the outcome is beneficial to the targets 'as judged by themselves' is both appealing as a political argument for policy and potentially contentious for any policy designer (Sugden, 2016).

Under the assumption of rationality, people are unitary actors seeking to satisfy their preferences, which are taken to be complete and consistent (see Saint-Paul, 2011, for a discussion). To give an oversimplified example in the context of intertemporal choice: A person may both want to eat healthily to enjoy a long life and really want to have that second cupcake now. The utility from the future outcome is weighted today according to the individual discount factor the person applies and the temporal distance (see Frederick, Loewenstein, & O'Donoghue, 2002, for a review on intertemporal choice models). This utility is compared to the immediate gratification from having the cupcake. The observed choice of the person to either have the cupcake or not is the revealed preference indicating the result of this decision-making process. Based on the revealed preference, any externalities created[11] and the welfare implications, the question whether policies should be instituted to interfere with the satisfaction of a given preference in a paternalistic manner is one of weighting normative arguments. In classic welfare economics accounts, the designer would be seen as a benevolent planner concerned with maximising the well-being of the entire society.[12] The degree of interference chosen is then a question of, for instance: whether the target for paternalism is knowledgeable about (the consequences of) an action and carries it out voluntarily (hard vs. soft paternalism); whether the paternalist interferes with an individual's actions directly or indirectly by targeting a third party (pure vs. impure paternalism); and the identity of the paternalist (broad vs. narrow paternalism) (Dworkin, 2017). In the case of the second cupcake, the resulting interference could, for instance, take the form of prescribing, incentivising or providing information on healthy eating. However, behavioural research has shown that people discount future events time inconsistently and are present biased (O'Donoghue & Rabin, 1999, 2000), that is, they discount future events at a rate they come to regret and disproportionally tend to prefer immediate gratification over delayed (but larger) benefits (Frederick et al., 2002). The future self comes to regret the decisions of the present self. Thaler and Sunstein (2008) argue that people's decisions thus not only create externalities but also 'internalities'. They see the individual split into two selves along the lines of the two-process theories of reasoning: the automatic and emotional self of System 1, which operates unconsciously, and the self-aware and reflective self of System 2. In their view, the bias-prone System 1 decisions negatively affect (i.e. have internalities for) the reflective System 2, which, given the chance, would have come to non-biased decisions. Importantly, internalities not only occur in an intertemporal context but also for any biased decision. Such a 'market failure', they argue, gives justification for the

[11] In the case of an unhealthy diet, for example, the costs of increased health care spendings to treat obesity and related health issues.

[12] See Sugden (2013) for a discussion of this point of view in welfare economics and of a contractarian alternative in which welfare maximisation results from mutually beneficial agreements between individuals without a central planner.

intervention of the social planner to not only maximise social well-being but also arbitrate inconsistent behaviour or incoherent preferences within the individual.

The meaning of 'better-off, as judged by themselves' thus deviates from the classic welfare considerations for paternalism both in the understanding of 'better-off' and in the novel introduction of 'as judged by themselves'. Starting from the latter, in a 'behavioural' world, people may not necessarily have complete and consistent preferences.[13] Preference reversals are a good example. If a choice is dependent on the frame, then it cannot be seen as indicative of the satisfaction of an ex-ante distinct preference. The absence of complete and consistent revealed preferences can be dealt with in different ways. First, it can be seen as a measurement problem, where welfare-relevant preferences can be separated from those that may be irrelevant to the social planner in improving social well-being. Such preferences could potentially be measured by focusing on unambiguous choices or choices with multiple rationales, reducing the influence of any single bias (see, for instance, Bernheim & Rangel, 2009; Grüne-Yanoff, 2009, for a discussion). Second, it can be argued that 'true' preferences exist and can be reconstructed despite the biases influencing decisions. If such 'purified' preferences are indeed context independent, they can serve as basis for welfare analysis (Infante et al., 2016). Along this line of thought, Thaler and Sunstein (2008) argue for the use of the decisions System 2 *would have made* as indicative of purified preferences, counteracting the 'sabotage' through the biases experienced by System 1. The question of 'better-off' then is one of satisfying those latent preferences.[14] As in the case of externalities, there are two parties whose concerns may be given different weights. It is in the interest of, say, the industrialist not to pay for the (non-taxed) pollution caused by a factory, but it is in the interest of the neighbours to have a clean environment. In the case of the internalities between the two systems, it is in the interest of System 1 to have that cupcake now and in the interest of System 2 to enjoy a healthy old age. Instead of applying a case-specific welfare analysis as for externalities, Thaler and Sunstein (2008) set the weight for System 1's interests to zero throughout. The conflict between the two selves in forming preferences is not solved but replaced by an assumption, which, in perfect circularity, requires System 2 to be (at least close to) the *Homo oeconomicus*: Only those preferences should be taken into account that System 2 reaches after being removed from stimuli, having ample time to deliberate and being able to rationally analyse the options (Infante et al., 2016; Rebonato, 2012; Schnellenbach & Schubert, 2015; Whitman & Rizzo, 2015).[15]

The use of the two-system process of reasoning in this context is problematic for multiple reasons. First, while moving people towards healthier lifestyles may seem like a

[13]Whether ex-ante preferences exist or preferences are elicited during the decision-making process is a separate point of debate (see, for instance, Fischhoff & Manski, 2000; Tversky & Thaler, 1990).

[14]See also the discussion by Qizilbash (2012) of the applicability of the informed desire account to the welfare considerations of Thaler and Sunstein (2008).

[15]Another objection lies in the understanding of preference ordering and individual-level maximising. Even if a policy designer wants to act benevolently, the two-system approach of Thaler and Sunstein (2003, 2008) aims at maximising the target's individual welfare (however defined), not at increasing the utility a rational target would maximise given the chance. The preferences a person holds, whether coherent or biased, are subjectively ordered and the utility a person gets from their fulfilment are impossible for a third party to judge or even presuppose. Without the instrument of revealed preferences, interventions are necessarily based on the interpretations of the planner which preferences are 'real' or meaningful and the chosen definition of welfare (Mitchell, 2005; Sugden, 2008).

good idea, individual welfare may be more dependent on the (emotional) contributions of System 1. As Mitchell (2005, p. 1268) argues,

> Many may agree in the abstract that better health is preferable to worse health, but when the choice is framed as enjoying life-shortening but intensely pleasurable vices during ones college days versus abstaining during college to gain a couple of extra boring years at an advanced age, then better health may not look quite as good.

Second, even if the preferences of the internal *Homo oeconomicus* are accepted as those that should be followed, it is not clear that they should serve as an argument for paternalism. Similarly, it is not necessarily clear that heuristics (i.e. listening to System 1) necessarily lead to inferior outcomes (Gigerenzer & Brighton, 2009; Gigerenzer & Gaissmaier, 2011). In the initial works of Tversky and Kahneman (1974), heuristics were seen as powerful devices. However, the biases they can invoke have been at the centre of attention in academic and popular writing, which, in turn, have increasingly been labelled as cognitive 'errors'. Finally, and importantly, without empiric justification, Thaler and Sunstein (2008) go much further than most psychologists, neuroscientists and behavioural science researchers in their treatment of the two selves. The latter generally see the two (or multiple) systems as far less clear-cut and as more interlinked and co-dependent. The now all too common approach to equate heuristics and System 1 thinking with error and suboptimal decisions overstates the original claims. As Evans and Stanovich (2013) state,

> Perhaps the most persistent fallacy in the perception of dual-process theories is the idea that Type 1 processes (intuitive, heuristic) are responsible for all bad thinking and that Type 2 processes (reflective, analytic) necessarily lead to correct responses.

