CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Nudge plus: incorporating reflection into behavioral public policy

Sanchayan Banerjee [ID] and Peter John* [ID]

Department of Geography and Environment, London School of Economics, London, UK and Department of Political Economy, King's College London, London, UK
*Correspondence to: Department of Political Economy, King's College London, Bush House, 40 Aldwych, London WC2B 4PH, UK.
E-mail: peter.john@kcl.ac.uk

Abstract
Nudge plus is a modification of the toolkit of behavioral public policy. It incorporates an element of reflection – the plus – into the delivery of a nudge, either blended in or made proximate. Nudge plus builds on recent work combining heuristics and deliberation. It may be used to design prosocial interventions that help preserve the autonomy of the agent. The argument turns on seminal work on dual systems, which presents a subtler relationship between fast and slow thinking than commonly assumed in the classic literature in behavioral public policy. We review classic and recent work on dual processes to show that a hybrid is more plausible than the default-interventionist or parallel-competitive framework. We define nudge plus, set out what reflection could entail, provide examples, outline causal mechanisms, and draw testable implications.

Keywords: nudge; nudge plus; think; dual-process theory

A nudge that incorporates an element of reflection might at first seem to be a contradiction in terms. After all, the whole point of a nudge is that it happens automatically without much conscious thought on the part of the individual. The acknowledgment of low cognitive capacity to make fully rational choices is thought to show its superiority over other policy instruments, such as information campaigns, laws, and taxes (Thaler & Sunstein, 2009). Nudge is supposed to work on fast and automatic type 1 processes, leaving the slow and reflective type 2 unengaged. Thaler and Sunstein stress that the individual could reflect and agree with a nudge after its delivery rather than before or during (Thaler & Sunstein, 2009, p. 244), with most citizens approving of this approach to designing public policies (Sunstein, 2016b, pp. 140–41).

Nonetheless, recent work in behavioral public policy suggests that a nudge could become more effective and legitimate if it incorporated an element of self-awareness and internal deliberation, which could generate long-term, persistent, and sustainable

behavior change (see Mühlböck *et al.*, 2020). John and Stoker (2019) made a start by coining nudge plus, which adapts their earlier contrast between the classic nudge and a purely deliberative 'think'. Nudge plus refers to an intervention that has a reflective strategy embedded into the design of a nudge. It can be delivered either as a one-part device in which the nudge and the reflective plus are intrinsically combined or two-part whereby the nudge is extrinsically combined with a deliberative instrument that prompts individual reflection on the nudge. Examples include the dual self-pledge involving multiple commitment contracts catering to an individual's short- and long-term preferences, and a GPS device combined with AI technology assistants.

Some existing nudges already have an element of self-reflection, which could be enhanced in a program of nudge plus. A commitment device, for example, is based on the idea that a precommitment default keeps people to a desired course of behavior (Thaler & Shefrin, 1981); yet, it also ensures that the individual has some autonomy and space to think through what is involved, which precedes the signature of the contract (Stutzer *et al.*, 2011). Reflection also appears in Sunstein's advocacy of educative nudges and contrasts between system 1 and system 2 nudges (Sunstein, 2016a; Sunstein & Reisch, 2019), leading to work on transparent nudges (Hansen & Jespersen, 2013) and deliberation tools, such as cooling-off periods (Yeung, 2012). Other examples include encouraging job seekers to think more slowly (Heller *et al.*, 2017) and training in youth self-investment using Cognitive Behavioral Therapy (Blattman *et al.*, 2017).

Recent research shows that classic nudges work better by bundling them with reflective elements to make them more salient. For instance, Bradt's (2019) study of insurance demands against flooding finds that providing information about the probability of inundation over a 30-year span increases the effect of the nudge. Visintin *et al.* (2020) test the efficacy of a 'thought-provoking' nudge for the implementation of disability insurance policies in Switzerland. Nudges may be enhanced if citizens are convinced of their ethical dimension (Engelen *et al.*, 2018). The engagement of the conscious brain also appears in the recent debate about whether being aware of a nudge affects its efficacy (Loewenstein *et al.*, 2015; Bruns *et al.*, 2018). Finally, work on capacity-building devices, such as boosts (Hertwig, 2017; Hertwig & Grüne-Yanoff, 2017), suggests that the individual needs preparation to make an effective choice.

