Introducing Behavioural Insights

During the last 10 years, the concept of *Behavioural Insights* (BI) has become an increasingly established one in the public policy vocabulary as well as in society at large. It was originally coined by the UK Behavioural Insights Team (UKBIT) to refer to an evidence-based approach to integrating insights and methodologies from the behavioural sciences in public policy in order to provide better and more effective regulation (Halpern, 2015). As this approach has spread wider into public policy circles, the resulting initiatives and outcomes are increasingly referred to as 'behaviourally informed public policy', or just 'behavioural public policy'. Only 10 years in, this new approach to the development, implementation and evaluation of public policy has been granted its own academic journals, associations, cross-institutional networks, and an ever-increasing number of institutions and teams coordinating and/or integrating BI into public policy around the world (for resources, see Appendix 1).

The core tenet of BI is the application of *behavioural insights* and methodologies from the behavioural sciences in the formulation, development, design and delivery of public policies. To be more precise, these insights are mainly taken from behavioural economics, cognitive- and social psychology, the study of judgement and decision-making, and similar disciplines sharing not only the inductive, but also the causal explanatory and experimental approach to the subject matter of human behaviour as well as their theoretical underpinnings adopted from dual process theories of human cognition.[1] The aspiration is to better understand why people act as they do in order to create less invasive and more effective public policies by applying more realistic models of human behaviour; in particular, models taking into account how our cognitive limitations and biases, as uncovered by these sciences, influence our behaviour.

Thus, Behavioural Insights stands in contrast to more traditional approaches to public policy making, which have tended to rely on more abstract models and ideal assumptions about human behaviour; models that do not factor in such limitation and biases. Instead traditional approaches have usually assumed that people's behaviour could be understood as if resulting from fully rational and deliberative thinking provided full information and the absence of constraints on time and attention. Consequently, at least according to proponents of BI, traditional policies easily end up being naïve, overly invasive and ineffective as they reflect assumed rather than actual behaviours. BI in contrast claims to provide more realistic models and assumptions about the psychological factors that shape human behaviour, tools for how to influence such behaviour and methods for how to investigate and measure actual behaviour and behaviour change.

Behavioural Insights, or just BI, is the evidence-based approach to integrating insights and methodologies from the behavioural sciences in public policy and beyond in order provide better and more effective regulation and initiatives aimed at causing behavioural change.

'Behavioural insights' (written in full), on the other hand, refers to the specific insights and methodologies from the behavioural sciences sought integrated. In particular, this latter concept of behavioural insights refers to a series of theories and empirical findings originating in the behavioural sciences about what shape real world human behaviour in predictable ways, including the methodologies for how to approach this subject matter.

---

[1] What exactly constitutes 'the behavioural sciences' is up for debate. Some prefer to define the behavioural sciences very broadly so as to accommodate almost any approach that relate to human behaviour, others prefer to define the term more narrowly so as to ensure at least some level of theoretical and methodological consistency. This book presents an approach that falls in the latter category. Either way, it is important to emphasise that the behavioural sciences do not by themselves constitute a unified field, but rather feature a plurality of sciences that do not readily lend itself for policy makers and practitioners to tap into. Still, BI tend to draw on a particular branch of psychological theories viz. those compatible with experimental methodologies (Lepenies & Malecka, 2018).

Saving for retirement

To illustrate what BI brings to the table of policy making, we may look at work carried out relative to people saving for retirement. In most countries saving for retirement is one of the most sensible things you can do if you expect to live beyond the age of retirement and do not want to live off your children's wages. Even so, currently roughly one third of UK adults are not saving at all for retirement and of those that are saving many are not saving enough to give them the standard of living they hope for when they retire. This is not only a problem in the UK. A study in the US found that only 32% of respondents believed their pension savings-rate was "about right", while 68% of respondents believed their savings-rate were "too low" (and 1% believed it to be "too high"). That means a swooping majority of 69% was not choosing optimally as judged by themselves – and that was amongst a sample of respondents that actually did save for retirement, which, in turn, is only true for about 55% of all non-retired Americans.

This type of behaviour is not only a problem for all those employees that end up without retirement-savings. It is also a societal problem. Pension contributions are usually directed into long-term investments that are good for societal development; proper retirement savings results in elderly people leading better, more active and healthier lives, reducing their negative impact on public health spending and positively impacting the social capital of their communities; and people who save sufficiently for their retirement enjoy a higher quality of life in general, the guarantee of which is one of the main purposes for public policies to pursue. Thus, if free markets do not seem capable at readily achieving societal and individual interests by their own devices, the problem that people do not save enough for retirement becomes a regulatory problem.

However, looking closer at a problem like the failure for employees to contribute sufficiently, or just contribute at all, to their retirement savings, traditional policy efforts such as creating awareness, providing information and campaigns and aligning incentives seem to be quite ineffective. For instance, a US study found that while 100% of employees participating in a seminar about pensions declared their interest in saving more for retirement, only 14% ultimately signed up for a plan increasing their savings-rate. Although this was much better than the mere 7% doing the same in a comparison group, it represents the more general finding in pension research, that lack of awareness, information or proper attitudes are not the main causes why employees throughout the world are failing to save enough for their retirement. Likewise, the problem does not seem to be caused by a lack of positive incentives for saving for retirement – or the implied negative consequences for failing to do so. As mentioned above, saving for your retirement is one of the most sensible things you can do, as employers in most developed countries provides quite generous matching contributions – e.g. 50% of the employee's monthly contribution up to 6% of the salary – on top of which is provided generous tax reliefs.

As such the widespread insufficient tendency of saving too little for one's retirement, does not appear to be a problem of lack of information, proper attitudes, sufficient incentives or regulation. Rather it appears to be an instance of what may be referred to as a *behavioural problem*.

> **A behavioural problem** is a pattern in behaviour, whether regarded in terms of attention, belief formation, choice, or determination that occur despite that people have good reasons to act otherwise; i.e. people know what they ought to attend to, hold proper beliefs, and face incentives and regulation, that in sum provide them with the necessary and sufficient reasons for acting differently.
>
> A subjective indication of irrational behaviour is that people enacting such behaviour use the word 'ought' when recognising that they should have done otherwise: e.g. "I ought to have seen the sign", "I ought to have believed what my parents said", "I ought to have chosen a healthier meal", and "I ought to have gone to the gym".
>
> In the psychological literature there has been heated discussion about whether to use the predicate 'irrational' in relation to such behaviours. One school of thought holds it to be a derogatory term. However, as will be argued later in this book, there is no such meaning to the term 'irrational', why it will be used interchangeably with 'behavioural problem' in this book. For more on this discussion, see Chapter 2.

When it comes to behavioural problems or 'irrational behaviours', such as not saving for retirement, the suggestion made by the behavioural sciences is that, such behaviours are often the result of people's judgement and decision-making being subject to the influence of *cognitive biases and heuristics* that may be activated by subtle contextual factors. One such well-known factor, is that of *default settings*. At its most simple a default setting is an option from a choice set that enters into force by default unless the person choosing, actively chooses an alternative option. What is interesting from a policy perspective is that the default setting *ought* not to influence people's behaviour; i.e., if people are purely rational creatures as usually assumed in traditional public policy the default setting should not be expected to have any effect on their behaviour. Instead, a rational person would choose whatever option he or she prefers the most regardless of the default setting. Yet, within the behavioural sciences, a multitude of studies and experiments have revealed that default settings quite significantly affect what people end up choosing (Jachimowicz, Duncan, Weber, & Johnson, 2019).

This has also been found to be the case when it comes to making pensions contributions. In many countries where people are observed not to save sufficiently, or at all, for their retirement the default thus turns out to be that of 'not making contributions to a pension fund unless actively choosing otherwise'. In the behavioural sciences this kind of default setting is also referred to as a so-called 'opt-in system' as it requires people to make an active choice to participate in a particular system; a system famous for getting people to stick to the status quo.

To study how default settings affect retirement savings, behavioural economists Madrian and Shea studied how changing an opt-in default to an opt-out default impacted the participation and savings rates in a 401(K) defined-contribution pension plan for employees at a large US cooperation (Madrian & Shea, 2001). At the company studied by Madrian and Shea, employees could participate in 401(k) when they had worked one or more years at the firm. Employees eligible for participation had the option of contributing up to 15% of their salary, with the first 6% receiving a 50% match by the employer. Besides (1) choosing whether to participate in 401(k) or not, and (2) selecting the contribution rate wanted, individuals also needed to (3) choose the investment allocation of the combined employee and employer contribution amongst nine different investment funds. Under this system 37% of employees had enrolled in 401(k) after one year, a number that slowly increased with tenure reaching 83%, but only after 20 years of employment (see figure 1). Further, the selected contribution rates were widely distributed, though with 30% clustering on the 6% contribution level thus maximising the 50% contribution by the employer (see figure 2). Finally, the investment was allocated amongst the different options with the majority being invested in stocks (75%), followed by bonds (17,8%), and money market (7,3%), (see figure 3).[2]

Cognitive biases and heuristics are psychological mechanisms and mental 'rules of thumb' that do not rely on thinking in the sense where one thought logically leads to another. When applied in judgement and decision-making they may potentially lead people to act irrationally.
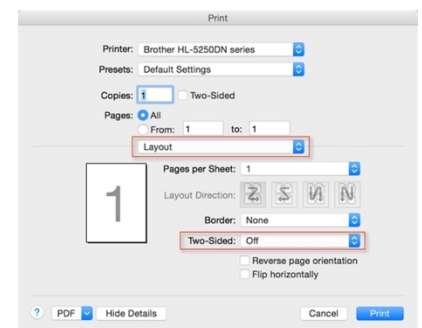
The perhaps best illustration of a basic default setting is known from printers where the default can be set to either one-side (simplex) vs. double sided (duplex) print. While changing the default from one-sided to two-sided print does not provide people with a reason to use less paper, it has been shown to significantly decrease their use of paper, see e.g. (Egebark & Ekström, 2013)

---

[2] These numbers are simplified somewhat from the analysis of Madrian and Shea. The 37% enrolment is taken from their 'WINDOW' cohort – a transition group of employees who was made immediately eligible for participation in 401(k), but was not automatically enrolled – and limited to (3-15 month of tenure). This group is chosen as it is the one used by Madrian and Shea as comparison with the effect of automatic enrolment, see (Madrian & Shea 2001: 1159). The same choice is made with regards to the contribution rates reported here, see (ibid: 1163), while the investment allocation is taken from their 'OLD' cohort, as this represents the allocations made in the original system (ibid: 1170) – albeit the allocations from the OLD and the WINDOW cohort are almost identical.

The reason why Madrian and Shea studied the company, though, was that by April 1, 1998, it changed the default of participation. With the new default the system was switched to an 'opt-out default system', where new employees were automatically enrolled in the 401(k) plan. Further, automatic enrolment also meant that additional defaults for contribution levels and investment allocation were needed. These were chosen as 3% and 'money market', respectively. While this change of defaults from a rational point of view should not make any difference for 401(k) participation, contribution rate and allocation, the impact of the change was huge. With automatic enrolment 86% of new employees was participating in 401(k) after approximately one year – that is, more than double the number for a comparable group under the old system (see figure 1) and more than participated in 401(k) in the old system after 20 years of employment. In addition, the increase equalised participation in 401(k) across various demographic groups such as gender and ethnicity.
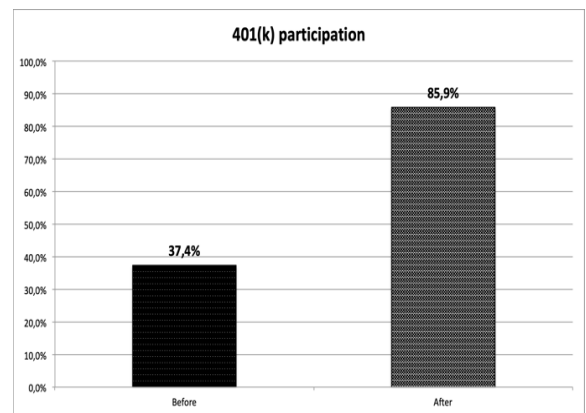

Figure 1 After changing the default from an opt-in to an opt-out system participation in 401(k) retirement savings-plan saw a significant increase. Madrian and Shea attribute this effect to the reversal of the friction involved in the pension program.
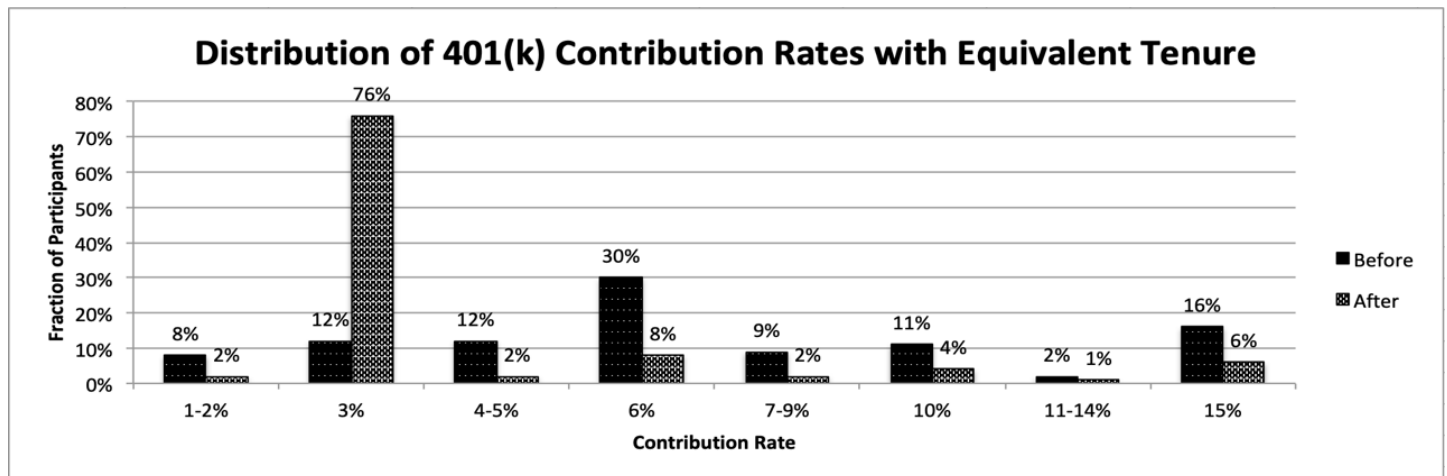

Figure 2 Changing the default to an opt-out system did not only reduce the friction of participating to zero. It also effected the distribution of contribution rates so that participants in general saved less than before the intervention. Madrian and Shea attribute this effect to the fact that the new default also suggests a contribution rate, which in this case was lower than optimal.