Alternative accounts of heuristic decision-making see it as less automatic than the two-system view would suggest (see Kruglanski & Gigerenzer, 2011) and even as learned behaviour far removed from cognitive bias (Goldstein & Gigerenzer, 2002; Grüne-Yanoff & Hertwig, 2016). Instead of taking heuristics as a burden, a more fine-grained, context-dependent and case-by-case analysis may be warranted when designing policy.[16]

## 5. Nudging

Within the libertarian paternalist approach, Thaler and Sunstein (2008) suggest behavioural intervention as a means to achieve policy goals, so-called nudging. Thaler and Sunstein (2003, 2008) suggest that the environment in which people operate – the 'choice architecture' – is never neutral and influences their decisions. Whoever is in the position to alter the environment becomes a choice architect. Thaler and Sunstein (2008) use the, by now famous, cafeteria manager Carolyn to open their book and demonstrate the idea as follows: Carolyn is responsible for a large number of school cafeterias and for the way food is displayed to the students. She discovers that the way she arranges the different items can have a large effect on the frequency with which they are purchased by the children. She is now torn between different options. Should she

[16]There is an ongoing debate on debiasing as a preferable approach to nudging to maintain freedom of choice and limit paternalistic intervention in response to cognitive biases, see, for instance, Gigerenzer (2015) and Mitchell (2005).

simply distribute the items randomly, maximise the cafeterias' profits, sell off the best positions for bribes,[17] try to match the choices the children would have made themselves or arrange them so that they are best-off, all things considered.[18] Unsurprisingly, though not explicitly having Carolyn choose it, Thaler and Sunstein (2008) suggest the paternalistic option improving the kids' welfare to be the most appealing.[19]

A multitude of different nudges in various areas of life have been suggested so far. Two examples may illustrate their general approach: framing social norms and default rules. Environmental protection is an important topic, and energy consumption is a classic example for externalities difficult to regulate through policy. In an experiment to nudge people to reduce their electricity use, private households were sent information on their energy consumption during the preceding weeks. One part of the sample received factual information on their use, a comparison of their usage with that of their neighbours and information on how to reduce consumption. The other part received the same information with an additional visual stimulus in the form of a smiley face (happy if the consumption was below average and sad if it was above) to make the social norm more salient. The factual information led those who were above average to reduce their consumption but had a boomerang effect for those consuming below average. It signalled that they could use more and still do well in comparison. Including the smiley face eliminated this boomerang effect (Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007). The effects of such interventions seem to be substantial and comparable to rather high (in one instance an estimated 11–20%) short-run price increases (Allcott, 2011). However, they may be ineffective or even backfire if the underlying norms are not shared by the entire target group, as Costa and Kahn (2013) find when political conservatives in the United States were nudged to save energy.

Default rules are a good example of the notion that choice architecture is never neutral. Germany and Austria are arguably quite similar in terms of norms and culture. Yet, only about 12% of Germans are registered organ donors, whereas almost 100% of Austrians are signed up to have their organs used after their death. The most important influence on the decision whether or not someone volunteers their organs seems to be the default rule. When it is an opt-in scheme and people need to actively consent (even if only by ticking a box), the rate is much lower than it is with presumed consent in an opt-out scheme (Johnson & Goldstein, 2003).[20] One can argue in the case of organs that any effort to change the default may outweigh the benefit, given that one will not be affected by the choice any more when the time comes. However, similar mechanisms have been widely discussed for retirement savings, which one would hope to still be able to enjoy. Saving optimally for retirement is an analytically complex issue. It requires at

[17]This is an interesting proposition in a world Thaler and Sunstein (2008) construct to show that the intervention itself would be highly effective and argue that governmental choice designers ought to be trusted.

[18]Note that leaving the arrangement as it currently is does not feature as an option. Thaler and Sunstein (2008) argue that any arrangement is taken to have a large effect, and therefore, the current display cannot be neutral either.

[19]This starting point for an argument for paternalistic policies is interesting also for discussing an intervention on, of all possible target groups, children's choices. They are probably the least contentious group to argue for paternalistic intervention (as the origin of the word may suggest) and the least appropriate to be steered towards choices that make them 'better off, as judged by themselves' (Thaler & Sunstein, 2008, p. 5). For completeness, it should be mentioned that an earlier version of the example features adults as customers (Sunstein & Thaler, 2003). See Rebonato (2012) for a more detailed discussion of this and other examples.

[20]Note that this is one example where a nudge would not qualify as being paternalistic, given that it by definition will not benefit the interfered with individual.

least an estimation of the remaining lifespan, foresight of the circumstances one expects to be in – including the anticipated income flows and investment opportunities – and changes to one's preferences. Even if an optimal retirement plan exists and can be found (a technically daunting exercise), one also needs the willpower to stick to it. Two striking examples in which nudges can benefit the individual are the initial take-up of a savings plan and the adjustment of the savings rate over time. Madrian and Shea (2001) show that automatic enrolment for savings plans at US companies drastically increases the participation rate. Employees also tend to remain with the default allocation of funds to different investments. Inversely, Benartzi and Thaler (2007) describe a UK case in which half the employees fail to sign up for an opt-in savings plan that has the full contribution paid by the employer. The libertarian paternalistic course of action would clearly be to change the default rule to automatic enrolment but to leave the well-informed employee the option to opt-out if they prefer to for any reason. To increase the (retirement) savings rate once an individual has decided to take up a savings plan, Thaler and Benartzi (2004) propose the Save More Tomorrow™ programme. It is designed to help people overcome their lack of willpower by pre-committing to increases in the savings rate with future wage increases. They never see their *nominal* wage go down but consistently save more. While such pre-commitment achieves the goal of increasing the savings rate, it is disputed whether the exploitation of the common confusion between nominal and real wages is actually in the interest of the nudged individuals as their consumption preferences may be negatively affected (see Rebonato, 2012).

### 5.1. *Effectiveness concerns*

The approach of libertarian paternalism and the nudges derived from it are a very distinct form of applying behavioural insights to policy. It relies on the same empirical research as behavioural economics but takes a much more proactive role in interpreting the observed biases as a foundation for paternalistic intervention. In particular, it takes them as a mandate to implement policy to reach normatively defined goals through behavioural instruments. Unfortunately, Thaler and Sunstein (2008) are not always cautious about aligning their examples with their definitions, which themselves are not always clear-cut (see, for example, Hausman & Welch, 2010). For instance, they present the provision of information as a nudge in multiple cases (e.g. improved labelling to communicate the dangers of smoking or the fuel efficiency of cars). It is not always obvious how such nudges go beyond classic knowledge instruments (see, for example, Gigerenzer, 2015; Rebonato, 2012).[21] Even when the broad mandate as claimed by Thaler and Sunstein (2008) is accepted, it is worth highlighting effectiveness issues particularly relevant to the designer's decision to use nudges rather than to rely on established tools.