Building on this contemporary empirical research, nudge plus implements such strategies systematically. Nudge plus, as a modification of a classic nudge, must involve an active trigger of reflection as the plus, as the potential for reflection is not sufficient to prompt deliberation and cause lasting behavior change. A commitment device, for instance, can be upgraded to a nudge plus when it also provides information about the underlying aims of the process or when it has a way of feeding back to the individual. Nudge plus needs to rest on a coherent and defensible account of cognition and to fit with dual-process models as advanced by Stanovich and West (2000), taken up by Kahneman (2012) and Thaler and Sunstein (2009). Using recent research in psychology, it is possible to show how type 1 and type 2 processes can be in play at the same time. To achieve this aim of grounding the tool of nudge plus, this article synthesizes classic and recent literature in social cognitive psychology, showing convincing evidence in favor of the hybrid dual-process theory and providing credibility to nudge plus. The later part of the article concerns reflection as embodied by the plus and conveys practical examples. These insights set out the mechanisms involved in the design of nudge plus and generate testable propositions.

From nudge to nudge plus

Nudge is low-cost signal or procedure that encourages, from the planner's point of view, a socially desirable change in behavior while preserving individual liberty (Thaler & Sunstein, 2009). Although there is considerable debate about definition (see Baldwin, 2014; Oliver, 2017), nudge is best thought of as an instrument involving a change in choice architecture. Nudge takes advantage of the biases of type 1 processes by changing the external environment (choice architecture) to help a person get to the socially optimal outcome.

One common criticism is that nudge can only deal with relatively minor public problems strictly under the guidance of the benevolent policy-maker (Marteau *et al.*, 2011). The size of the challenge of achieving sustainable behavior change may require more profound and long-lasting solutions that build on the consent of individuals. There also needs to be a way to address the common criticism that nudge manipulates individuals, reducing their autonomy and bypassing their explicit consent (Bovens, 2009; Glod, 2015). Nudge is often thought to 'work better in the dark', undermining its legitimacy (Bovens, 2009). Thaler and Sunstein have justified nudge from type transparency in line with Rawls' publicity principle: a watchful agent may identify the underlying choice architectural change and opt out, making nudge 'in principle token transparent' (Sunstein, 2015; Lades & Delaney, 2019). Yet, this leaves the problem of a lack of autonomy at the point when nudge is being delivered.

These objections are addressed in an alternate program of 'think', which implies that debate and deliberation can help individuals achieve their objectives (John *et al.*, 2019). Think is always open to public scrutiny and respects freedom of choice. But it is hard to scale up to the general population. The individual has to spend considerable time they may not be willing to give, and it relies on a strong commitment. To be closer to nudge, John and Stoker (2019) propose nudge plus, which incorporates an element of reflection and autonomy, yet is cognitively easy to uptake. In combining the nudge and think, nudge plus promises to make nudge token transparent, such that all receivers, regardless of watchfulness, are conscious of nudge with its deliberative prompt, respecting the autonomy of the individual who can decide what is best, even if that does not entail the socially optimal choice. Nudge plus gets autonomy at minimal cost.