However, the change of default also had adverse effects. These included that the modal contribution rate fell from 6% (maximising employer match) to the 3% contribution rate suggested by the new default (see figure 2) as well as a change in the investment allocation such that more than 80% was now allocated to the money market (see figure 3). Thus, the change of default not only worked to get employees to join the 401(k) plan, but also directed their subsequent contribution and investment allocation decisions.

So how does one explain these data? According to Madrian and Shea saving up for pensions requires some determination. One needs to pull together and consider whether to participate, decide what contribution rate to opt for, how to allocate the resulting investment and finally go
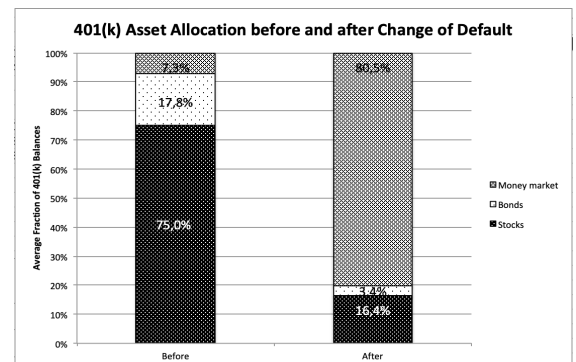

Figure 3 Likewise, the new default also had the adverse effect of suggesting an asset allocation based solely on the Money market.
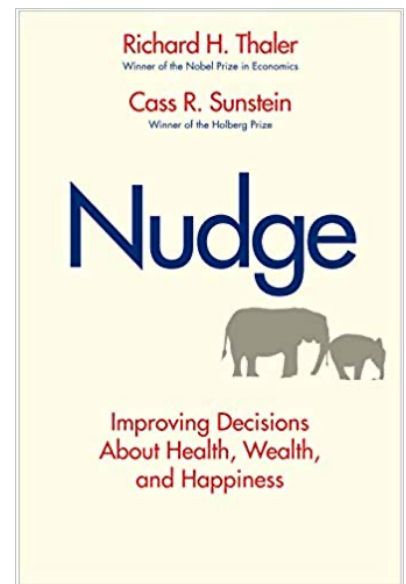
through the process of implementing those decisions. All of these factors may be described as 'friction costs', i.e. the costs associated with the means of implementing a choice, rather than with the choice itself. In the case of pension, changing the default from an opt-in to an opt-out system reverses the friction costs over options so that it now becomes zero for participation with some friction costs imposed instead on non-participation.[3]

But this is just one of multiple factors. In observing that the change of default did not only have an effect on participation in 401(k), but also on subsequent contribution rates and investment allocations, Madrian and Shea abduces that the new default also imposes friction costs on choosing alternative contribution levels and investment allocations than that chosen by default. If that is not all, the new default also seems to perform the further function of "suggesting" to employees that the 3% contribution rate with full investment allocation to money markets is the sensible or even recommended course of action. That is, this additional function of the default results from the complexity of forming beliefs about what the optimal contribution rate and investment allocation might be when joining a 401(k)-pension plan. In conclusion, then, even though defaults ought not to influence people's decision to save for retirement, they clearly do so, resulting in the irrational behaviours observed; but by changing the default one may also change these behaviours, albeit not their irrational nature.

Nudge and Libertarian Paternalism

The case of saving for retirement provides a paradigm example of how behavioural insights may be used to inform public policy with the result of providing more effective and less invasive policies, than those devised by traditional regulation. In their popular book *Nudge – Improving Decisions about Health, Wealth and Happiness* (2008), behavioural economist Richard Thaler and law professor Cass Sunstein referred to such interventions, which achieve their behaviour changing effect without banning options or making use of further incentives, as 'nudges'. Their suggestion was that if a behavioural or decision-making pattern is sensitive to the influence of subtle contextual factors of the *choice architecture*, these contextual factors may be adjusted so as to "nudge" people's choices in ways that promote a more preferred behaviour, rather than obstruct it. In particular, they argued that such nudges, if done by public policies may avoid some of the challenges and potential pitfalls of traditional regulation, such as costly procedures and ineffective campaigning, unintended effects of incentivizing behaviours, and invasive choice regulation, such as bans. The advantage, they claimed, of applying behavioural insights in terms of nudges is that public policy makers might thus supplement – or, perhaps, at times even replace (Thaler & Sunstein, 2008, p. 14) – traditional regulation with nudges so as to influence people's everyday choices and behaviours in cheaper, less invasive, and more effective ways.

The front cover of the first edition of 'Nudge: Improving Decisions About Health, Wealth and Happiness' was published in 2008. An updated and revised edition was published soon thereafter in 2009, which is the edition that one should refer to.

Choice architects and choice architecture. In their book *Nudge* Richard Thaler and Cass Sunstein introduced the notions of 'choice architecture' and 'choice architects'. By choice architecture they meant the totality of factors – contextual, cognitive and rational – that influence judgement and decision-making. By choice architects, they meant the people responsible for or capable at controlling those factors and thereby influencing the judgement and decision-making of other people, see (Thaler, Sunstein, & Balz, 2010).

---

[3] Of course, one could attempt to explain the massive effect of changing the default in terms of a rational response to the reduction of friction costs. Yet, Madrian and Shea holds that the massive effect on participation can hardly be rationalised completely in this way. Research has shown time and again how even just a tiny bit of friction tend to make people procrastinate. For this reason, Madrian and Shea suggest, it is an irrational, rather than a rational effect of such friction costs, that is partly responsible for the status quo behavioural problem of people not saving enough, or at all, for their retirement; and that changing the default system from an opt-in to an opt-out system is all it might take to get people to join.

In this sense, nudges may be described as interventions falling within a regulatory paradigm that Thaler and Sunstein refers to as *Libertarian Paternalism*. The concept of Libertarian Paternalism was originally introduced in a 2003 essay of the same name (Thaler & Sunstein, 2003). Here they defined a policy 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" where they intend by "better off" that this be "measured as objectively as possible" (p. 175). According to Thaler and Sunstein, while many economists believe the term paternalistic to be derogatory because they think paternalism always involves some kind of coercion, this is not necessarily the case. Policies may be selected with the goal of influencing the choices of affected parties in a way that will make those parties better off, but where there is no coercion involved. They refer to this kind of paternalism as libertarian paternalism and ultimately define it as "… an approach that preserves freedom of choice but authorizes both private and public institutions to steer people in directions that will promote their welfare." According to Thaler and Sunstein an approach like that of libertarian paternalism "should be acceptable to even the most ardent libertarian".

In *Nudge* the notion of libertarian paternalism is further refined. It is described as a "movement" or "strategy" recapturing common sense from dogmatists (Thaler & Sunstein, 2008, p. 5). The libertarian aspect of the strategy is said to lie in "the straightforward insistence that, in general, people should be free to do what they like – and to opt out of undesirable arrangements if they want to do so" (ibid). Thaler and Sunstein say that "libertarian paternalists urge that people should be 'free to choose'" and strive to "design policies that maintain or increase freedom of choice" (ibid). In particular, they say that by modifying the term 'paternalism' with 'libertarian' they simply mean liberty preserving, adding that "Libertarian paternalists want to make it easy for people to go their own way; they do not want to burden those who want to exercise their freedom" (ibid). According to Thaler and Sunstein the paternalistic aspect lies in the claim that "it is legitimate for choice architects to try to influence people's behaviour in order to make their lives longer, healthier, and better" (ibid). Hence, they argue for a self-conscious effort by institutions "to steer people's choices in directions that will improve their lives" (ibid). However, they also modify their understanding of paternalism compared with their 2003 paper, now holding that, "a policy is 'paternalistic' if it tries to influence choices in a way that will make choosers better off, as judged by themselves"(ibid) – the so-called *JBT-Criterion* – rather than interpreting "better off" as earlier to be "measured as objectively as possible" (Thaler & Sunstein, 2003, p. 175). According to Thaler and Sunstein "libertarian paternalism is a relatively weak, soft, and nonintrusive type of paternalism because choices are not blocked, fenced off, or significantly burdened" (ibid). Yet, it does count as paternalism because "private and public choice architects are not merely trying to track or to implement people's anticipated choices. Rather, they are self-consciously attempting to move people in directions that will make their lives better" (Thaler & Sunstein, 2008, p. 6). In particular, Thaler and Sunstein say that by doing this "They nudge" (ibid).

The definition of nudge.
Various revisions have been provided in the academic literature in order to clarify conceptually as well as ethically relevant aspects of the definition of a 'nudge', such as whether nudges are intentional interventions and how nudges involve the active use of non-rational psychological mechanisms (Hausman & Welch, 2010), (Hansen, 2016). Whereas the original definition of Thaler and Sunstein (2008, p. 6) says that:

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

a more technical definition suggested by (Hansen 2016) asserts that:

"A *nudge* is any function of an attempt at influencing people's judgment, choice or behaviour in a predictable way (1) that is made possible because of cognitive boundaries, biases, routines and habits in individual and social decision-making posing barriers for people to perform rationally in their own declared self-interests and which (2) works by making use of those boundaries, biases, routines, and habits as integral parts of such attempts."

Which definition one prefers depends upon what level of detail and precision is needed for the purpose at hand.

Highlight: Four small nudges with big impacts

Improving hand hygiene at hospitals by making the decision to sanitise salient. Every year thousands of patients die from hospital acquired infections due to bad hand hygiene amongst staff, visiting family and patients themselves. Many of these infections could easily be avoided if people took care and sanitised their hands using the sanitisers set up at entrances and the various sections of hospitals. However, applied behavioural science studies have found again and again that compliance levels are extremely low amongst all groups involved due to the low level of visible and insistent encouragement to sanitise. People are simply expected to pay attention to, and then do the obviously right thing in the situation. Experiments by iNudgeyou (DK) have shown that by merely placing sanitisers in front of entrances with salient signs significantly nudges visitors compliance; e.g. an Increase from 1% to 15% were measured at a main entrance of a large hospital and an increase from 0-1% to 67% at a hospital section (Aarestrup, Moesgaard, & Schuldt-Jensen, 2016), (Hansen, Schilling, & Malthesen, 2019).

Improving traffic safety using visual illusions. Fast paced traffic was not a factor in most of human evolutionary history. It is not surprisingly then, to find that we have grave difficulties assessing speed, control and other risk factors when driving in our cars. The result is that thousands of people die every year in traffic accidents that could have been avoided. A place where this is the case is the Lake Shore Drive in Chicago, where a tight turn makes it one of the city's most dangerous curves. Trying to limit accidents the city painted a series of white lines perpendicular to traveling cars such that the lines get progressively narrower as drivers approach the sharpest point of the curve. This creates the illusion of speeding up, which – by hypothesis – should make drivers lift the foot from the speeder to compensate for possible illusions of control and overconfidence. The result: there were 36 percent fewer crashes in the six months after the lines were painted compared to the same six-month period the year before. (Thaler & Sunstein, 2008).

Making energy efficient appliances more attractive. Many choices that are attractive in the short term has non-attractive consequences in the long run. However, in making such choices the long-term consequences are often intangible, abstract and forced into the background of decision-making. This is especially true when consumers buy white appliances and other energy consuming devises, where the shelf-price is up front, while consumption costs need to be calculated from technical specifications. In an attempt to change this, GreeNudge (N) implemented a sticker showing the assessed lifetime costs of appliances next to their price tag in a large electronic storehouse. They then calculated the effect of this by comparing sales with 10 storehouses in the same chain in the same period. The effect of adding the sign with lifetime costs up front with the shelf-price led to the average white appliance sold being 4,9% more energy effective than in the other stores. If rolled out across the EU the CO_2 emissions saved from this policy would amount to removing 2 mill. cars of the streets (Kallbekken, Sælen, & Hermansen, 2013).

Helping job-seekers stay focused and motivated in their job-search. Some goals cannot be achieved by making a single choice. They require making the right choice over and over again. One example of this is when searching for a job. While initially fuelled by motivation, this search might quickly go sour and become about documenting minimal compliance at the job centre. To help job-seekers stay focused and motivated in their job search the UKBIT trialled a 'commitment pack' at UK job centres drawing on the research on implementation intentions (see Chapter 6) that shifted the focus of the relationship between job-seekers and coaches from one of documenting compliance to working together to agree on specific goals in the immediate future. First testing this intervention on a small-scale trial in one job centre, and then scaled up the commitment packs across 12 job centres UKBIT found that 3% job seekers in the treatment group across these 12 centres became independent of income support faster than the control group over the 11 months of the trial, (Briscese & Tan, 2018).

Challenging Traditional Public Policy

To understand how profound Thaler and Sunstein's suggestion in *Nudge* was, where it came from and why it was received as it was, we need to travel back a bit in the intertwined history of classical economics and public policy.

The emergence of Homo Economicus

During the last 150 years or so, economic theory has come to define the basic object of public policy *viz.* the portrayal of citizens as '*Homines Oeconomici'*. This all began in the late 19th Century when economists started to build mathematical models on the assumption that decision-makers act like Homo Oeconomicus; models that came to dominate economics as well as public policy as mathematical theories of rational choice developed.