[21]See also Kuehnhanss (2018) for a discussion of nudging relative to information-based tools. The distinctions between different 'kinds' of nudges will not be discussed in this article. See, however, Baldwin (2014) and Hansen and Jespersen (2013) for categorisations and discussions of nudges according to their impact on the autonomy of the target and their transparency, respectively.

### 5.1.1. *Backfire and dependency effects*

In general, nudges may suffer from a range of methodological drawbacks. They may be inefficient because, by the same behavioural biases exploited by the nudges, they tend to be sticky. In the example of the default manipulation for retirement savings plans, an individual may get nudged into signing up, but without further intervention the default savings level may not be appropriate for that person. Because the default level used in the scheme reported by Madrian and Shea (2001) was rather low, it stands to reason that the savings rate decreased at least for some participants, reducing the benefit of the intervention. The savings level may also not be adjusted over time when an initial default is chosen due to inertia or lack of willpower. In another case, Fonseca and Grimshaw (2017) report experimental evidence that the effect of pre-populating tax forms as a means to reduce non-compliance is heavily dependent on the accuracy of the information quality used for the default values. When the defaults are set too low, they can significantly reduce compliance and hence tax revenue (see also Mullane & Sheffrin, 2012, for further examples of backfiring and ineffectiveness of nudges).

As Lunn (2014, p. 45) argues, 'Designing a good policy is made more awkward by the fact that the reasons why defaults are so powerful are not fully understood'. Defaults may work because they (i) are seen as a recommendation by the authority that determines them, (ii) signal the choices of others, (iii) help the procrastination or inertia of the chooser and (iv) act as a reference point against which other choices are evaluated. As long as it is not clear which factor drives the behaviour in each specific case, resulting policies may be suboptimal. From a broader perspective, De Haan and Linde (2017) show that when defaults lead to 'better' outcomes, they may create dependencies once people are used to them. When the default in a similar decision is then suboptimal (e.g. randomly determined), it may lead to worse outcomes overall. One potential solution is mandated choice. Forcing the chooser to make a decision may remove some of the concerns with hidden defaults. While it has been shown that mandated choice brings the number of organ donors closer to the rates observed under opt-out systems (Behavioural Insights Team (BIT), 2011), it can also have detrimental effects when the decision is more complex, as Sweden was unfortunate to discover when it left people the open choice between hundreds of different retirement funds (Thaler & Sunstein, 2008, p. 151 ff.).

### 5.1.2. *The equal incompetence assumption*

Thaler and Sunstein (2008) claim nudges to be choice preserving by leaving people the possibility to opt out or to make their own choices from an essentially unrestricted set of options. If you do not want to save for retirement, you can opt out of the plan. If you do not like the healthy fruit displayed in a convenient spot in the cafeteria, you can go looking for the sweet dessert. Mitchell (2005) argues that the assumption of rationality gets replaced by an assumption of equal incompetence: people need to be seen as systemically 'falling prey' to cognitive biases to justify the actions of a 'planner' to rectify the misled decisions of individuals to increase their (and social) welfare. However, not all people are equally rational or irrational and education, training and debiasing can substantially reduce the susceptibility to biases. Conversely, for someone who lacks a clear preference on an issue, Thaler and Sunstein's reassurance of leaving the choice set unrestricted is not convincing. When the proposed nudges are indeed as effective as

claimed, leaving an opt-out option for the more rational decision-maker will for most people be ineffective. The libertarian paternalistic nudge becomes simply a paternalistic intervention (Mitchell, 2005).

In the case of the cafeteria, Thaler and Sunstein (2008) suppose that it is generally preferable to live healthily and, therefore, have the fruit and not the dessert. Why this norm is superior in terms of individual welfare or necessarily and systematically economically more efficient is left open. Basing a nudge on norms, such as eating healthily not to become obese, can cause a conservative bias in policy-making. Social norms tend to stabilise the behaviour of people by inviting peer pressure on deviators. If a social norm is the basis for policy, it levies an 'emotional tax' on those who do not follow it (think, for example, of the negative societal response to obese people or smokers) without generating any form of tax revenue (Glaeser, 2006; Lucas, 2013). The reinforcement of existing, habitual and potentially welfare suboptimal social norms thus already lowers the welfare of those who dissent by increasing the allowable level of social pressure (Binder, 2014; Schnellenbach, 2012). It may also lower the welfare of the intended target. If a rational decision-maker, based on coherent preferences, has decided to want that dessert, the nudge imposes an additional cost. Thaler and Sunstein (2008) argue that interventions can only be counted as nudges if they are easily reversible. While they seem to be supportive of a cost–benefit analysis of paternalistic intervention proposed by Camerer, Issacharoff, Loewenstein, O'Donoghue and Rabin (2003), they leave unanswered the questions of what constitutes easy reversibility and how small the related cost must be to count as such, particularly if multiple small costs sum up (see Sugden, 2009).

Even if the cost is indeed negligible, the intervention may still have a redistributive effect. Assuming that the supposed social norm is accepted as a viable basis, a nudge may benefit some but still harm those who rationally have another preference. As discussed above, the goal of libertarian paternalists is to lead people to decisions they would have made if they had the full capacity to decide rationally. Redistribution away from the already rational, while potentially justifiable on moral arguments such as fairness, runs counter to both classic and libertarian paternalist notions of welfare maximisation (Mitchell, 2005; Schnellenbach, 2012).

Finally, an assumption of equal (in)competence may miss important influences of the socio-economic environment on the decision-making of individuals (Crawshaw, 2013). For instance, Mols, Haslam, Jetten and Steffens (2015) argue that the nudge approach misses the important dynamics of social identity and norm internalisation when tailoring nudges to individual behaviour. They urge for a recognition of the importance of social groups in defining and enacting norms.

#### 5.1.3. *Legal implications of nudges*

An important point in the regulation of behaviour in liberal societies is the preservation of the ability of the target to understand, adjust to and if necessary contest the regulation. Any restriction of freedom must be justified and should be as mild as possible (Aaken, 2015). While nudging promises to be a mild intervention, it raises concerns about the target's ability to know whether an intervention is taking place, how it works and what can be done against it. Thaler and Sunstein (2008) object to subliminal messaging and argue for a public disclosure of nudges. However, it is not

clear in how much detail a government should disclose its nudging activities. Is it sufficient to inform the public about the use of nudges in general, should the exact mechanism be disclosed (risking that the nudge becomes ineffective or even backfires) or are other legal safeguards warranted (Bovens, 2009; Lepenies & Małecka, 2015; Mols et al., 2015)? In either case, a nudge is aimed at altering choice behaviour when the target is not extending enough attention, deliberation and willpower to the decision at hand. In such a context, it is doubtful that the target will heed a nudge (Baldwin, 2014). The effect size of the nudge is also hard to gauge for the target. In the case of coercion or a tax, the effect is clear, but what part of not having that second cupcake can be ascribed to a true preference and what part to the nudge? As Aaken (2015, p. 94; footnotes omitted) argues,

> This is highly problematic and very different from the transparency requirements for state action. It impacts the rule of law to a considerable degree since law, and measures based thereon, must be public and accessible to those targeted or affected by the measure. This is a prerequisite for them to be challengeable in courts: if the individual does not know she is nudged, she cannot challenge the measure.