Let us consider reducing obesity. How would nudge plus work differently to nudge? Nudge works by tapping into people's biases by changing the choice architecture only; for instance, defaulting the obese into buying healthier meals yet not necessarily leading to the uptake of healthy eating behavior. A default can even backfire as individuals experience moral warm glow and compensate by binge-eating other meals. Alternatively, a think strategist recommends education, for instance, consultation with a dietician to draw up a detailed plan. In contrast, nudge plus is a hybrid nudge-think strategy that combines nudge, in this case a default, with an active mechanism device, like a pledge that enables reflection on future meal choices, including the possibility of binge eating. While the obese may choose to override the motive of the nudge, the pledge makes the nudge transparent in that individuals own the process of behavior change. While the claim that nudge plus is more efficient than nudge remains to be validated, nudge plus has the benefit of restoring consumer autonomy

and agency. Note that nudge plus is different to boost, which works by enabling citizens to use their heuristics smartly. For example, a quick rule, such as to eat frequent yet smaller meal portions or to combine junk food with healthier options (temptation bundling), does not necessarily involve reflection and autonomy as must happen with nudge plus.

Dual-process theories and nudge plus

Nudge plus is based on a different approach to cognition than nudge. It deploys a hybrid framework that incorporates both heuristic and reflective processes, a bifurcation often referred to as dual-process theories. But even within dual-process theories, there can be a subcategorization depending on the type of the response mechanism involved. Although there can be different interaction mechanisms of the dual processes, one may suit the role better. Furthermore, dual processes can act in conjunction. To address these issues, we review dual-process theories, in particular, recent evidence from social cognitive psychology and neuroscience that supports nudge plus as a hybrid nudge-think strategy.

Dual-process theories posit that 'there are two distinct processing models available for cognitive tasks: one (type 1) that is fast, automatic and non-conscious, and another (type 2) that is slow, controlled, and conscious' (Frankish, 2010, p. 914). They date back to the 1960s and have been evolving ever since. Although different schools of thought have emerged, cognitive processes have been clearly distinguished into an intuitive (or, heuristic) and analytical (or, systematic rule-based) type that might interact with one another and take precedence, depending on the nature of the task. Earlier labeled as system 1 and system 2, and extensively used following Stanovich (1999) and Stanovich and West (2000), they were popularized by Kahneman (2012). While these theories came to existence independently, an attempt to combine them into a more structured and generalized framework, based on the common traits of these models, was made much later (for a summary, see Gawronski & Creighton, 2013, chapter 14). Evans and Stanovich (2013) argue that these processes share multiple features, but they are all not defining. They put forward a necessary and sufficient condition for each of the dual processes: type 1 processes must have autonomy and type 2 processes must satisfy cognitive decoupling for hypothetical thinking, that is, being able to differentiate an assumption from a belief and back up a rational decision with a thought experiment. These cognitive processes interact with one another and resolve conflict in different ways: they might take precedence sequentially or operate in parallel depending on the nature of the task.

The former conflict resolution strategy is commonly referred to as the default-interventionist model. It posits that both brain processes can dominate one another; see A-dominating (Lindsay & Jacoby, 1994) or C-dominating (Jacoby, 1991) process dissociation models. However, classic default-interventionist models (Evans, 2010; Kahneman, 2012; Evans & Stanovich, 2013) assume a corrective role for the rational processes at all times, that is, the type 2 processes will override type 1 processes if there is a conflict. Contrarily, the latter conflict resolution model is referred to as the parallel-competitive model in which the dual processes fight constantly to gain dominance (for details, see Evans, 2007). Which of these are more effective and suit the role better remained an ontological concern until recently as 'all models enjoy[ed] implicit

support from dual process theorists' (Evans, 2007, p. 10). For instance, Epstein's cognitive-experiential self-theory (1994) presumes that the two types of processes (called systems incoherently) might occur in parallel, while Kahneman and Frederick (2002, p. 51) endorse a default-interventionist structure as they write, 'we assume, system 1 quickly proposes intuitive answers to judgment problems as they arise, and system 2 monitors the quality of these proposals, which it may endorse, correct or override.' There is no situation where any type of processes is passive.