Originally rational choice theory had emerged as the result of mathematical musings over economic decisions, especially gambling problems, beginning in the latter part of the 18th Century. In the early 20th Century it reached its maturity with the work of John von Neumann and Oskar Morgenstern (von Neumann & Morgenstern, 1944). In order to work out how people ought to make decisions posing risks as well as uncertainty, the theory came to examine decision-making using mathematical models build around the assumption of individual decision-makers trying to figure out what choices to make to maximise their expected subjective utility. In doing this, decision-makers would be advised to reason according to some fundamental self-evident axioms defining rationality (see below), plus formal mathematical rules for deriving their implications. This would allow them to figure out what choices would serve as optimal means, or 'instruments', to the end of such utility maximisation. That is, if one wanted to maximise expected utility, one was advised to reason according to standards of *instrumental rationality*; rules which, by the way, says nothing about what ends to pursue, only about how one ought to pursue them (Hume, 1748 (1999)). Thus, rational choice theory emerged as a 'normative theory' of instrumental rational decision-making – i.e. a normative theory about the process proscribed for optimally and reliably arriving at one's goals, whatever they might be.

Of course, a pre-condition for rational choice being a successful approach to decision making is that decision-makers act on some information about the world, including themselves, their options and the quality of goals. Being a normative discipline primarily interested in what it meant to act rationally, little interest, though, was paid in economics to what constituted reliable and rational information acquisition. For purposes of convenience, decision-makers where as a starting point assumed to hold perfect information about their own preferences, their available choices and the quality of the predicted outcomes associated with these. Also, their belief-formation was assumed to adhere to sensible rules of epistemic rationality and Boolean logic. While probabilistic models, such as Bayesian inference, where indeed incorporated along the way to deal with issues of risk and uncertainty, the assumptions of perfect self-knowledge of preferences, information, and epistemic rationality were taken as given. The model of Homo Oeconomicus that resulted from combing these assumptions with those of instrumental rationality, came to define what constitutes 'perfect rationality'.


The term 'Homo Oeconomicus' was first used in the late 19th Century by critics of John Stuart Mill (1806 - 1873) to refer to his proposal of "an arbitrary definition of man, as a being who inevitably does that by which he may obtain the greatest amount of necessaries, conveniences, and luxuries, with the smallest quantity of labour and physical self-denial with which they can be obtained" (Mill, 1836); see (Persky, 1995).


*L'Uomo Vitruviano (Vitruvian Man) by Leonardo da Vinci ca. 1490 is often used to depict the idea of man as Homo Oeconomicus. While there are many theories that pins out rationality in slightly different ways the main characteristics of assuming man to be rational usually apply to four basic domains: **Attention** is a matter of choice and thus people should focus on what is most important to them; **Belief-formation** should be subject to the rules of logic and rationality; **Choices** should be made according to the axioms of rational instrumental rationality and may thus be explained and predicted by subjective expected utility functions; and **Determination**, requires that people should stick to their choices ceteris paribus.*

The axioms of rational choice. A choice is 'instrumental' when it serves as a means to a further end. For instance, both apples and lemons may serve as food (a).

(a)   (b)
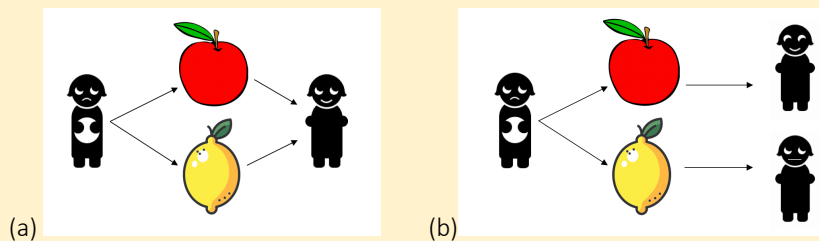
For a choice to further qualify as Instrumentally rational it has to be the optimal given one's preferences combined with one's beliefs about the available choice options, and one's expectations about the outcomes associated with choosing each of these options. For instance, given that one believes that one can choose between apples and lemons, and one's expectations about one's experiences of the taste of apples relative to lemons, one may prefer to eat an apple to a lemon (b).

(c)   (d)
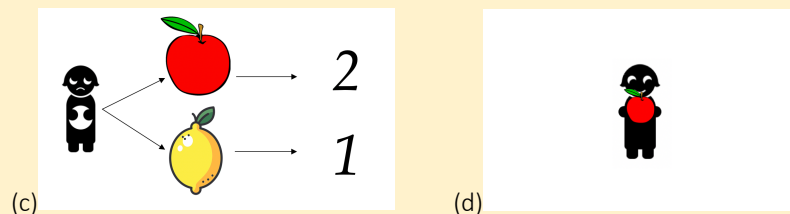
More formally, a strict preference for an option x relative to an alternative y may be expressed by a preference relation >, such that if x is preferred to y, then x > y. If the chooser is indifferent between two options x and why, then x = y. This also allows for a numerical representation, where x is assigned a higher value or 'utility number' than y based on the choosers subjective expectations, e.g. 2 and 1, such that these assignments capture the preference relation, i.e. 2>1 (c). Mathematically preferences may then be described by a utility function, which maps each choice option to a utility number.

In turn, this allows for saying that an 'agent' (this is how one refers to the 'chooser') is choosing instrumentally rational if she maximises her subjective expected utility function in the choice scenario (d). In particular von Neumann and Morgenstern proved that, for an agent to maximise her subjective expected utility function in general, she must be following a series of self-evident axioms of rational choice, as well as if she follows this series of self-evident axioms of rational choice, her choices may be represented by a subjective expected utility function. Assuming that the agent has three choice options – x, y, z – these axioms may be described formally as follows (see side-notes for an illustration):

The preferences of an agent choosing instrumentally rational should be
a. *reflexive,* i.e. x = x;
b. *complete*, i.e. either x > y, y > x or x = y);
c. *transitive*, i.e. if x > y and y > z, this implies x > z;
d. *independent*, i.e. if x > y > z, then x > y, y > z and x > z;
e. and *continuous*, i.e. if x > y > z, there always exists a probability *p* such that $xp + z(1 − p) = y$.

During the mid-20th Century, however, the lines between the normative proscriptions of economics and rational choice theory, on the one hand, and the description of human behaviour, on the other hand, became blurred. Since the axioms defining rational choice were self-evidently reasonable it was only a short step to thinking that no one would intentionally deviate from these standards except by mistake; and, although mistakes might occur, they would surely be random, thereby annulling themselves in the aggregate, as well as become weeded out over time, as learning and experience would take hold – if not by deliberative means, then by the evolution of behavioural rules and thinking strategies progressing through trial-and-error, such that individuals would be making their choices 'as if' being rational. Ultimately then, when individuals repeatedly act and interact, a sensible way to approach the outcome would be by analysing not only individual behaviour as rational, but also social behaviour as the aggregate result of the individual choices of rational decision-makers. Thereby the foundation of classical economics and its portrayal of citizens, consumers and family members as Homines Oeconomici had been cast. The proscriptions of rationality became empirical predictions as well as ex-post tools of analysis, and since then mathematical models of rational decision-making have formed the backbone of the way economists, and in turn most public-policy makers and practitioners, understand as well as predict human behaviour.

The Rational Foundations of the Traditional Public Policy Approach
In morphing from a normative ideal to a descriptive truth, the theory of rational choice and its conceptualisation in terms of Homo Economicus came to provide scientific structure, and precision to what may be referred to as *traditional public policy*, the general regulatory approach of which may be captured in terms of the 'traditional policy escalator' for behavioural regulation, see (figure 4); a generalisation of how public policy efforts tend to proceed in terms of the provision of information, campaigns aimed at rational persuasion, provision of positive incentives and negative sanctions, as well as hard regulatory measures such as bans and prohibitions.

According to the traditional policy escalator, when observing people to act in ways they ought to change, the most charitable thing to assume is that they are indeed rational, but just not aware that they ought to act differently and why; i.e. the most charitable thing to do is to assume the undesired behaviour is a case of an *informational problem*; a type of problem that traditionally may be methodologically documented by measuring the level of information through standard surveys. If the assumption of an informational problem is found to be validated, the first strategic step of the escalator is, then, to get peoples' attention by creating awareness and inform them about how to act and what the reasons are for this. Practically, this strategy of creating general awareness and providing information is, in most instances, also a quite cheap and very little invasive strategy to pursue that few citizens would oppose.

Normative vs. descriptive theory.
A descriptive theory is a theory that attempts to describe and explain the world in a value-free and objective way. A normative theory is a theory that attempts to prescribe how things ought to be, what is right and what is wrong, what is acceptable and what is not.

A normative theory about rational choice, specify how a perfectly rational agent ought to choose and thus establishes a standard for evaluating actual behaviour much like the idea of a perfect circle provides a standard for evaluating an actual circle you might draw on the blackboard.

In rational choice theory the move from a normative to a descriptive theory consists in assuming people to be rational, i.e. to be Homines Oeconomici. Under this assumption, observing a person to pick an apple from a fruit-bowl makes for inferring that this is the fruit that the person preferred the most. This move is also termed *revealed preference theory*.

One problem with this move is that almost any behaviour viewed in isolation may be explained as rational by making fitting assumptions about an agents beliefs, thereby making the theory unfalsifiable; or more precisely, making the explanation pseudoscientific (Popper, 1963).

Revealed preference theory (RPT) *was pioneered by economist Paul Anthony Samuelson (1915 – 2009) as a method of analysing choices made by individuals, and mostly used for comparing the influence of policies on consumer behaviour. Taking economics to be an empirical subject, there is an issue that one cannot observe preferences. In other words, according to advocates of revealed preference theory "It is not what you say, it is what you do that reveals what you want". Revealed preference models assume that the preferences of consumers can be revealed by their purchasing habits. Revealed preference is a way to infer the preferences of individuals given their observed choices. This contrasts with attempts to directly measure preferences or utility, for example through stated preferences.*

Figure 4: The traditional policy escalator: in pursuing policy goals regulatory efforts tend to escalate through a series of steps aimed at achieving rational behaviour change. As efforts escalate, the policy measures tend to become increasingly invasive.
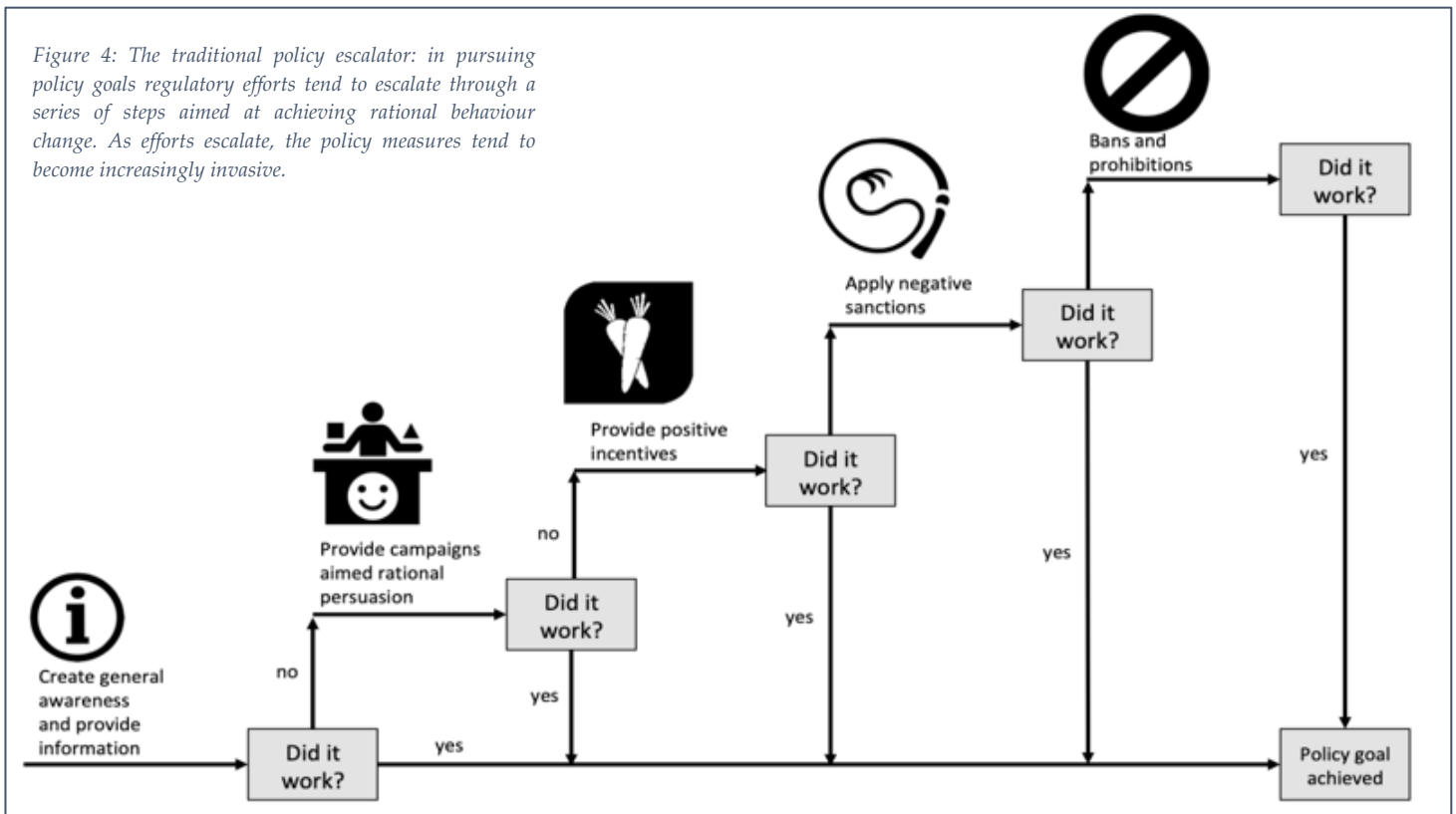
However, should this initial strategy not work, the traditional policy escalator invites for assuming instead that people perhaps did not act on the information provided because they do not hold the proper attitudes; which is just another way of saying that people probably do not agree and are thus not intrinsically motivated to change their behaviour. Methodologically such *attitudinal problems* may, given certain precautions, quite easily be documented using standard survey techniques such as Likert-scales. If the lack of proper attitudes turns out to be the problem, the strategic measure suggested by the escalator is to provide public campaigns that rationally argues the case in order to persuade people to change their behaviour. On a practical level, while more expensive and a bit more invasive than the pure provision of information, rational persuasion aspires to the democratic ideal of getting people themselves to connect the dots of information with their priorities and pre-existing preferences and then make up their mind, i.e. engage in belief-formation, and decide whether they agree to what is being asked of them or not.