Finally, many nudges may be enacted administratively (see Alemanno & Spina, 2014). Beyond the difficulty in estimating the existence and effect of a nudge, this is an additional barrier for citizens to hold their government and representatives accountable (Glaeser, 2006). Relatedly, policy designers concerned with effectiveness also need to take a long-term view on the implications of the use of specific tools on the design process itself. Legal uncertainty and potential implications for the trust relationship between citizens and governments may have destabilising effects jeopardising any short-term effects. While Reisch and Sunstein (2016) report strong support for nudges in some European countries, citizens in, for instance, Denmark and Hungary are more critical. Also among Americans, support differs drastically depending on the type of nudge and the dispositions of individuals (Jung & Mellers, 2016). Albeit nudge policies often being designed to avoid attracting attention, such divergences in public opinion raise the question how relevant they may be to the electoral chances of governments using them (see, for example, Schubert, 2017).

#### 5.1.4. *The ever-benevolent planner*

One of the most criticised aspects of libertarian paternalism is the role of the social planner (e.g. Gigerenzer, 2015; Glaeser, 2006; Rebonato, 2012; Sugden, 2008). Thaler and Sunstein (2008, p. 238 f.) acknowledge that choice architects may want to further their own interests. However, they dismiss this problem rather bluntly. Paraphrasing their argument: Just because one has problems with a real architect, one does not stop building more houses but uses incentives and monitoring devices. They focus on the specific case of interest capture by the private sector and possible disclosure requirements for politicians and bureaucrats to circumvent this possibility. They subscribe to the view that electoral control and oversight should ensure policy to be in line with the public's interest. Unfortunately, their argument misses some fundamental problems. Public choice scholars have for half a century highlighted the restrictions and costs of controlling public agents (e.g. Buchanan & Tullock, 1962). Even if public choice had identified a satisfactory monitoring, incentive and sanctioning scheme, it would likely rest on the assumption that

public agents behave rationally (see Schnellenbach & Schubert, 2015, on why voter behaviour does not give much hope for such a scheme). Together with the rise of behavioural insights on the side of citizens, the actual behaviour of governments (Lodge & Wegrich, 2016), and of individual politicians and bureaucrats, has also moved in the focus of researchers. Policy decisions may, for instance, be subject to framing effects (see, for example, Kuehnhanss, Heyndels, & Hilken, 2015; Kuehnhanss, Murdoch, Geys, & Heyndels, 2017; Kuehnhanss & Heyndels, 2018, and references therein), and policy-makers may exhibit the same biases as ordinary people (see e.g. Linde & Vis, 2017). Research areas such as behavioural public choice (Lucas & Tasić, 2015), behavioural political economy (Schnellenbach & Schubert, 2015) and behavioural public administration (Grimmelikhuijsen, Jilke, Olsen, & Tummers, 2017) have recently emerged and specifically address such issues.

When a (libertarian) paternalistic policy is enacted based on behavioural insights, the balance between public and private errors must be considered. Glaeser (2006) argues that private citizens are in a better position to overcome biases through feedback and learning mechanisms than public actors, particularly when the latter are the target for private interest capture.[22] Besides limitations on the decision-making of policy designers, classic concerns such as self-serving motives are not decreased. Why, for instance, should a government refrain from using a new policy tool that can be used rather secretively to solidify its position or promote its preferred policies at lower political costs than traditional deliberation and persuasion (see Schubert, 2017, for a discussion)?

## 6. Capacity building and positive integration

To transpose behavioural insights from the academic study of behaviour into policy-making, the concept and findings need to be taken into the political and administrative process. If they are to be employed in formulating and implementing policy, the responsible organisations and actors require the necessary competences and capabilities, that is, the policy capacity to integrate them in the performance of their functions (see Howlett, 2015; Wu, Ramesh, & Howlett, 2015). While anticipating the success of a behavioural insights informed policy is not necessarily straightforward, the operational and political capacities of administrations have over recent years been expanded across a number of countries.

After the publication of *Nudge*, Richard Thaler and Cass Sunstein started working in or advising the public administrations in the United Kingdom and the United States, respectively, strengthening the organisational capabilities for nudging in those countries. Sunstein became advisor to Barack Obama and in 2009 head of the White House Office of Information and Regulatory Affairs (OIRA), an agency with substantial influence on the drafting and implementation of government policies. In terms of political capacity, clear support came from President Obama (2011) in the form of the Executive Order 13563. In line with the normative suggestions of libertarian paternalism, it urges agencies to

[22]See Klick and Mitchell (2006) for a discussion of the effect of paternalistic policies on, among other factors, the learning capabilities of its targets.

> [...] identify and consider regulatory approaches that reduce burdens and maintain flexibility and freedom of choice for the public. These approaches include warnings, appropriate default rules, and disclosure requirements as well as provision of information to the public in a form that is clear and intelligible.

In Executive Order 13707, Barack Obama (2015) further directs all agencies to apply behavioural insights to the design of their policies. The question of individual and organisational competences is, however, left open. In 2015, the centralised Social and Behavioral Sciences Team (SBST, 2016) was established with support of the Executive Office of the President and has since engaged in the empirical testing of policies. Under the new Trump administration, the services of the SBST have, however, been suspended.

In 2010, taking a slightly different approach, the UK coalition government around David Cameron founded the Behavioural Insights Team (BIT or 'Nudge Unit') based in the Cabinet Office.[23] With Richard Thaler on its advisory board, its central goal is close to the normative approach taken in the United States: to 'find intelligent ways to encourage, support and enable people to make better choices for themselves' (BIT, 2011, p. 4). The BIT has from the start integrated the empirical testing of potential policy interventions as one of its core functions. It relies heavily on running randomised control trials (RCTs) to pretest policy interventions before policy is based on their recommendations. RCTs have become a favourite tool for behavioural insight teams in government to test different policy designs. An RCT requires the random split of a sample population into a treatment and a control group. The treatment group is exposed to a policy intervention (or a nudge), whereas the control is not. For both groups, the behaviour of interest is assessed, and conclusions on the effectiveness of the intervention are drawn based on observed differences between the two groups. Examples of RCTs the BIT lists as successfully carried out include the increasing of on-time tax payments and payments of fines by highlighting social norms (e.g. informing people about the tax compliance rate among their neighbours), testing different information methods to encourage debt repayments and a reduction in the amount of unnecessary prescription of antibiotics by doctors through informing them of their peers' behaviour (BIT, 2016; Haynes, Service, Goldacre, & Torgerson, 2012).[24] When carried out properly, RCTs provide a strong argument for the (non-)effectiveness of the tested interventions. While also a common methodology in, for instance, medicine, and increasingly in the social sciences, RCTs are not without problems. Even if randomisation is done properly, they can face severe selection problems, statistical inferences may not be straightforward and the validity of the findings is easily overestimated (see Deaton, 2010; Deaton & Cartwright, 2018). Gigerenzer (2015) points out that the effectiveness of nudges may also be overstated through a publication bias. Successful RCTs and interventions may gain a disproportionate amount of attention. In its report on behaviour change, the House of Lords' Science and Technology Select Committee

[23]Note that the previous New Labour government had already taken first steps towards using behavioural insights in some of its strategy papers and specific policies concerning the environment, finance and health (Halpern, Bates, Beales, & Heathfield, 2004; Whitehead, Jones, & Pykett, 2011). The Behavioural Insights Team was privatised in the form of a social purpose company in 2014. It is part owned by the Cabinet Office.