However, situations exist where people have both been provided the necessary information and hold the right beliefs and attitudes, relative to a desired behaviour, yet still do not do it. Sticking to the assumption that people are rational, the traditional policy escalator assumes for such situations that the necessary extrinsic incentives for doing the right thing is not in place; i.e. it makes for assuming it to be a *problem of incentivisation* that keeps people from doing the right thing. Methodologically, this can again be sought validated, this time in terms of economic analysis of the incentive
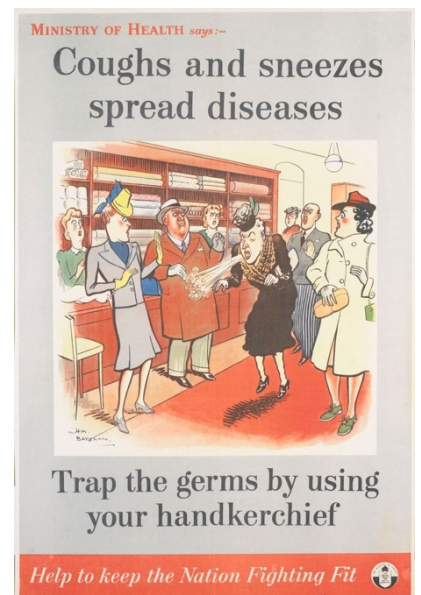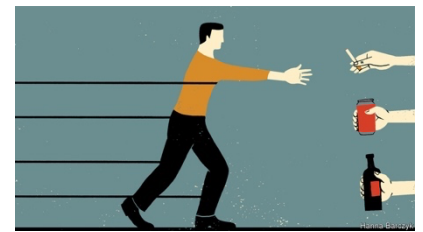

Public health campaigns are a classic example of behaviour change interventions that assumes people need to be persuaded to do the right thing. Here by appealing to patriotism and the wish to not be frowned upon.

structure surrounding the behaviour. If this reveals incentives are not aligned with the desired behaviour, i.e. that reasonable incentives for doing the right thing are not in place, then the obvious policy strategy to pursue is that of aligning incentives with the preferred behaviour. However, on a practical level, pursuing such alignment through positive or negative incentivisation, this strategy is usually more costly and controversial than the two first strategies on the escalator. Alignment requires monitoring conformity, additional administrational costs and different from the mere provision of information and rational persuasion, policy makers will also be rewarding those conforming to behaviour preferred by them, as well as sanctioning those that do not. This will make things quite controversial if it turns out that nonconformity is not only a matter of non-aligned incentives, but also an attitudinal problem, as it means that policy makers basically will be favouring those citizens who agree with them and punishing those who do not. Still, policy makers may easily be attracted to apply negative financial incentives, especially since such taxes create substantial tax revenues.
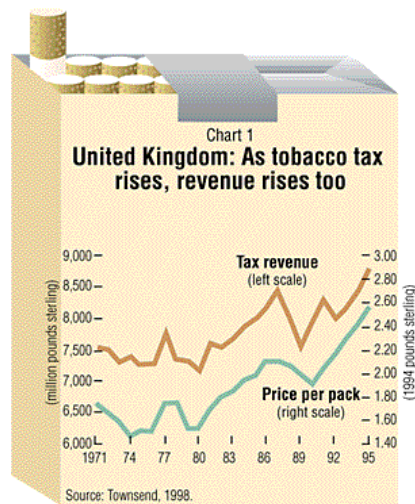
Tobacco tax is a classic example of a behavioural change intervention that works on the assumption that an undesired behaviour persists due to an incentivisation problem. While such taxes have been shown to be effective in getting people to quit smoking, the tax levels needed to create such effects often have to be incredulously high. This also means that taxes such as those imposed on tobacco create substantial revenues for the state that ironically may disincentive policy makers from taking further steps aimed at changing people's behaviour too fast. A similar paradox exists relative to other sin taxes on consumption goods such as gasoline, energy for housing, sugar and fat, road pricing, and speeding in traffic. In one interpretation, this makes policy makers turn up consumption taxes in a pace that maximises revenue, without chasing citizens away from consumption.

Sin taxes: A sin tax is an excise or sales tax specifically levied on certain goods deemed harmful to society and individuals, for example alcohol and tobacco, candies, drugs, soft drinks, fast foods, coffee, sugar, gambling and pornography. Two claimed purposes are usually used to argue for such taxes. Sin taxes are used to increase the price in an effort to lower their use, or failing that, to increase and find new sources of revenue. Increasing a sin tax is often more popular than increasing other taxes. However, sin taxes have been criticized for, amongst other: increasing smuggling and black markets; discriminate against lower classes; leading people to consume lower quality and higher risk products; creating a reliance of government on the undesired behaviour; being used to finance unrelated policies or self-defeating programs, such as when tax on cigarettes are used to finance stop smoking campaigns. Source: Wikipedia "Sin Tax".

Finally, if neither providing information, rational persuasion or aligning incentives work to change people's behaviour, the assumption that people are rational may seem somewhat at a loss. In that case the traditional policy escalator, escalate policies by applying hard regulatory measures *viz.* prohibitions and bans. From a strategic perspective, such measures have several rational functions. The common-sensical one is that bans and prohibitions formally takes away choices that policy makers dislike. However, as the actual removal of choice is easier said than done, or may be impractical for wider purposes, hard measures such as bans and prohibitions usually serve their main function as signals to potential dissenters that authorities are willing to apply negative and social sanctions up to the level where one's the freedom to choose beyond the fences of jail are removed and one is ostracised from the group that makes up society as such. Needless to say, though, the strategy based on bans and prohibitions is usually a last option applied only when a behaviour poses a risk of harm to others. Not only is it costly – policing the streets, maintaining a legal system as well as providing detention does not come for free – but it is also as invasive to individual freedom as can be.

For decades governments around the world have fought cannabis smoking with prohibitions. However, as have previously been the case with nation-wide alcohol prohibitions, the war on drugs, and underage drinking the effects of banning the behaviour are often argued to be limited compared to the costs. In addition, bans and prohibitions have also given rise to increased crime levels, smuggling, consumption of lower quality and higher risk products, and for some behaviours, even seems to have added to the attractiveness of engaging in the behaviour for certain groups.

Anomalies in rational regulation

Despite the fact that the rational regulation of traditional public policy directly echoes the classical economical model of man as Homo Oeconomicus and self-evident axioms about what it is to act rationally, one may still ask – especially in light of the side notes above – whether the approach is effective beyond the blackboards of economic professors and, if so, what strategies work best.

Needless to say, this is a question that is almost impossible to answer. Regulatory efforts usually mix strategies when targeting a given behaviour making it difficult to compare their relative effectiveness. This is partly because policy efforts tend to add up layer upon layer of increasingly harder rational regulatory strategies over time; and partly because the reasons for citizens not changing behaviour are often heterogenous, i.e. in a given population, the reasons for people to act, or not act, in a particular way may for some be an informational problem, for others an attitudinal problem, for others again, a problem of incentivisation, and yet for others a lack of hard measures to rein their temptation.

That said, the best answer to whether rational regulation works and what strategies are effective is: "it seems to depend". Beginning with creating general awareness and information, as the lifeguards at the famous Bondi Beach in Australia says, "nothing clears the surf at Bondi faster than the shark alarm"; but then again, writing "smoking kills" on cigarette packages have had no similar effect on smokers. Relative to rational persuasion, years of intensive campaigning have convinced most people of the stupidity in drinking and driving, yet some of the very same people still do not apply sun-screen when going to the beach, despite similar campaigning efforts relative to this behaviour. Next, positive and negative incentivisation. People may put in the extra hours at work for additional money and pay for parking in fear of getting a ticket; however, despite generous government subsidies in many countries for citizens making energy efficient home improvements, this has usually had little effect, and likewise goes for fines against jaywalking.[i] Finally, the ban on indoor smoking has worked incredibly well in most countries; yet similar bans on smoking Marihuana, whether indoor or outdoor… not so good.

Given the rational underpinnings of the traditional policy escalator such regulatory failures come to count as *anomalies* in need of explanation. An attractive answer is just to point to cases of regulatory failure and try to explain them away by arguing that these did not work out as predicted, because the information and arguments were not presented clearly enough; that the incentivisation scheme was not rewarding or deterrent enough; or that the bans or prohibitions put in place were not sufficiently policed and sanctioned. Thus, what explains the anomalies, on this account is an undersupply of rational regulation. The reason why this answer is so attractive is that it actually follows from sticking to the assumption that people are rational, as this is what allows one to deduce the insufficiency of regulatory efforts from their ineffectiveness.

Yet, there are two dangerous illusions that this account may give rise to: one is, it is immune from criticism if made a priori; the other is, that it may give rise to the adoption of increasingly unnecessary invasive policies if subsequently used to inform policy. To elaborate; the undersupply of rational regulation can always be used to explain away

any anomaly if undersupply is merely deduced from the lack of impact from rational regulation plus the assumption that people are rational. Of course, the cost of this is immunisation from falsification, which renders the claim empirically impotent from a scientific point of view. Yet, like everyone else, policy makers tend to fall prey to such illusions of infallibility. The problem is, however, that this may mindlessly give rise to the adoption of further rational regulation that tend to make itself a self-fulfilling prophecy if left to its own devices. Ultimately, held at gun-point most people are willing to engage almost in whatever behaviour change you ask of them. Yet, the fact that you can open a door with an elephant doesn't make it a key; similarly, that you can push regulation to the point where it works, does not prove a pre-existing undersupply.

To avoid these attractive illusions when claiming that rational regulatory undersupply may explain anomalies, one needs to ask instead whether *reasonable* levels of information, reasons, incentives and hard regulation, respectively, are in place in the relevant case – that is, reasonable in the sense that it would be sufficient for a Homo Oeconomicus to change behaviour. Rather than deducing a priori from the lack of regulatory impact and the assumption of citizens being perfectly rational, that this is not the case, this means that the argument to be made becomes a quasi-empirical one. In doing this, the approach opens up for a more detailed account of regulatory undersupply that may explain anomalies of rational regulation in terms of 'regulatory mismatch'; that is, by claiming that regulatory success depends on whether a regulatory effort falls in the strategic domain that matches problems correctly. Incentivising further without awareness or agreement; creating further awareness and making rational arguments when incentives are not aligned; or implementing bans or making prohibitions without informing people, would be cases of mismatches that could explain rational regulatory failures. Matching informational problems with creating more awareness and providing information; attitudinal problems with rational persuasion; incentivisation problems with alignment of incentives and possibly bans or prohibitions, would then explain successes of rational regulation.

The empirical approach of asking whether *reasonable* levels of regulation exists for each strategic domain relative to a given behaviour allows for mapping potential undersupply in the relevant dimensions. That is, by asking for each domain whether information, reasons, incentives and existing bans and prohibitions are in place sufficiently for a rational agent to change his or her behaviour, one can identify those domains that suffer from undersupply and direct further efforts to these.

But this is not all. What such a mapping may also reveal is that for many anomalies of rational regulation, from cigarettes over sunscreen to filing your tax on time, reasonable levels of information, arguments, incentives and regulation are in place, yet people do still not perform the desired behaviour. That is, behaviours exist, where people continue to exhibit a certain problematic behaviour even though the rational regulatory potential has been exhausted. What we have on our hands in such cases are true anomalies to rational regulation. Such anomalies challenge not only traditional regulation, but also the practical usefulness of rational choice theory, as they should not exist from a theoretical point of view.

The Behavioural Turn

However, while rational public policy is only starting now to show signs of a significant crisis due to an accumulation of anomalies, the challenges to rational choice theory and the ensuing theoretical crisis already started to reveal itself in Academia almost as soon as the theory had ascended to the throne as the dominant paradigm informing public policy. Following (Cartwright, 2011) these challenges came in three waves that taken together with the publication of *Nudge* gave rise to what will be referred to here as the Behavioural Turn in Economics, and in turn, in Public Policy as well.

In the first wave, it was pointed out that the increasingly complex rational computations assumed by rational choice theory on behalf of people in order to explain even some of the most trivial behavioural phenomena were unrealistic (Simon, 1955). In the second wave, it was revealed that even if, allowing for such complex computations, the theory suffered from internal inconsistencies and problems rendering it helpless to explain on its own, simple behaviours of social coordination and collaboration fundamental to social organisation, the so-called 'multiple-equilibrium selection problems' (Schelling, 1960). Third, and more constructively, it was shown that when studying human behaviour empirically, even under controlled experimental conditions, people did not always reason and act as the theory predicted (Tversky & Kahneman, 1974). As the insights provided disseminated more broadly the result became the behavioural turn in a series of disciplines, especially economics and law. With the fourth, and ongoing wave, public policy has become subject to this behavioural turn. In the following, four basic pillars of this turn are briefly summarised.

Our rationality is bounded

The notion of *bounded rationality* was developed by the American psychologist and economist Herbert Simon (1916-2001). It refers to the general insight that the reasoning involved in human judgment and decision-making is constrained and moderated by the cognitive limitations of decision-makers as well as the finite amount of time and information available to them when making their decisions.

To illustrate, consider the game of chess. Chess is a game of complete information and it is known that there exists an optimal strategy for playing a game of chess. The only problem is that the human brain does not have the computational capacity and power to calculate what this is. This point is easy to grasp. Imagine yourself being in the position of White and thus about to make your opening move. As chess is a game of strategy you might decide to try to think ahead, but since there are 20 different ways of making the first move, just considering what your second move might be, requires you to keep 20 x 20 = 400 different games (20x20) in mind. In fact, after the second pair of turns, there are 197.742 possible games to keep in mind, and after the third pair of turns, 121 million. Then imagine that a time constraint such as 1 minute for each move is added. You don't even need to imagine what the implications would be if further adding incomplete information about your opponent's moves, as is often the case in situations of everyday decision-making, to get the point. Our cognitive limitations and the finite amount of time and information available when making even simple choices keeps us from reasoning our way to optimal, i.e. rational decisions.
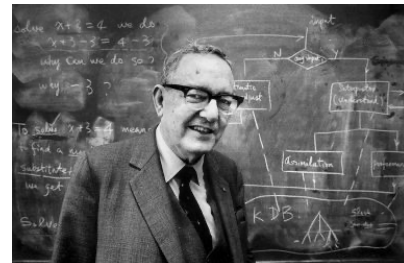
Herbert Alexander Simon (1916 –2001) was an American economist, political scientist and cognitive psychologist. He received the Nobel Prize in Economics in 1978.
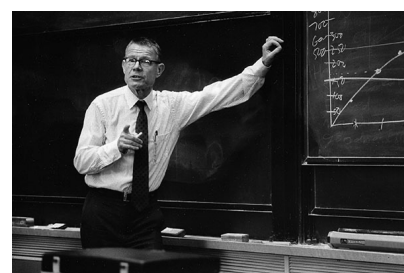
Thomas Crombie Schelling (1921 – 2016) was an American economist. He was awarded the 2005 Nobel Memorial Prize in Economic Sciences (shared with Robert Aumann).

Left: Amos Nathan Tversky (1937 –1996) was an Israeli-American cognitive and mathematical psychologist, a student of cognitive science, and collaborator of Daniel Kahneman; Right: Daniel Kahneman (1934) is an Israeli-American psychologist and economist notable for his work on the psychology of judgment and decision-making, as well as behavioural economics, for which he was awarded the 2002 Nobel Memorial Prize in Economic Sciences (shared with Vernon L. Smith). In 2016 Michael Lewis published the book 'The Undoing Project' (2016) about the fascinating friendship and work of Tversky and Kahneman.

We rely on heuristics in judgment and decision making

If our bounded rationality usually prevents us from reasoning our way to making optimal judgments and decisions, then how do we navigate the world? According to the behavioural sciences we rely heavily on heuristics (Kahneman, Tversky, & Paul Slovic, 1982). Broadly speaking the concept of a *heuristic* refers to any practical approach or method that one may employ to solve a problem that does not rely on watertight logical and rational reasoning, yet still is sufficient for reaching a satisfying conclusion (Myers, 2010). Herbert Simon referred to the use of such heuristics as 'satisficing' in contrast to rational 'optimising' (Simon, 1956). Again, think of chess. Chess players apply heuristics in playing the game, such as the Sicilian Defense, King's Gambit or the Venice Attack. In psychology, the concept of heuristic has come to denote simple and efficient, hard-wired or learned mental rules and shortcuts that ease the cognitive effort involved in making judgment and decisions. A heuristic may be the seemingly hardwired propensity for humans to seek-out, focus upon and better remember negative information or learned stereo-types for initial categorisation of people or situations, such as the game of chess you are playing.

> A **heuristic** is a simple and efficient, hard-wired or learned mental rule or shortcut that ease the cognitive effort involved in making judgment and decisions. Such 'rule-of-thumb' strategies decrease the time it takes to make decisions and allow people to deal with the thousands of potential decision points they encounter every day. Heuristics are in general helpful and accurate, but they can also lead to systematic mistakes aka cognitive biases.

Our thinking is subject to cognitive biases

The concept of *cognitive bias* refers to the third general insight, that human judgment and decision making may systematically come to deviate from the predictions based on the rules of epistemic and instrumental rationality because our reasoning processes rely on heuristics and other simplifying psychological mechanisms. To illustrate, think of a person, who apply a heuristic in decision making called *social proof* (Cialdini, 1984), saying that when facing choices involving informational uncertainty, then look to what other people say and do because they will tend to have relevant information and thus know what they are talking about or doing. While this heuristic might generally work quite well to efficiently arrive at satisfactory conclusions, it may also lead people systematically astray when the boundary condition – "other people tend to have relevant information and thus know what they are talking about or doing" – is not satisfied or the social cue informing the heuristic is manipulated. This might for instance happen, if other people irrationally come to fear a vaccination programme for causing autism, or if a mediocre book is deliberately placed on the top of the best-seller list in turn turning it into a bestseller (Hendricks & Hansen, 2016). It is the continuously interplay in our reasoning between what ought to be, from the perspective of rationality, irrelevant heuristics and psychological mechanisms with contextual features of complex decision-making that may make our judgment and decision-making systematically deviate from what is predicted by rational choice theory. Confusingly, the different ways this happens are referred to as *cognitive biases* as well.

The availability heuristic: In their classic paper Judgment under Uncertainty: Heuristics and Biases (1974) Kahneman and Tversky described three heuristics that are employed in making judgements under uncertainty: the 'representativeness heuristic', the 'anchoring and adjustment heuristic', and 'the availability heuristic'.

Representativeness and anchoring-and-adjustment are discussed later, here we look at the availability heuristic.

The availability heuristic refers to a mental short-cut where the likelihood of something is based upon how easy it is to bring something to mind. For instance, if you are considering whether to fly or drive on Holidays you may suddenly come to think of a number of recent airline accidents and then feel that air travel is too dangerous because those examples of air disasters came to mind so easily. That is, being more readily available in your memory, examples of air disasters make you judge these outcomes as being more common or frequently-occurring than is the case making you choose to go by car.

In the months after 9-11 where two planes crashed into The World Trade Center in New York many US travellers seemingly applied the availability heuristic and decided to travel by car. This caused an increase in traffic accidents, as travelling by car is far more dangerous than by plane (Gigerenzer, 2004).

As history would have it, the three insights addressed above were largely established as a result of debates within economics; debates which ultimately resulted in the birth of the discipline of behavioural economics (Cartwright, 2011). This also explains why behavioural economics later has been seen as the main contender of traditional rationality based public policy. Why did these challenges not come from psychology? According to Kahneman this is because few psychologists had ever thought to bother with the theory underpinning the Homo Oeconomicus model informing traditional public policy, as this was so obviously out of tune with what they could readily observe (Kahneman, 2011). However, as the behavioural turn has picked up speed, psychologists have joined the party bringing decades of work to the table. Thus, the fourth pillar of the behavioural turn is firmly rooted in psychology.

We are creatures of habit

The fourth basic pillar of the behavioural turn comes directly from psychology and is thus rooted outside the debate in economics about rational choice theory. This pillar concerns that much of our routine behaviour occurs without the intentional control of processes or the consideration of pros and cons according to the rules of rationality. Instead such routine behaviour is habitual meaning that its performance reflects the routine repetition of past acts that is cued by stable features of the environment. This may be the habit of where you put your keys after arriving back home, how you tend to start a conversation, reaching out for your phone during dinner, or the route you take to work. Keeping the question aside of whether behaviours and habits can be quantified at all, psychologist Wendy Wood and colleagues have assessed that approximately 43% of our daily behaviours are performed out of habit (Wood, Quinn, & Kashy, 2002).

Beyond its description as a routine repetition of past acts, a *habit* may be defined more precisely as a learned automatic response to a contextual cue the likelihood of which increases with the number of repetitions up to a certain limit and which is insensitive to short term changes in goals (Wood, Labrecque, Lin, & Rünger, 2014). First of all, this means that habitual behaviour needs to be learned or trained – a process referred to as *habit formation*, through which it becomes automatized. Second, it means that its execution is conditioned by an internal or external cue without which the behaviour is not activated. Third, it means that while habit formation is goal-dependent, its performance might persist even if goals changes or if better alternative measures for goal achievement appear on stage. Consequently, habitual behaviour does not necessarily conform to the predictions of rationality-based theories, nor reflect the mindful internal workings of the model of man suggested by Homo Oeconomicus.

Taken together these four pillars documents how human behaviour often systematically differs from the predictions of rational choice theory. Human behaviour and decision making it turns out, is often much better understood in terms of realistic models of Humans as we know them from our everyday lives, rather than the rational ideal of Homo Oeconomicus. But what does these more realistic models look like? The answer to this is to be found in the so-called dual process theories of reasoning, judgment and social cognition, which the behavioural turn has adopted from cognitive- and social psychology.

Dual-process theories of human cognition

To structure these general insights of the behavioural turn theoretically, behavioural scientists have looked to cognitive- and social psychology from where they have adopted alternative models to the traditional rationality-based ones. The core theories underpinning these models are usually referred to as *Dual Process Theories of Cognition*. Common to these theories are that they seek to understand the processes involved in actual human reasoning, judgment and social cognition by positing how thought can arise from (at least) two different types of information processing in the mind (Evans J. S., 2008). In turn such dual process theories may explain why and how, and predict when, thinking does not reflect normative models of rational judgement and decision making.

Historically dual process theories have emerged from largely disconnected literatures and experiments in cognitive and social psychology. In the BI literature it has received attention from the general public with Daniel Kahneman's popular intellectual autobiography *Thinking, fast and slow* (2011) covering his work with Amos Tversky's, which led to Kahneman receiving a 'Nobel Prize in Economics' (Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel) in 2002 (shared with experimental economist Vernon L. Smith). Dual-process theories vary greatly but generally share the overarching structure of positing two types of human information processing — automatic (System 1) and nonautomatic (System 2) processing — in explaining and predicting human behavior (Evans J. S., 2008). In the most widespread version of dual process theories, Kahnemans and Tversky's *Dual Systems Theory*, System 1 processes are characterised by being fast, parallel, automatic, intuitive and based on leaning by doing, while System 2 processes are characterised by being slow, serial, controlled, deliberative, rule-governed, and subject to conscious rule (Kahneman, 2002), (see the figure below).
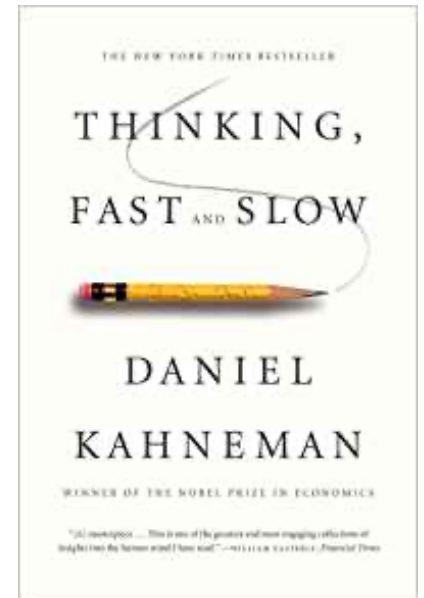
In 2011 Daniel Kahneman published his intellectual autobiography Thinking, Fast and Slow (Kahneman, 2011). 'Fast thinking' refers to System 1 processing and 'Slow thinking' refers to System 2 processing. The book popularised the use of dual process theories to explain cognitive bias in judgement and decision making and stayed at the top of bestseller lists around the world for years.

Kahneman and Tversky's Dual Systems Theory as presented in Kahneman's Nobel Prize Lecture 'Maps of Bounded Rationality' (Kahneman, 2002)

The main purpose of dual process theories is to study the interplay of non-rational features, automatic processes and reflective reasoning (Gawronski, Sherman and Trope, 2014). System 2 represents those reflective reasoning processes where one thought logically leads to another and which seemingly aspire to the ideals of rationality. Yet, the computational power, working capacity, mental resources and time required for reflective thinking to operate is limited and hence captures the insight that rationality is bounded. It also explains why System 2 processes are not activated automatically: they are too costly to run uncontrolled. This also explains the division of labour that exists between System 2 and System 1 thinking. The processes involved in System 1 thinking are fast, energy-efficient, processes that may relieve the pressure on and need for reflective thinking. The habits of the mind, whether habitual or heuristic, constitute exactly such processes allowing for a highly efficient autopiloting of attention, judgment, decision-making and self-control. Most of the time these automatic processes work satisfactorily and may even align with what would have been the average outcome of reflective or rational thinking. For this reason, habits and heuristics are often characterised as being *ecologically rational*, in the sense that they are on average optimal given individuals' priorities, circumstances and constraints (Gigerenzer, Todd, & The ABC Research Group, 1999). Yet, they may also come to influence our decisions and behaviour in sub-optimal ways resulting in cognitive biases, i.e. the systematic deviations from the predictions of rationality, that may lead us astray.

Compared with traditional rationality-based theories of human behaviour, such as rational choice theory, dual process accounts of human cognition currently seem to be superior to these on several traditional parameters for evaluating scientific accounts. While rationality-based theories may be intuitively attractive because they always seem to allow one to come up in retrospect with a story of how a behaviour fits them – "He failed his diet because he *really* didn't prioritise it" and "She didn't follow procedure because she *actually* hadn't understood the information" – such immunisation to falsifiability is not a virtue in the empirically grounded behavioural sciences. Not that dual process theories are directly falsifiable either – no theory is – but they are in some sense simpler; often possess better explanatory as well as predictive power; both of familiar and unfamiliar anomalous phenomena that has puzzled economists and behavioural scientist; and it fits our evolutionary accounts of human nature.

Yet, one should also be careful, as dual process accounts are prone to cause interpretational mistakes and invites for overinterpretation. For instance, one may easily be fooled into thinking that System 1 and 2 refers to two distinct 'systems', that a behaviour is driven by either one or the other, that system 1 thinking is a mere servant or substitute of system 2 thinking and that it is System 1 that causes all the biases in behaviour to occur. While attractive, these are mistaken beliefs. For instance, the processes involved in System 1 precedes System 2 processes in evolutionary time. Consequently, rather than System 2 being in charge, it often serves the needs and aspirations of our more basic natures, e.g. by making us prone to believe what fits our wants as well as cunningly act on temptation and temperaments. But much more on this in Chapter 2.

The Bat-and-Ball puzzle.

(1) A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball.

How much does the ball cost? _____ cents

In a famous experiment Daniel Kahneman and Shane Frederick (Kahneman & Frederick, 2002) presented this question which they had informally found most people to answer wrong by saying "10 cents", while the right answer is actually "5 cents". They used this as an illustration of System 1 vs. System 2 thinking arguing that the answer "10 cents" results from intuitive system 1 thinking, while the correct answer "5 cents" is arrived at through applying analytical reflective thinking processes characterising System 2 processing.

Since then the bat-and-ball puzzle has been posed to thousands of university students around the world, with shocking results. More than 50% of students at Harvard, MIT, and Princeton have been found to provide the intuitive, but incorrect, answer. At less selective universities, the frequency of incorrect answers has been found to be in excess of 80% (Kahneman, 2011).