[24]Note that information provision is not necessarily a nudge and the mentioned examples have been argued to reflect classic persuasion and signalling of authority (Mols et al., 2015).

(2011) criticised the UK government for overly relying on novel behavioural interventions, even if little tested. The report further argues that the empirical testing of behavioural interventions is usually not representative enough of the entire population and that other traditional policy instruments are neglected in the analysis (see also Bubb & Pildes, 2014).

Following the lead of the United States and the United Kingdom in bringing nudging into the political sphere and improving the corresponding policy capacity, other countries such as Australia, Canada, Denmark, France, Germany, Saudi Arabia and Singapore have started to consider and build similar structures. By now, at least 135 countries have in at least some form used nudge-type policies, 51 of which have 'central state-led policy initiatives that have been influenced by the new behavioural sciences' (Whitehead, Jones, Howell, Lilley, & Pykett, 2014, p. 9). The Organisation for Economic Co-operation and Development (OECD) has also shown interest in behavioural insights (Lunn, 2014).

From a different starting point, the European Commission provides a good example of capacity building for integrating behavioural insights into evidence-based policy-making efforts. It began applying behavioural insights in 2009 with the Consumer Rights Directive, limiting the allowable use of default rules by companies in consumer contracts. The directive was followed by a court case against Microsoft and its default installation of its own proprietary Internet browser, demanding consumers be given a mandated choice during installation. Since then, the Commission has set up a framework contract for conducting behavioural studies, and the Joint Research Centre of the European Commission (JRC) launched its own Behavioural Insights Unit. It has created a useful taxonomy for the application of behavioural insights to policy (Louren¸Co, Ciriolo, Almeida, & Troussard, 2016, p. 6):

> [. . . a] classification of initiatives according to whether they are behaviourally-tested (i.e. initiatives based on an ad-hoc test, or scaled out after an initial experiment), behaviourally-informed (i.e. initiatives designed explicitly on previously existing behavioural evidence), or behaviourally-aligned (initiatives that, at least a posteriori, can be found to be in line with behavioural evidence).

The same report identifies more than 200 policy initiatives (at least partially) based on behavioural insights in 32 countries.[25] Many of these cases do not drastically alter the regulatory approach but include a behavioural component that makes an existing policy tool more powerful, such as adjusting defaults. They do not alter the underlying laws and incentives but may still have positive effects at little cost.

To strengthen individual capabilities, the European Commission has, for instance, offered a summer school in behavioural economics for policy-makers and issued guidelines on applying behavioural insights (Van Bavel, Herrmann, Esposito, & Proestakis, 2013). Similarly, the UK BIT has developed a guide called 'MINDSPACE' to popularise the most robust behavioural insights among policy-makers (BIT, 2010; Dolan et al., 2012). While the above examples highlight a centralised approach to building policy capacity for behavioural insights, the hiring of behavioural economists throughout administrations is becoming more frequent. In some countries, for example, Norway

[25]See http://blogs.ec.europa.eu/eupolicylab/portfolios/biap-country_overviews/ for an actively maintained overview.

and Denmark, non-governmental organisations have also started to actively promote behavioural policy design (Lunn, 2014).

## 7. Conclusion

Given the considerable political support, the use of behavioural insights in the design of policy will likely only increase. However, the term comprises a wide range of different practical approaches and different normative starting points. On one hand, behavioural economics and the use of scientific methods to understand, for instance, consumer behaviour have led to stronger consumer protection in the European Union (EU) and may help overcome some biases in people's decision-making. On the other hand, the use of, in particular, nudges by governments to steer citizens' behaviour towards social goals has met criticisms ranging from the very derivation of its normative justification (Whitman & Rizzo, 2015) to concerns over legal implications and privacy (see Kapsner & Sandfuchs, 2015, for the latter point of critique).

In this context, it is a particular challenge to anticipate the success of policies down the line and respond with design effectiveness. With decision-makers themselves subject to the use of heuristics and related biases, and with the success of nudges being an empirical question both before and after implementation, pre-empting potential problems and different contexts in policy design is difficult. Nevertheless, designs allowing systematic empirical testing may be effective if sufficient safeguards for issues such as legitimacy and accountability are implemented.

Current developments, however, do not seem to fulfil the latter criterion. While operational capacity building is ongoing and more policy interventions are based on behavioural science results, the options for citizens to know about, engage in and contest behavioural interventions do not seem to follow suit. Stated as one of the main functions for citizens to be sure that libertarian paternalism does not turn into straight paternalism (Thaler & Sunstein, 2008), public oversight has so far been limited to the issuance of often not very detailed reports about RCTs after completion. To the best of my knowledge, nudges and their normative justifications have yet to come under legal scrutiny.

This review has focused on the origins of behavioural insights and their use in policy to steer the behaviour of individual citizens. An entirely different aspect will be their role in regulating firm behaviour and in shaping the environment in which firms operate (see, for example, Barr, Mullainathan, & Shafir, 2012; Frey & Eichenberger, 1994).

As far as feasible, separating the normative from the positive aspects of behaviourally informed policy design may substantially facilitate the discussion of its further development. At least it would allow more clarity about the assumptions and understanding of individual autonomy and governments' role in shaping people's lives. It may thus again give weight to the fundamental discussion (see, for example, Saint-Paul, 2011), of what notion of individual freedoms and preferences policy ought to be built on in the first place.

## Acknowledgments

I wish to thank two anonymous reviewers, Luigi Bobbio, Gilberto Capano, Bruno Heyndels, Michael Howlett, Benny Geys, Jona Linde, M Ramesh, Rasmus Wiese and the participants of the

following meetings for their comments and suggestions on earlier drafts: the GDN II Workshop on Effective Policy Design (Florence, November 2016), the Annual Meeting of the European Public Choice Society (Budapest, April 2017) and the 3rd International Conference on Public Policy (Singapore, June 2017).

## Disclosure statement

No potential conflict of interest was reported by the author.

## Funding

This work was financially supported by the Research Foundation Flanders (FWO) in the form of a PhD fellowship [Grant No. 11V2117N].

## Notes on contributor

*Colin R. Kuehnhanss* is a postdoctoral researcher at the Department of Applied Economics of the Vrije Universiteit Brussel, Belgium. His research interests include political economy, behavioural economics, and political decision-making. Previous work was published in the European Journal of Political Economy, JCMS: Journal of Common Market Studies, Journal of Economic Psychology, and Public Administration.

## ORCID

Colin R. Kuehnhanss http://orcid.org/0000-0002-3840-5674

## References

Aaken, A. V. (2015). Judge the nudge: In search of the legal limits of paternalistic nudging in the EU. In A. Alemanno & A. L. Sibony (Eds.), *Nudge and the Law: A European Perspective*, Chapter 4, 83–112. Oxford: Hart Publishing.