What is noticeable though is an issue that will occupy us in Chapter 2; not everyone gets the answer wrong. This is called 'individual variance' and has been found to correlate with various factors. One controversial example is the finding that people who provides the intuitive answer are more likely to believe in religion, while those who prove analytical answers are less likely to believe in religion. Conversely others have found that the ball-and-bat puzzle is really just a decision-trap triggered by a very specific word-illusion. In asking the almost identical question "A bat and a ball cost $110 in total. The bat costs $10 more than the ball" experimenters found most that most subjects answered correctly. This suggests that, while the puzzle is illustrative of system 1 vs. system 2 thinking, its real-world relevance is also quite limited.

Some methodological characteristics and implications

Different from a rationality-based approach to studying human behaviour, the uncovering of behavioural insights as conducted by the behavioural sciences as well as their conceptualisation within dual process theories of cognition cannot be done from the armchair. What it means to act rationally and what this implies for the instrumental rational pursuit of one's goals under different conditions may be examined by pure reason. As the rational choice paradigm of the 20th Century revealed, rationality can be axiomatized and its consequences (usually) deduced under various assumptions. Studying how and why human behaviour may systematically deviate from the norms of rational reasoning and predictions, though using rationality as a baseline, is an empirical effort.

Can we rely on self-reports about behaviour?

However, the effort to empirically study behavioural insights is complicated by the fact that the study of cognitive biases and heuristics as well as dual process theories of cognition posits constructs that is neither readily observable to the eye – whether speaking of the inner eye or the regular ones in people's foreheads – nor easily described in everyday language. This has some huge and crucial methodological implications. For one, if System 1 processes are characterised by being fast, intuitive, automatic, and, not to mention, by and large unconscious, and system 2 processes are characterised by being limited in capacity and memory, then it follows that people cannot easily self-report on what prompts and drives their behaviour. Second, even when subjects are asked, the theory holds that answers are likely to be warped by biases and characterised by ex-post rationalisations. Third, and finally, since System 1 processes and mechanisms as well as their interaction with System 2 processes, defies the rules of rationality and are not structured linguistically, people are hard challenged to make sense of and provide valid accounts of their behaviour in everyday language. Taken together, all of this fly in the face of the methods usually applied by rationality-based approaches, especially the widespread use of self-reports, including interviews, focus groups and surveys. Add to this, that since the link between thinking and behaviour is non-trivial on dual process accounts, one cannot either take observed behaviour as reliably revealing our preferences and beliefs as usually done in economics.

For these theoretical reasons, it is not coincidental that the behavioural sciences that BI rely upon share not only the inductive, but also the causal explanatory and experimental approach to the subject matter of human behaviour (nor is it coincidental that such a methodological approach produces such theories). In particular, the behavioural insights pursued have been brought to life in controlled laboratory settings, where subjects can be studied in true experimental designs, meaning that they are randomised into treatment and control groups and are measured and observed on a series of variables allowing for studying and testing psychological theories and insights. Thus, the Randomised Controlled Trial (RCT) is held as the golden standard within the behavioural sciences, as defined here, that allows for the controlled study of what psychological mechanisms mediate between independent variables and behavioural effects as well as what factors may condition and moderate such relationships.

Having covered part of the substantial practical and theoretical background for the behavioural turn in economics and public policy, as well as of the science underpinning the book *Nudge*, we can now return to the story of BI movement.

Behavioural Insights spreads around the world

Different from most suggestions made by scientists, Thaler and Sunstein's idea was quickly picked up in policy circles. A contributing factor to this happing was the promise of cheap, effective and measurable improvements in the wake of the financial crisis. Historically, the institutionalisation of the systematic approach to applying behavioural insights and thus the beginning of the BI movement began with the creation of the Behavioural Insights Team (UKBIT) in the UK in 2010. UKBIT initially acted like an internal consultancy for UK policy makers (Lunn, 2014) and was institutionalised as part of the UK Prime Minister's Office and the Cabinet Office. Headed by David Halpern, former Director of the UK Institute for Government, it was the first government institution dedicated to applying behavioural insights to policy and public administration (Halpern, 2015). With originally seven members, the UKBIT was famously set three objectives to achieve to avoid triggering a 'sunset clause' that would see the team shut down on its two-year anniversary: (1) transform at least two major areas of policy; (2) spread an understanding of behavioural approaches across Whitehall; and (3) achieve at least a tenfold return on cost (Sanders, Snijders, & Hallsworth, 2018). UKBIT succeeded in doing this by translating the best-evidenced interventions from the behavioural science literature with a focus on revenue-producing and money-saving projects, thereby providing proof of concept for the application of behavioural insights in public policy development, design and delivery.

From the beginning the UKBIT had a strong focus on nudges. This is witnessed by the fact that the team has gone under the nickname "The Nudge Unit" since its establishment. However, the UKBIT also made additional contributions to the emerging field of behaviourally informed public policy. In particular, the team committed very strongly to an evidence-based approach to policy development, where policies were not only behaviourally inspired or informed, but also subject to the evidential standards of behavioural science, in particular the experimental approach.

Chief Executive David Halpern (right) and Managing Director Owain Service (left) of the UKBIT. The UKBIT started in 2010 as a 7-member team part of the UK Prime Minister's Office and the Cabinet Office. In 2014 the team rolled out of government to become a private company. Today it employs almost 200 people in the UK, Australia, Singapore, the US and Canada.
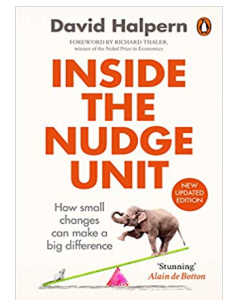
The fantastic first years of the journey of UKBIT is described in David Halpern's well-written 'Inside the Nudge Unit' published in 2015.
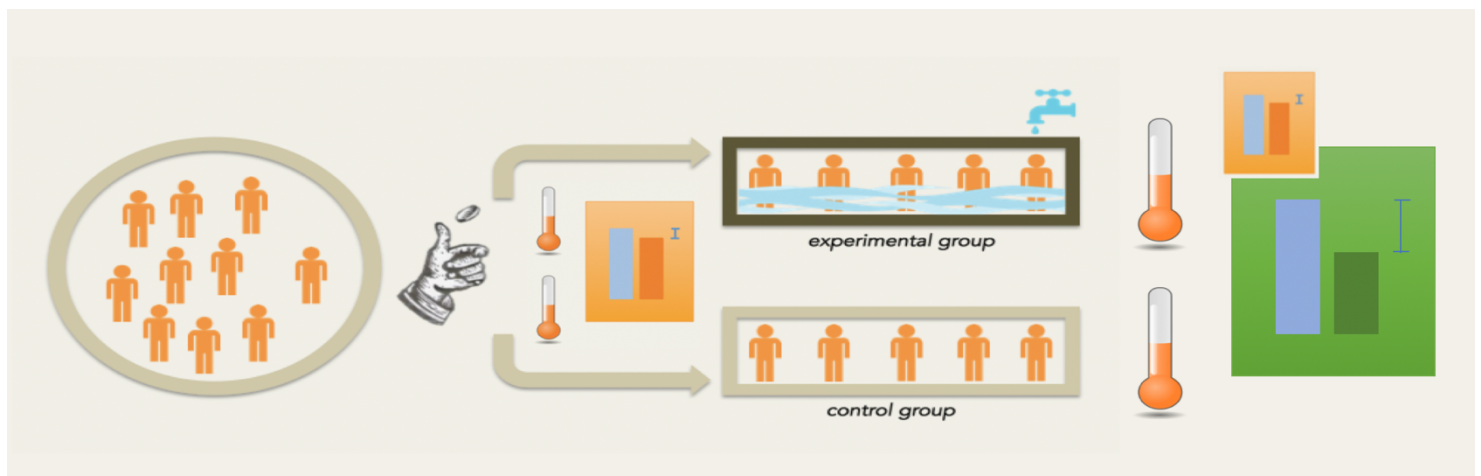
The Randomised Controlled Trial (RCT). *The RCT is the golden standard in the experimental approach. It consists of randomising experimental subjects from a sample into at least two groups; an experimental or 'treatment' group which is exposed to the treatment, e.g. a policy, and a control group which is left untouched. This experimental design allows for measuring the effect of the treatment by comparing the subsequent behaviour of the experimental group with the behaviour of the control group (potentially adjusted for pre-existing differences between groups, when sample size is small).*
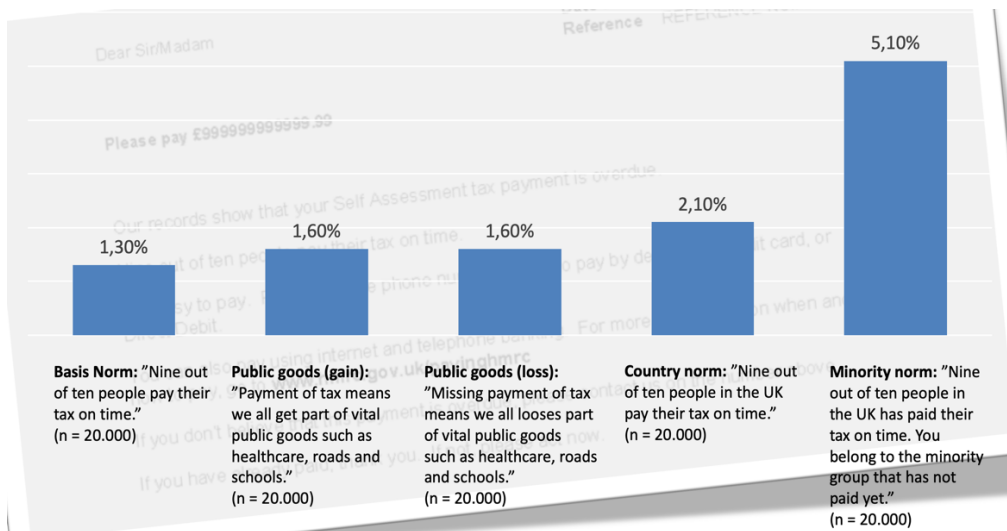
One of the most famous UKBIT experiments aimed at nudging UK citizens to file tax on time by tweaking one of the opening sentences in the beginning of a letter from the tax authorities. The various tweaks were tested in a randomised controlled field-experiment involving more than 100.000 citizens in 5 treatment groups. In agreement with the theory of social proof, the most succesful intervention was the one that illicited the in-group-out-group social norm by telling that most people filed their tax on time in the group with which you identify (UK tax payers). Compared to the population control this led to 5,1% increase in the number of people who filed their tax on time.

Richard Thaler (1945), Chicago Booth, 'the god-father' of behavioural economics and recipient of the 'Nobel prize' in Economics 2017. Together with Cass Sunstein he wrote 'Nudge' (2008) after years of conducting experiments and publishing research about how psychological factors influence economic decision making in ways that make people act irrationally. His intellectual autobiography 'Misbehaving' was published in (2015).
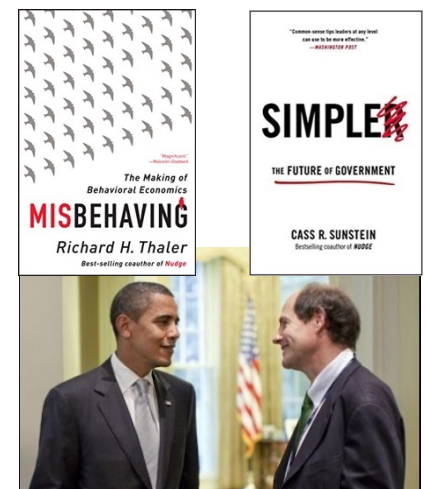
Cass R. Sunstein (right), (1954), Harvard University, is perhaps the worlds most quoted living legal scholar. Sunstein taught at Chicago University when he met and co-authored 'Nudge' with Richard Thaler. In 2009 Cass Sunstein was appointed the head of the White house Office of Information and Regulatory Affairs (OIRA). This appointment led to a series of initiatives, experiments and, ultimately, a presidential decree for integrating behavioural insights in US regulation. The ways that behavioural insights informed public policy under Sunstein's appointment is described in his book 'Simpler: The Future of Government' (2013).

Approximately at the same time that UKBIT was established, Cass Sunstein was appointed Administrator (2009-2012) of the White House Office of Information and Regulatory Affairs in the Obama administration (OIRA), see (Sunstein, 2013). During his mandate, OIRA promoted a series of disclosures and important simplifications in the regulatory process based on behavioural insights as described by Sunstein in the book *Simpler: The Future of Government* (2013). This work became the precursor for the creation of The Social and Behavioral Sciences Team (SBST) in the White House, which like the UKBIT, was aimed at translating findings and methods from the social and behavioural sciences into improvements in US Federal policies and programs. The SBST functioned from 2015 to 2017 carrying out work in a way similar to UKBIT and resulted in a Presidential Executive Order #13707 directing Federal Government agencies to apply behavioural science insights to their programs to better serve the American people, as well as two annual reports (2015 and 2016) before dissolving in 2017 as a result of a new government administration.

From these two initiatives, UKBIT and SBST, it is clear that the emerging BI movement was to begin with a top-level centralised effort to integrate behavioural insights in public policy. A positive effect of this approach was the ability of initial efforts to focus on prime candidate projects for the application of behavioural insights, build buy-in, demonstrate effect, increase awareness, and not to be ignored, use prominent English-speaking platforms and spokespersons for promoting the approach. As a result, researchers, public policy-makers and officials around the world soon became inspired to aspire to pursuing efforts to integrate BI in their work as well, making the BI movement and the idea of BPP spread quickly around the world.