Alemanno, A., & Spina, A. (2014). Nudging legally: On the checks and balances of behavioral regulation. *International Journal of Constitutional Law*, *12*(2), 429–456.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9–10), 1082–1095.

Amadae, S. M., & Bueno de Mesquita, B. (1999). The Rochester school: The origins of positive political theory. *Annual Review of Political Science*, *2*(1), 269–295.

Angner, E. (2015). To navigate safely in the vast sea of empirical facts. *Synthese*, *192*(11), 3557–3575.

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *Quarterly Journal of Economics*, *118*(1), 73–106.

Baldwin, R. (2014). From regulation to behaviour change: Giving nudge the third degree. *The Modern Law Review*, *77*(6), 831–857.

Bali, A., Capano, G., & Ramesh, M. Forthcoming. Anticipating policy success in policy design. *Policy and Society*.

Barr, M. S., Mullainathan, S., & Shafir, E. (2012). Behaviourally informed regulation. In E. Shafir (Ed.), *The Behavioral Foundations of Public Policy*, Chapter 26, 440–461. Princeton, New Jersey: Princeton University Press.

Behavioural Insights Team. (2010). *MINDSPACE: Influencing behaviour through public policy* (Technical Report). London: Cabinet Office. Retrieved from http://www.behaviouralinsights.co.uk/publications/mindspace/

Behavioural Insights Team. (2011). *Annual update 2010-11* (Technical Report). London: Cabinet Office. Retrieved from https://www.gov.uk/government/publications/behavioural-insights-team-annual-update

Behavioural Insights Team. (2016). *Update report 2015-16* (Technical Report). London: Behavioural Insights Ltd. Retrieved from http://www.behaviouralinsights.co.uk/publications/the-behavioural-insights-teams-update-report-2015-16/

Bemelmans-Videc, M. L., Rist, R. C., & Vedung, E. O. (1998). *Carrots, Sticks, and Sermons: Policy Instruments and their Evaluation* (5th ed.). New Brunswick: Transaction Publishers.

Benartzi, S., & Thaler, R. H. (2007). Heuristics and biases in retirement savings behavior. *Journal of Economic Perspectives*, *21*(3), 81–104.

Bernheim, B. D., & Rangel, A. (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics*, *124*(1), 51–104.

Bernoulli, D. (1738/1954). Exposition of a new theory on the measurement of risk. *The Econometric Society*. https://www.jstor.org/stable/1909829

Binder, M. (2014). Should evolutionary economists embrace libertarian paternalism? *Journal of Evolutionary Economics*, *24*(3), 515–539.

Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes*, *63*(2), 131–144.

Bobrow, D. B., & Dryzek, J. S. (1987). *Policy Analysis by Design*. Pittsburgh, PA: University of Pittsburgh Press.

Bovens, L. (2009). The ethics of nudge. In T. Grüne-Yanoff & S. O. Hansson (Eds.), *Preference Change*, Chapter 10, 207–219. Netherlands, Dordrecht: Springer.

Bubb, R., & Pildes, R. H. (2014). How behavioral economics trims its sails and why. *Harvard Law Review*, *127*(6), 1593–1678.

Buchanan, J. M., & Tullock, G. (1962). *The Calculus of Consent: Logical Foundation of Constitutional Democracy*. Indianapolis: Liberty Fund, Inc.

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

Camerer, C., Loewenstein, G., & Prelec, D. (2005). Neuroeconomics: How neuroscience can inform economics. *Journal of Economic Literature*, *43*(1), 9–64.

Camerer, C. F., Loewenstein, G., & Rabin, M. (Eds.). (2004). *Advances in Behavioral Economics*. Princeton, New Jersey: Princeton University Press.

Capano, G., & Howlett, M. (2015). *Policy design and non-design in policy-making: Policy formulation and the changing dynamics of public policy*. Warsaw: ECPR Joint Sessions.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, *117*(3), 817–869.

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, *105*(5), 1–33.

Costa, D. L., & Kahn, M. E. (2013). Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, *11*(3), 680–702.

Crawshaw, P. (2013). Public health policy and the behavioural turn: The case of social marketing. *Critical Social Policy*, *33*(4), 616–637.

De Haan, T., & Linde, J. (2017). 'Good Nudge Lullaby': Choice architecture and default bias reinforcement. *Economic Journal*, *128*(610), 1180–1206.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science*, *313*(5787), 684–687.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, *48*(2), 424–455.

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2–21.
DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, *47*(2), 315–372.
Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, *33*(1), 264–277.
Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, *47*(2), 268–298.
Dworkin, G., (2017). Paternalism. Stanford encyclopedia of philosophy. Retrieved February 25, 2018, from https://plato.stanford.edu/entries/paternalism/
Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *The American Psychologist*, *49*(8), 709–724.
Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.
Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and Reasoning*. Hove: Psychology Press.
Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, *14*(3), 159–182.
Fehr, E., Kirchsteiger, G., & Riedl, A. (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, *42*(1), 1–34.
Felsen, G., & Reiner, P. B. (2015). What can neuroscience contribute to the debate over nudging? *Review of Philosophy and Psychology*, *6*(3), 469–479.
Fischhoff, B., & Manski, C. F. (Eds.). (2000). *Elicitation of Preferences*. Dordrecht: Springer Netherlands/Kluwer Academic Publishing.
Fonseca, M. A., & Grimshaw, S. B. (2017). Do behavioral nudges in prepopulated tax forms affect compliance? Experimental evidence with real taxpayers. *Journal of Public Policy & Marketing*, *36*(2), 213–226.
Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*(2), 351–401.
Frey, B. S., & Eichenberger, R. (1994). Economic incentives transform psychological anomalies. *Journal of Economic Behavior & Organization*, *23*(2), 215–234.
Gigerenzer, G. (2015). On the supposed evidence for libertarian paternalism. *Review of Philosophy and Psychology*, *6*(3), 361–383.
Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*(1), 107–143.
Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482.
Gilovich, T., Griffin, D. W., & Kahneman, D. (Eds.). (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge, MA: Cambridge University Press.
Glaeser, E. L. (2006). Paternalism and psychology. *The University of Chicago Law Review*, *73*(1), 133–156.
Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75.
Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral public administration: Combining insights from public administration and psychology. *Public Administration Review*, *77*(1), 45–56.
Grüne-Yanoff, T., 2009. Welfare notions for soft paternalism. Papers on Economics and Evolution (0917). . Evolutionary Economics Group, MPI Jena. https://www.econstor.eu/handle/10419/32665
Grüne-Yanoff, T., & Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory? *Minds and Machines*, *26*(1–2), 149–183.
Halpern, D., Bates, C., Beales, G., & Heathfield, A. (2004). *Personal responsibility and changing behaviour: The state of knowledge and its implications for public policy*. London: Cabinet Office.

Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation*, *4*(1), 3–28.
Hausman, D. M., & Welch, B. (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy*, *18*(1), 123–136.
Haynes, L., Service, O., Goldacre, B., & Torgerson, D., (2012). *Test, learn, adapt: Developing public policy with randomised controlled trials* (Technical Report). London: Behavioural Insights Team. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/62529/TLA-1906126.pdf
Howlett, M. (1991). Policy instruments, policy styles, and policy implementation: National approaches to theories of instrument choice. *Policy Studies Journal*, *19*(2), 1–21.
Howlett, M. (2000). Managing the "hollow state": Procedural policy instruments and modern governance. *Canadian Public Administration/Administration Publique Du Canada*, *43*(4), 412–431.
Howlett, M. (2005). What is a policy instrument? Tools, mixes, and implementation styles. In P. Eliadis, M. M. Hill, & M. Howlett (Eds.), *Designing Government: From Instruments to Governance*, Chapter 2, 31-50. Montreal: McGill-Queen's University Press.
Howlett, M. (2014). From the 'old' to the 'new' policy design: Design thinking beyond markets and collaborative governance. *Policy Sciences*, *47*(3), 187–207.
Howlett, M. (2015). Policy analytical capacity: The supply and demand for policy analysis in government. *Policy and Society*, *34*(3–4), 173–182.
Infante, G., Lecouteux, G., & Sugden, R. (2016). Preference purification and the inner rational agent: A critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology*, *23*(1), 1–25.
Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science*, *302*(5649), 1338–1339.
Jolls, C., Sunstein, C. R., & Thaler, R. H. (1998). A behavioral approach to law and economics. *Stanford Law Review*, *50*(5), 1471–1550.
Jones, B. D. (1999). Bounded rationality. *Annual Review of Political Science*, *2*(1), 297–321.
Jones, B. D. (2001). *Politics and the architecture of choice: Bounded rationality and governance*. Chicago: The University of Chicago Press.
Jung, J. Y., & Mellers, B. A. (2016). American attitudes toward nudges. *Judgment and Decision Making*, *11*(1), 62–74.
Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. *Organizational Behavior and Human Decision Processes*, *51*(2), 296–312.
Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, *93*(5), 1449–1475.
Kahneman, D. (2013). Foreword. In E. Shafir (Ed.), *The Behavioral Foundations of Public Policy*, vii–ix. Princeton, New Jersey: Princeton University Press.
Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, *5*(1), 193–206.
Kahneman, D., & Tversky, A. (1979). Prospect Theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292.
Kahneman, D., & Tversky, A. (2000). *Choices, Values, and Frames*. New York, NY: Cambridge University Press.
Kapsner, A., & Sandfuchs, B. (2015). Nudging as a threat to privacy. *Review of Philosophy and Psychology*, *6*(3), 455–468.
Klein, R., & Marmor, T. (2008). *Reflections on Policy Analysis*. Oxford: Oxford University Press.
Klick, J., & Mitchell, G. (2006). Government regulation of irrationality: Moral and cognitive hazards. *Minnesota Law Review*, *90*(6), 1620–1663. .
Korobkin, R. B., & Ulen, T. S. (2000). Law and behavioral science: Removing the rationality assumption from law and economics. *California Law Review*, *88*(4), 1053–1144.
Kőszegi, B., & Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, *121*(4), 1133–1165.

Kőszegi, B., & Rabin, M. (2007). Reference-dependent risk attitudes. *American Economic Review*, *97*(4), 1047–1073.
Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review*, *118*(1), 97.
Kuehnhanss, C. R. (2018). Nudges and nodality tools: New developments in old instruments. In M. Howlett & I. Mukherjee (Eds.), *Routledge Handbook of Policy Design,* Chapter 15, 227-242. Abingdon: Routlegde.
Kuehnhanss, C. R., & Heyndels, B. (2018). All's fair in taxation: A framing experiment with local politicians. *Journal of Economic Psychology*, *65*, 26–40.
Kuehnhanss, C. R., Heyndels, B., & Hilken, K. (2015). Choice in politics: Equivalency framing in economic policy decisions and the influence of expertise. *European Journal of Political Economy*, *40*(pt.B), 360–374.
Kuehnhanss, C. R., Murdoch, Z., Geys, B., & Heyndels, B. (2017). Identity, threat aversion, and civil servants' policy preferences: Evidence from the European Parliament. *Public Administration*, *95*(4), 1009–1025.
Kurzban, R. (2010). *Why Everyone (Else) is a Hypocrtite: Evolution and the Modular Mind*. Princeton, New Jersey: Princeton University Press.
Lepenies, R., & Małecka, M. (2015). The institutional consequences of nudging – Nudges, politics, and the law. *Review of Philosophy and Psychology*, *6*(3), 427–437.
Linde, J., & Vis, B. (2017). Do politicians take risks like the rest of us? an experimental test of prospect theory under MPs. *Political Psychology*, *38*(1), 101–117.
Lodge, M., & Wegrich, K. (2016). The rationality paradox of nudge: Rational tools of government in a world of bounded rationality. *Law & Policy*, *38*(3), 250–267.
Loewenstein, G., Read, D., & Baumeister, R. (Eds.). (2003). *Time and Decision: Economic and Psychological Perspectives of Intertemporal Choice*. New York, NY: Russel Sage Foundation.
Loewenstein, G. F., & O'Donoghue, T., 2004. Animal spirits: Affective and deliberative processes in economic behavior, SSRN Working Paper. https://ssrn.com/abstract=539843.
Lourenço, J. S., Ciriolo, E., Almeida, S. R., & Troussard, X., (2016). *Behavioural insights applied to policy* (European report 2016). European Commission Joint Research Centre. Retrieved from http://publications.jrc.ec.europa.eu/repository/bitstream/JRC100146/kjna27726enn_new.pdf
Lucas, G. M., Jr. (2013). Paternalism and psychic taxes: The government's use of negative emotions to save us from ourselves. *Southern California Interdisciplinary Law Journal*, *22*(2), 227–302.
Lucas, G. M., Jr, & Tasić, S. (2015). Behavioral public choice and the law. *West Virginia Law Review*, *118*, 199–266.
Lunn, P. (2014). *Regulatory policy and behavioural economics*. Paris: OECD Publishing. doi:10.1787/9789264207851-en
Madrian, B. C. (2014). Applying insights from behavioral economics to policy design. *Annual Review of Economics*, *6*, 663–688.
Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics*, *116*(4), 1149–1187.
May, P. J. (2012). Policy design and implementation. In B. G. Peters & J. Pierre (Eds.), *The SAGE Handbook of Public Administration*, 2nd ed, Chapter 17, 279–291. London: SAGE Publications.
Mitchell, G. (2005). Libertarian paternalism is an oxymoron. *Northwestern University Law Review*, *99*(3), 1245–1277.
Mols, F., Haslam, S. A., Jetten, J., & Steffens, N. K. (2015). Why a nudge is not enough: A social identity critique of governance by stealth. *European Journal of Political Research*, *54*(1), 81–98.
Moxey, A., O'Connell, D., McGettigan, P., & Henry, D. (2003). Describing treatment effects to patients. *Journal of General Internal Medicine*, *18*(11), 948–959.
Mullane, M., & Sheffrin, S. (2012). *White Paper: Regulatory nudges in practice*. Tullane University. Retrieved from http://murphy.tulane.edu/files/events/Regulatory_Nudges_feb_24.pdf