Looking back, 10 years later, one may observe that the BI movement has not only spread but also evolved extremely fast. In particular, it is notable that it has emerged in a wide variety of institutional settings as well as forms quite different from an initial setup of a centralized governmental unit in the UK. A recent mapping from the OECD thus identifies 200+ institutions inside as well as outside government applying BI to public policy around the world. These institutions include intra- and extra-governmental teams and networks, policy-academic collaborations and specialised consultancy teams amongst others, which has produced an impressing number of examples integrating behavioural insights in public policy around the world. This is particularly evident in the 2017 OECD case-collection that exhibits more than 150 reported cases of BPP (OECD, 2017). Thus, although the centralized governmental institutionalisation of BI might historically have been a catalysing force showcasing how behavioural insights may apply in public policy, it increasingly shows to be just one form of institutionalisation, which has been superseded by a wider movement beginning to penetrate public policy at all levels of government, public institutions and societies at large.
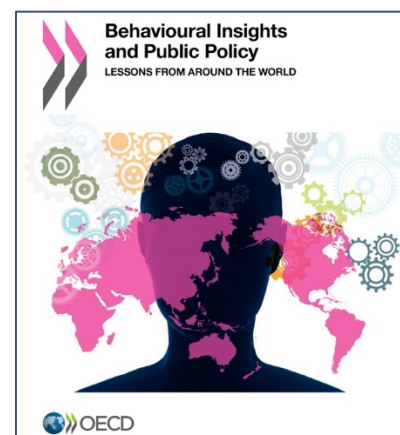
In 2017 OECD published the report "Behavioural Insights and Public Policy: Lessons around the world" summarising more than 100 self-reported examples of Behaviourally Informed Policy initiatives.

Frameworks and reports

Besides the promise of cheap, effective and measurable policy improvements and a long list of very well-written books on behavioural science, a central driver in the fast spread and evolution of the BI movement has been a series of frameworks, reports and paradigm results published by behavioural insights teams around the world. These tools of dissemination have helped to spread best practices, proofs of concepts and methodological standards, such as the important role played by randomised controlled trials (RCTs), to policy-makers and upcoming practitioners who would otherwise never have thought about venturing into the niche of evidence-based policy that BI constitute.

Most prominently, perhaps, is the MINDSPACE report (Dolan, Hallsworth, Halpern, King, & Vlaev, 2010). This report was published already in 2010 by the UK institute for government, then headed by the later Chief Executive of UKBIT David Halpern, and provided and early checklist for thinking about how nine well-evidenced behavioural principles may inform public policy design and delivery. Two years later the UKBIT published the report *Test, learn, adapt: Developing public policy with randomised controlled trials* (Haynes, Service, Goldacre, & Torgerson, 2012). This gave an easily accessible introduction to the basics of using randomised controlled trials in evaluating policy interventions. In 2014 OECD entered the scene by publishing the report *Regulatory Policy and Behavioural Economics* (2014) written by Pete Lunn, the founder and head of the ESRI Behavioural Research Unit in Ireland. The report provided a timely overview of the increasing attention paid to behavioural economics in policy design and delivery throughout the OECD countries. The EAST framework published by UKBIT in 2014 provided a simple framework for thinking about how behavioural insights may help design policies leveraging convenience, social aspects of decision making and the attractiveness and timeliness of polices (2014). In 2015 The World Bank chipped in with the impressive flagship report *Mind, Society, and Behavior* (2015) which gave a comprehensive overview of how the BI approach is of relevance to development policy. In addition, throughout all of these years, beginning with *The Behavioural Insights Team Update report 2010-2011* (UKBIT, 2011) UKBIT has also continuously published so-called 'update reports' summarising their work and experiments. Update reports has since then been published almost annually by UKBIT, see (UKBIT, 2015), (UKBIT, 2017), (UKBIT, 2018). The update format was also adopted by the US Social and Behavioral Science Team for its two years of existence, see (SBST, 2015), (SBST, 2016) and latest by iNudgeyou – The Applied Behavioural Science Group (iNudgeyou, 2019).

Finally, in 2017 the OECD published the comprehensive overview report *Behavioural Insights and Public Policy: Lessons from Around the World* (OECD, 2017), containing more than 150 case studies of the use of behavioural insights in public policy throughout the world (see above). This work by the OECD was followed up in 2019 with the report *Tools and Ethics for Applied Behavioural Insights: The BASIC Toolkit* (OECD, 2019) written by the present author, iNudgeyou and the OECD Behavioural Insights Team. Based on 10 years of experience with applied behavioural science, BASIC provides a framework and toolkit for practitioners for how to apply behavioural insights in public policy effectively and responsibly; the very same framework which is elaborated and described in this handbook.

MINDSPACE (2010), an early report presenting 9 behavioural insights of relevance to public policy.

EAST (2014) presents a series of behavioural insights for making policies Easy, Social, Timely and Attractive.

BASIC (2019) presents a comprehensive framework, including detailed tools, for applying BI to public policy.

Four approaches in Behavioural Public Policy

Tacking stock of the first 10 years of the BI movement in public policy reveals a series of lessons learned as well as challenges that have surfaced. Above was mentioned the lesson that BI may be institutionalised in many different ways. Another and perhaps more important one is that behavioural insights may inform public policy in multiple ways. Initially practitioners found it hard to answer policy makers what types of policies behavioural insights may give rise to and what it means for policies to be behaviourally informed. Now the contours of an answer are starting to crystallise. While nudges might constitute the most famous type of BI application to public policy, a series of other behaviourally informed approaches may also be identified as behaviourally informed, cf. (Hansen, 2019).

The Traditional Approach

Beginning with *the traditional public policy approach* explored earlier in this chapter, this approaches behaviour as the outcome of rational deliberation and decision-making, assuming people to have unbounded attention, processing capacity and willpower. It approaches behaviour change as the result of lack of information, absence of proper attitudes or lack of sufficient incentives and motivation. As a consequence, the traditional public policy approach approaches behavioural regulation by providing rational reasons for behaviour change, such as creating awareness and providing information (informational campaigns), presenting and arguing the case (persuasion campaigns), providing incentives (reliefs, rebates, taxation, and fees), and hard regulation (formalised prescriptions and prohibitions sanctioned by law).

During the first 10 years of BI, the traditional public policy approach has very much played the part of strawman as well as scapegoat. Many practitioners coming to the field of BI, especially those with few preconditions, have tended to frame the traditional public policy approach as the opposite of whatever they take BI to be. Thus, it has largely been portrayed as an inefficient, expensive and unrealistic approach to behavioural regulation compared to the efficient, cheap and realistic approaches that BI may appear to offer.

Although this is an attractive way to portray things, it is a misrepresentation of facts. The dual process theories of cognition underpinning BI do not discount the existence and importance of rational and reflective thinking. As a species we are just as much characterised by our ability to think rationally, as we are by our propensity to act irrationally. Thus, the provision of information, rational persuasion, incentivisation and hard regulation are strategies that have their place within BI. In particular, there is a long-standing *evidence-based policy* tradition, which have shown standard public policies to be effective in many situations. Informational campaigns do work to inform people; persuasion campaigns do work to influence people's attitudes; reliefs, rebates, taxation, fees and fines do work to make certain actions more attractive, and others less; and prescriptions and prohibitions sanctioned by law do work to change people's behaviour. Sometimes. 10 years of BI have shown that it is knowing when this 'sometimes' is, which is important. Applying the traditional approach should thus be viewed as part of BI when grounded in empirical evidence.

Evidence-based policy and BI.
In medicine it was quite early on recognised that a treatment better go through rigorous scientific testing before being approved for the market. From this idea of evidence-based medicine, the idea of evidence-based policy followed quickly: public policies should also be tested according to scientific standards before being implemented.

Today Evidence-Based Policy (EBP) is a strong tradition in public policy evaluation. It has ordered types of evidence into hierarchies and devised standards for testing policies and evaluating their impacts. This evidence shows that traditional rational regulatory approaches aimed at behaviour change often work. In so far as these policies integrate insights from the behavioural science about how people respond rationally to interventions, such work qualifies as BI and the resulting policies as BPP.
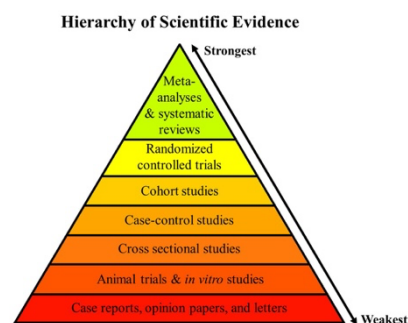
Hierarchy of Scientific Evidence: EBP orders types of scientific evidence in a hierarchy according to their strength. RCT goes into the top together with Meta-analysis, while case reports and opinion papers go into the bottom.

The Boost Approach

Different from a traditional, but behaviourally informed approach to public policy is the Boost approach. The boost approach originates in *the fast-and-frugal-heuristics school* (Gigerenzer, Todd, & The ABC Research Group, 1999), (Gigerenzer, Hertwig, & Pachur, 2011). This observes that behaviours and choices may deviate from the predictions of rationality-based theories, not because people's behaviour is irrational as such, but because they simply lack the necessary information, skills and competences (Hertwig, 2017). Consequently, the Boost approach focuses on how to make it easier for people to exercise their own agency in making choices by 'boosting' individuals' own decision-making skills and competences as well as adapting policies to our cognitive limitations, biases and habits. It ranges from strategies that require little time and effort on the individual's part to strategies that require substantial amounts of training, effort and motivation. Providing people with statistical skills, or presenting information to them in ways that make them less likely to be influenced by cognitive biases are instances of boost politics.

The Nudge Approach

The third behaviourally informed approach to public policy is *the Nudge approach*. As mentioned above, this approach has been a core driver at the centre of the BI movement. The nudge approach targets problems on the hypothesis that they result from cognitive biases and heuristics, which ultimately may prevent people from aligning their behaviour with their own desires and intentions. In particular, the Nudge approach aims to influence such behaviours by intentionally applying cognitive biases and heuristics, not only in the analysis of such behavioural problems, but also as strategic means to achieve behaviour change. It does this by integrating particular 'nudges' into aspects of the choice architectures within which crucial decision points are embedded. Unlike boosts, then, nudges target immediate individual behaviours, rather than building new decision competences or fostering existing ones.

The Protectionist Approach

Finally, there also exists the 'protectionist' approach in behaviourally informed public policy. This also works on the assumption that humans are subject to cognitive biases and heuristics in ways that may lead to poor decision-making. What the former approaches do not address, though, is the fact that the fallible nature of human cognition is often exploited by ill-intended actors. For this reason, the protectionist approach in behavioural public policy focuses on identifying contexts where behavioural insights are used to mislead, fool and manipulate people with the aim of providing legislation so as to protect their exploitation. The protectionist approach thus brings two new things to the table. First, and importantly, by adopting the view that we are humans rather than perfectly rational agents, it provides the theoretical foundation for understanding how citizens may be systematically exploited by ill-intended people, companies and organisations. Second, the protectionist approach seeks to devise legislative interventions that protect consumers from such exploitation. Especially, the EU has been at the forefront of this use of BI, e.g. by banning defaults, where customers are auto-subscribed to newsletters and the like, when making purchases online.
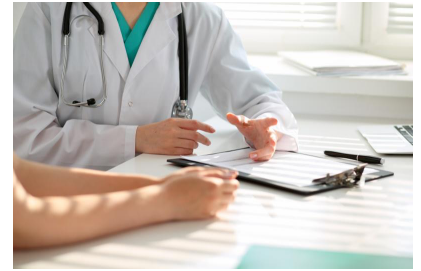
Studies have shown that people are better at making optimal decisions when statistical information is presented in frequencies rather than probabilities. Thus, the decision making of medical doctors and their patients may be boosted if a premise containing statistical information is put in the form of "Ten out of every 1,000 people die from this operation" rather than "The probability that one dies from this operation is 1%."

(1) Ten out of every 1,000 people die from this operation

(2) 990 out of every 1,000 people survive this operation

However, studies have also shown that peoples choices are significantly influenced by how they are framed, i.e. how a choice is formulated. In particular, people have been found to be 'loss averse' meaning that they seek to avoid options that described in terms of potential losses. In the case of the decision making of medical doctors and their patients this means that choices may be nudged, if a premise containing information about potential negative consequences as in (1) are framed instead in a positive formulation as in (2) to persuade patients,, provided that the choice architect has good reason to believe that having the operation with increase the health, wealth, and happiness of the patient.

Finally, one may also integrate behavioural insights in policy making by deciding to ban against its misuse. In the case of doctors and patients, policy makers may for instance decide, that doctors are not allowed to frame a given treatment positively.

Current Challenges to Behavioural Public Policy

However, while BI and BPP is experiencing increasing public attention it is also becoming evident that policy-makers and professionals, behavioural practitioners and researchers alike have a hard time of orienting themselves critically within this fast-evolving scientifically based paradigm. This is especially evident from four current challenges in BI.

Applying Behavioural Insights beyond proof-of-concept

In a recent issue of the journal *Behavioural Public Policy* celebrating the 10-year anniversary of UKBIT, Sanders, Snijders and Hallsworth (2018), all from the UKBIT, writes the following about the current situation in BI:

> There is much for proponents of behavioural science to be pleased with here, but it is difficult to argue that it constitutes a revolution. There have been behavioural insights teams that have failed to get off the ground, or which have been launched and failed to make a meaningful contribution – whether through contingent factors or deficiencies in ability. While behavioural science is much more widely used than it was, it has yet to sit alongside economics as a discipline dominant in the thinking of policymakers… there is a danger that behavioural science is seen to offer merely technocratic tweaks, rather than the more wide-ranging reassessment of public administration that could be possible.

The quote reveals a possible downside of the initial success enjoyed by implementing BI through high profiled centralized governmental units with a strong focus on *proof-of-concept* and intra-governmental cost-effectiveness (Hansen, 2018). This approach to implementing BI, together with the emergence of a series of easily digestible frameworks such as MIDSPACE and EAST, has led to a widespread perception amongst policy-makers and policy-professionals, that BI is best institutionalised as centralized governmental "nudge-units" that enters like *Mad Men* at the end of the policy cycle to deliver persuasive nudge-tweaks to increase the impact of already finished policy products. This idea is most likely to be wrong if BI is to deliver the more wide-ranging reassessment of public administration requested by Sanders, Snijders and Hallsworth; yet, despite the closure of SBST in 2017 and the fact that UKBIT rolled out of government in 2014 to become a private company, a centralized unit applying nudges persuasively remains the 'natural ambition' for many policy-makers and professionals, thereby continuing to leave out a proper discussion of how to apply BI as a real institutional capacity in public policy beyond proof-of-concept.