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes*, *39*(1), 84–97.
Öberg, P., Lundin, M., & Thelander, J. (2015). Political power and policy design: Why are policy alternatives constrained? *Policy Studies Journal*, *43*(1), 93–114.
O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American Economic Review*, *89*(1), 103–124.
O'Donoghue, T., & Rabin, M. (2000). The economics of immediate gratification. *Journal of Behavioral Decision Making*, *13*(2), 233–250.
Obama, B. (2011). Executive order 13563: Improving regulation and regulatory review. *Federal Register*, *76*(14), 3821–3823. Retrieved from https://www.whitehouse.gov/sites/default/files/omb/inforeg/eo12866/eo13563_01182011.pdf
Obama, B. (2015). Executive order 13707: Using behavioral science insights to better serve the American people. *Federal Register*, *80*(181), 56365–56367. Retrieved from https://www.gpo.gov/fdsys/pkg/FR-2015-09-18/pdf/2015-23630.pdf
Page, E. C. (2006). The origins of policy. In M. Moran, M. Rein, & R. E. Goodin (Eds.), *The Oxford Handbook of Public Policy,* Chapter 10, 207-227. Oxford: Oxford University Press.
Pareto, V. (1906/2014). *Manual of Political Economy*. Oxford: Oxford University Press.
Posner, R. A. (1998). Rational choice, behavioral economics, and the law. *Stanford Law Review*, *50*(5), 1551.
Qizilbash, M. (2012). Informed desire and the ambitions of libertarian paternalism. *Social Choice and Welfare*, *38*(4), 647–658.
Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, *83*(5), 1281–1302.
Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, *36*(1), 11–46.
Rabin, M. (2013). Incorporating limited rationality into economics. *Journal of Economic Literature*, *51*(2), 528–543.
Rachlinski, J. J. (2011). The psychological foundations of behavioral law and economics. *University of Illinois Law Review*, *2011*(5), 1675–1696.
Rebonato, R. (2012). *Taking Liberties: A Critical Examination of Libertarian Paternalism*. Houndmills: Palgrave Macmillan.
Reisch, L. A., & Sunstein, C. R. (2016). Do Europeans like nudges? *Judgement and Decision Making*, *11*(4), 310–325.
Saint-Paul, G. (2011). *The Tyranny of Utility: Behavioral Social Science and the Rise of Paternalism*. Princeton, New Jersey: Princeton University Press.
Salamon, L. M. (2002). The new governance and the tools of public action. In L. M. Salamon & O. V. Elliott (Eds.), *The Tools of Government a Guide to the New Governance,* Chapter 1, 1–47. New York, NY: Oxford University Press.
Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59.
Sanfey, A. G., & Chang, L. J. (2008). Multiple systems in decision making. *Annals of the New York Academy of Sciences*, *1128*(1), 53–62.
Sanfey, A. G., Loewenstein, G., McClure, S. M., & Cohen, J. D. (2006). Neuroeconomics: Crosscurrents in research on decision-making. *Trends in Cognitive Sciences*, *10*(3), 108–116.
Schneider, A., & Ingram, H. (1990). Behavioral assumptions of policy tools. *The Journal of Politics*, *52*(2), 510–529.
Schnellenbach, J. (2012). Nudges and norms: On the political economy of soft paternalism. *European Journal of Political Economy*, *28*(2), 266–277.
Schnellenbach, J., & Schubert, C. (2015). Behavioral political economy: A survey. *European Journal of Political Economy*, *40*(pt.B), 395–417.
Schubert, C. (2017). Exploring the (behavioural) political economy of nudging. *Journal of Institutional Economics*, *13*(3), 499–522.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, *18* (5), 429–434.
Science and Technology Select Committee. (2011). *Behaviour change*. London: House of Lords. Retrieved from https://www.publications.parliament.uk/pa/ld201012/ldselect/ldsctech/179/179.pdf
Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69* (1), 99–118.
Simon, H. A. (1957). *Models of Man: Social and Rational*. New York, NY: John Wiley & Sons Ltd.
Simon, H. A. (1986). Rationality in psychology and economics. *The Journal of Business*, *59*(4, pt.2), S209–S224.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119* (1), 3–22.
Smith, A. (1776/2010). *An Inquiry into the Nature and Causes of the Wealth of Nations* (3rd ed.). Petersfield: Harriman House Ltd.
Smith, A. (1790/2005). *The Theory of Moral Sentiment*. New York, NY: Prometheus/MetaLibri. Retrieved from https://www.ibiblio.org/ml/libri/s/SmithA_MoralSentiments_p.pdf
Sobel, J. (2005). Interdependent preferences and reciprocity. *Journal of Economic Literature*, *43* (2), 392–436.
Social and Behavioral Sciences Team. (2016). *Annual report 2016*. Executive Office of the President. Retrieved from https://sbst.gov/assets/files/2016SBSTAnnualReport.pdf
Stanovich, K. E. (1999). *Who is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), 645–665.
Sugden, R. (2008). Why incoherent preferences do not justify paternalism. *Constitutional Political Economy*, *19*(3), 226–248.
Sugden, R. (2009). On nudging: A review of nudge: Improving decisions about health, wealth and happiness by Richard H. Thaler and Cass R. Sunstein. *International Journal of the Economics of Business*, *16*(3), 365–373.
Sugden, R. (2013). The behavioural economist and the social planner: To whom should behavioural welfare economics be addressed? *Inquiry*, *56*(5), 519–538.
Sugden, R. (2016). Do people really want to be nudged towards healthy lifestyles? *International Review of Economics*, *64*(2), 113–123.
Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, *70*(4), 1159–1202.
Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy*, *112*(S1), S164–S187.
Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *American Economic Review*, *93* (2), 175–179.
Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness*. New Haven: Yale University Press.
Thorsteinson, T. J. (2011). Initiating salary discussions with an extreme request: Anchoring effects on initial salary offers. *Journal of Applied Social Psychology*, *41*(7), 1774–1792.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.
Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, *59*(S4), S251–S278.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.
Tversky, A., & Thaler, R. H. (1990). Anomalies: Preference reversals. *Journal of Economic Perspectives*, *4*(2), 201–211.

Van Bavel, R., Herrmann, B., Esposito, G., & Proestakis, A. (2013). *Applying behavioural sciences to EU policy-making*. Seville: European Commission Joint Research Centre. doi:10.2788/4659.

Von Neumann, J., & Morgenstern, O. (1947). *The Theory of Games and Economic Behavior*. Princeton, New Jersey: Princeton University Press.

Whitehead, M., Jones, R., Howell, R., Lilley, R., & Pykett, J., (2014). *Nudging all over the world: Assessing the global impact of the behavioural sciences on public policy* (Technical Report). Economic & Social Research Council Report. Retrieved from https://changingbehaviours.files.wordpress.com/2014/09/nudgedesignfinal.pdf

Whitehead, M., Jones, R., & Pykett, J. (2011). Governing irrationality, or a more than rational government? Reflections on the rescientisation of decision making in British public policy. *Environment and Planning A*, *43*(12), 2819–2837.

Whitman, D. G., & Rizzo, M. J. (2015). The problematic welfare standards of behavioral paternalism. *Review of Philosophy and Psychology*, *6*(3), 409–425.

Wu, X., Ramesh, M., & Howlett, M. (2015). Policy capacity: A conceptual framework for understanding policy competences and capabilities. *Policy and Society*, *34*(3–4), 165–171.