Applying Behavioural Insights: Persuasion or Problem-Oriented Solutions

Part of the craze and criticism that BI have been received with is due to its perception as a persuasion toolbox similar to that applied in PR, marketing and advertisement, where the main purpose is to get people to buy, say 'yes!' or conform to what one desires. The framing of behavioural insights as tools of persuasion or influence has a recent tradition stemming from the work of world famous social psychologist Robert Cialdini; in particular his book *Influence: The Psychology of Persuasion* (1984) and some existing frameworks such as MINDSPACE may be read in this vein. In recent years, however, there has been an increasing focus in the BI community on behavioural insights as tools for solving problems, i.e. BI are seen as offering a toolbox of problem-oriented solutions, rather than an approach to persuasion.

As Behavioural Insights and Behavioural Public Policy has grown academic journals devoted to this field are also starting to appear, one of the most prominent ones being the journal Behavioural Public Policy.

Proof of concept (PoC) is a realization of a certain method or idea in order to demonstrate its feasibility or a demonstration in principle with the aim of verifying that some concept or theory has practical potential. A proof of concept is usually small and may or may not be complete.

Mad Men *was a famous TV series running from 2007 to 2015. The drama follows one of New York's most prestigious ad agencies at the beginning of the 1960s which was the golden time of advertising industry. It was also a time where little thought when into what one was trying to persuade customers into buying. Thus, adds appeared with doctors advertising for cigarettes, babies for drinking soda and celebrities promoting alcohol and gambling. While doctors and babies are no longer allowed in these advertising roles, celebrities still promote*

Keeping the science in Behavioural Insights

Another challenge to the application of BI to public policy is the threat of the paradigm being watered down to comprise not just disciplines sharing the inductive causal explanatory and experimental approach as well as dual process theories of human cognition; but just about any policy activity that has to do with behaviour, based on any cluster of theories or disciplines, from anthropology over design thinking to big data analysis, using any kind of empirical methodology, from interviews, over focus group workshops to desk analysis. This water-down may partly be due to *the popularisation of behavioural science* and BI; partly because of social factors, such as newcomers entering the field preferring to have their pre-existing competencies qualify as BI competencies thereby making them immediate experts without additional knowledge acquisition; and partly because it may be easier and cheaper to apply pre-existing methodologies, rather than those involved in the behavioural sciences.

The reality of this threat may be illustrated by reference to another recent paper *Learning lessons: how to practice nudging around the world* by Osman et al (2018). The paper scrutinizes the 111 cases summarized, out of 159 submitted, in the aforementioned OECD publication *Behavioural Insights and Public Policy: Lessons from Around the World* (2017). Reluctantly adopting the broad definition of Behavioural Insights used by OECD, and similar to the one adopted here in this handbook, Osman et al finds that only 93 of cases reported included a form of experimental manipulation in the study that would constitute an actual experiment (i.e. not a survey, desk top literature review, workshop, interview, economic modelling). Further, only 65% of these included a control or baseline condition from which to compare the interventions that were implemented. That is, nearly 2 out of 3 cases submitted to the OECD by practitioners as BI work failed to qualify as such even within a broad definition of this. It may thus be concluded that while current BI activities aim to build on behavioural science, reality seems to be a bit less rosy. In general, many of the activities that describe themselves as rooted in the behavioural sciences often bare a very weak link to these and is, at best, described as 'inspired' by them.

The problem of this watering down of BI is that it threatens to undermine what the behavioural sciences brings to the table of public policy: realistic models of human judgment and decision making; the scientific rigour of the replicable causal experimental approach carried out in consistency with these theories; and the knowledge about what psychological mechanisms shapes behaviours and under what conditions, allowing for evidence-based public policy design, delivery and evaluation as well as the transfer and scalability of such policies. Of course, in science the watering down of a paradigm would usually not be an issue. There, progression of a paradigm leads to its crystallisation with clearer definitions, theories, methodologies, and standards of reporting as a result. Yet, the individual, social and organisational motives for working with BI combined with its many attractive features – novel, innovative, non-standard, embracing the fallibility of humans, purportedly 'cheap and easy to implement' and effective – invites for a fate similar to that of the commercialisation of the cupcake. Everyone wants to get onboard the boat to be part of the shebang or make a profit, but potentially with little wish for acquiring new skills ultimately leading to an interest in widening and watering down the concept until it loses its bite.

What happens to science when it is popularised? The popularisation of behavioural science through pop-sci books and easily digestible public reports that suggest lists of 'copy-paste' behavioural tweaks to influence people's behaviour, has played a central part in the world wide spread of BI. But while such science popularization fulfils the important task of making scientific knowledge accessible for policy-makers, practitioners and the lay public in general, the simplification of information required to achieve this accessibility also has its downsides.

In general, studies have shown that people reading popularised articles agrees more with the knowledge claims contained and underrate their dependence on experts than when presented for academic articles (Scharrer, Rupieper, Stadtler, & Bromme, 2017). Also, popularisation may make the knowledge and activities seem simpler than they are and thus easier for the lay person to master.

Queueing for cupcakes. At the turn of the millennium the tv-series Sex and The City popularised the cupcake as a fashionable cake. Soon people were cueing for cupcakes at the store in Bleeker Street, New York, which featured in the series. A worldwide cupcake craze then broke out, with trendy cupcake stores popping up in all cosmopolis. However, as everyone jumped the boat, the quality of cupcakes went to an unappetizing low level, leaving the social craze to meet its inevitable end with an undeservedly bad legacy.

As was mentioned above, a central driver in the fast spread of BI around the world has been a series of frameworks, reports and paradigm results of which the perhaps most prominent as well as illustrative of these is the MINDSPACE report (Dolan, Hallsworth, Halpern, King, & Vlaev, 2010). Like other subsequent successful BI frameworks such as *EAST: Four Simple Ways to Apply Behavioural Insights* (2014), MINDSPACE is an acronym covering a series of behavioural insights presented after a brief exposition of the state of the art in the behavioural sciences and followed by illustrations of their use in policy.

To be specific, MINDSPACE is an acronym referencing the *Messenger*-effect; the framing of *Incentives*; the influence on behaviour of *Norms*, *Defaults*, and making choices or aspects thereof *Salient*; the possibility of *Priming* peoples' actions, or using *Affect*, *Commitments* or appeals to *Ego* to do the same. Above we saw how one of the UKBITs most famous interventions tested the elements from MINDSPACE in the context of tax collection. In the most successful condition information about social norms was incorporated into the letter by writing:" Nine out of ten people in the UK has paid their tax on time. You belong to the minority group that has not paid yet." This use of localised social norms resulted in a 5,1% increase in the number of tax payers paying their tax on time compared to the standard letter, not only adding evidence to the existing literature on the effectiveness of social norms messages in raising tax compliance, cf. (Wenzel, 2004), but also making for a paradigm example of the value and potential of applying BI to public policy.

However, when the Irish tax authorities subsequently took a similar social norms approach to tax collection in a series of five RCT field experiments, they ultimately had to conclude that:

> Overall, Revenue research suggests that social norms do not significantly influence taxpayer behaviour in Ireland… the results for social norms on their own are not strong or consistent across different trials, as expected based on the literature… The social norms with geographic specificity trials comprehensively demonstrate that localised social norm letters are not effective at improving filing compliance… Similar trials in the UK have shown that localised social norms significantly increase compliance, which may suggest that Irish and UK taxpayers respond differently to social norms. (Kennedy, O'Carroll, Shirran, & Walsh, 2017)

What this indicates is that despite the easy appearance to the contrary, the effective application of behavioural insights to public policy is not merely a matter of copy-pasting behavioural strategies that have worked in other places. In general, existing frameworks tend to downplay this important point by portraying behavioural insights much as tools of persuasion (see Box p. 29). That is, they say very little, if anything about the importance of understanding the conditions under which particular behavioural insights are relevant, what moderators and side effects to expect or how to go about matching types of interventions to behavioural problems. Only careful reading reveals that the most prominent frameworks such as MINDSPACE and EAST are intended by their authors as inspirational, rather than as actual suggestions for systematically and effectively applying behavioural insights in public policy. Yet, careful reading is not a widespread art in public policy or fast paced consultancy. Nor does it answer the question of how behavioural insights may be applied systematically and effectively rather than through a blind copy-paste-trial-and-error process.

---

9 Central Behavioural Insights according to MINDSPACE

Messenger: we are heavily influenced by who communicates information.

Incentives: our responses to incentives are shaped by predictable mental shortcuts such as strongly avoiding losses.

Norms: we are strongly influenced by what others do.

Defaults: we "go with the flow" of pre-set options.

Salience: our attention is drawn to what is novel and seems relevant to us.

Priming: our acts are often influenced by sub-conscious cues.

Affect: our emotional associations can powerfully shape our actions.

Commitments: we seek to be consistent with our public promises, and reciprocate acts.

Ego: we act in ways that make us feel better about ourselves.

What criteria has led to exactly these 9 behavioural insights being chosen to form MINDSPACE as well as when and how to apply them in public policy, is not discussed in the report.

The final challenge to be mentioned here is the current lack of substantive ethical guidelines for the application of BI to public policy. The need to address ethical issues is evident from a recurrent political and normative criticism that holds BI, and especially nudging, to encourage abuse of power by policy makers and professionals (Farrell & Shalizi, 2011) because it works by "manipulating people's choices" through informal interventions exploiting mechanisms that "work best in the dark" the effects of which are likely to disappear if nudges become transparent (Bovens, 2008) and which impairs our autonomy and ability to make moral choices for ourselves (Furedi, 2011).

While the Academic literature has largely put such concerns to rest as exaggerated or just plainly mistaken – see e.g. (Hansen & Jespersen, 2013) and (Sunstein, 2016) – it is undeniably true that BI pursue behavioural regulation and policies in ways that are non-standard and for which the ethical and legal implications are largely unexplored (Sibony & Alemanno, 2015).

Thus, it is also alarming when the OECD survey *Lessons around the world* (OECD, 2017) reports that ethical issues were not highlighted as a concern for the respondents who were mostly policy professionals new to applying BI to public policy. Whereas the OECD report on the survey downplays this disturbing fact by interpreting it as a sign that professionals and practitioners probably had it covered by applying standard organisational frameworks of ethics, it may just as well be an indicator that ethical issues are typically not considered sufficiently, or at all. This point is also reflected by the OECD itself in the *OECD Regulatory Policy Outlook 2018,* (2018) which says:

> The ethical application of behavioural insights remains an important topic amongst the behavioural community of government officials, academic practitioners, and private sector advisors. The use of academic partners who work in institutions with established codes of ethics and the use of existing ethic codes within the public sector (which are not necessarily adapted to experimental approaches) have so far helped address possible ethical concerns. Nonetheless, the behavioural community has identified the need to establish a code of ethics for behavioural practitioners that promotes the responsible application of behavioural tools and adheres those working in the field to certain standards when designing and running experiments in a public sector context, and reporting on experiments by governments.

This reposition, though, is not a coincidental one. It was the direct result of a meeting on 'Behavioural Insights in Public Policy' held in Paris in 2017. Here OECD hosted more than 150 behavioural practitioners from government, academia, private sector and international organisations (OECD, 2017). As the summary of this meeting states "Attendees agreed that behavioural practitioners need to understand and adhere to certain standards when identifying behavioural issues and possible applications, as well as when making recommendations or decisions, and to ensure that conclusions derived from experiments are clearly presented and based on a variety of behavioural approaches" (OECD, 2018).

The only problem is, of course, that what those standards should be is a matter of discussion, since, as was just mentioned above, BI pursue behaviour change in ways that are non-standard and for which ethical and legal implications are largely unknown (Sibony & Alemanno, 2015). The best illustration of this problem in relation to ethics is perhaps to be found in the debate around *nudging organ donation* (see side notes).

Nudging organ donation: While the legal and ethical implications of conducting experiments as part of public policy development is fairly well studied, the ethical implications of applying behavioural insights as part of public policy is not.

Perhaps the best illustration of this is found relative to organ donation. A famous paper published by behavioural scientists Eric Johnson and Daniel Goldstein in 2003 in the renowned academic journal Science documented, that in countries where citizens have to actively choose to register as potential donors, that is in countries with opt-in systems for registering as organ donors, fewer citizens are registered than in countries with opt-out systems (Johnson & Goldstein, 2003).

This has led many policy-makers and professionals to suggest that the default for registering as a potential organ donor should be changed from opt-in to opt-out systems holding that people who don't want to become donors can just opt out. While this suggestion has proved popular amongst many laymen, one might pause to consider why most policy-makers, professionals and citizens, find it unethical for companies to auto-subscribe consumers into purchasing goods, sharing data and receiving newsletters.

Even more considerate is that some of the most prominent behavioural scientists in the BI community staunchly oppose the use of opt-out systems relative to organ donation for ethical reasons, se e.g. (Thaler & Hansen, 2012). Their take on it is that an opt-out system works not only by making it easier for people to sign up, but also by making it more difficult sign off, plus sends misleading signals from government officials to citizens about what is correct behaviour in a deeply personal matter. Their take on it, is that prompting citizens, e.g. when acquiring one's driver's license, is a far more ethical approach, plus that opt-out systems may backfire.

The BASIC approach to applying Behavioural Insights

In order to meet these four challenges to current behaviourally informed public policy, this book provides a lengthy and technical introduction aimed at helping aspiring practitioners to apply behavioural insights effectively and responsibly. In particular, the book seeks to help practitioners:

- Grasp the basic theories and concepts adopted from the behavioural sciences,
- Learn the processes and tools involved when integrating behavioural insights in public policy,
- Understand the scientific methodologies that is applied in validating and testing behavioural public policies, and
- Establish ethical guidelines on how to work responsibly within this paradigm.

It does so by presenting 'BASIC: A Framework for Applying Behavioural Insights to Create Real World Behaviour Change' developed by the present author and originally published in a shorter form by the OECD. BASIC is a framework consisting of five stages that a BI project usually progresses through:

- Behaviour
- Analysis
- Strategies
- Intervention
- Change

Each stage consists of a series of tools and guidelines that helps practitioners in systematically formulating, developing, designing and delivering behaviourally informed policies beyond proof-of-concept, and in the most efficient and responsible way, without watering down the scientific underpinning of this approach.