There is more to cognitive processes than following a sequential (default-interventionist) or simultaneous (parallel-competitive) conflict resolution mechanism (Lurquin & Miyake, 2017; Pennycook, 2018). Nudge plus requires an integrated theory of cognition, one that is rooted in a more involved interaction of brain processes, essentially a hybrid approach. Recent evidence from neuroscience qualifies simple dual-process conflict resolution strategies by advocating the role of a third and superior type of brain processes, called executive functions, that monitors the heuristic and reflective processes (Varga & Hamburger, 2014; Grayot, 2020). Furthermore, these dual processes are not discrete. As Grayot writes, 'Although most researchers prefer to believe that system 1 and system 2 are arranged sequentially, there isn't sufficient empirical evidence to validate either the default-interventionist model or the parallel-competitive model of system interaction. Recent meta-analyses and replications indicate that neither model is singularly equipped to predict and explain how individuals' reason and make decisions' (Grayot, 2020, p. 115). These dualistic conflict resolution models might be flawed. As Camerer *et al.* argue, 'human behaviour requires a [much more] fluid interaction between controlled and automatic processes' (Camerer *et al.*, 2005, p. 11).

A hybrid framework suggests a more involved role of dual processes. (De Neys, 2012, 2014; De Neys & Pennycook, 2019). It overcomes the shortcomings of the cognitive theories. Gronchi and Giovanelli (2018, p. 2) argue that a 'shallow analytic monitoring process is always active to detect potential conflicts between the two systems, and an optional deeper processing stage is activated once an actual conflict between fast and slow thinking is found.' DeNeys suggests two types of system 1 responses: one that is heuristically driven (the so-called intuitive processes); the other that is logically intuitive. The two system 1 processes are activated in parallel followed by the system 2 processes, which act to validate and justify the conflict resolution put forward by the logically intuitive processes. DeNeys upgrades the role of the system 1 processes, and he acknowledges that the system 2 processes are more of the validator than the corrector as has been incorrectly assumed by the prior scholarly literature on dual-process accounts. Another account by Bohl and Bos (2012) suggests complementarity between two neural systems, each embodying a type of dual process.

Most available behavioral instruments endorse either the parallel-competitive or the default-interventionist conflict resolution strategies; for instance, a system 1 nudge is theorized to correct for failures of the automatic cognitive processes, while a system 2 nudge facilitates deliberation in the agents and corrects for conscious biases. Nudge is strictly defined to work following a default-interventionist conflict resolution strategy where either type of the cognitive processes assumes the role of the rectifier in sequence but never work together simultaneously. Contrarily, boosts

work by upgrading an individual's repertoire of decision-making skills, the adaptive toolbox. In so doing, the boost closely resonates with a unified theory of cognitive processes, one where there is no distinction between the fast and the slow mind. However, recent evidence from cognitive psychology and neuroscience suggests that different sections of the brain might be activated in response to a common stimulus; for instance, Karlan *et al.* (2019) explain how a charitable-giving nudge can work either through impulse or deliberation. The functionality of this nudge, however, depends on the context. As such, behavioral change interventions based only on the default-interventionist or a parallel-competitive account are always unable to explain fully the interplay of cognitive resources. This makes a hybrid nudge-think tool like nudge plus so compelling that it becomes a means by which to effectuate behavior change.

Nudge plus is flexible in that it shares features of both the parallel-competitive and default-interventionist dual-process accounts, depending on the context of application. A simple way to think about this would be using Kahneman's characters, slightly renamed, Bobbie (type 1) and Joey (type 2). Let us assume that Bobbie and Joey participate in a pub quiz as a team. Bobbie and Joey get a minute to answer questions in a round. Both are normatively rational in that they have their own thematic strengths, allowing them to selectively sort out rounds based on their expertise. If this is the case, either Bobbie or Joey reacts more quickly, leaving the other to validate the response. Nonetheless, there may arise situations when both need to consult each other and work out the answer. The former, the default-interventionist model of conflict resolution, supports a sequential nudge plus mechanism whereby the plus precedes or follows the nudge, letting automatic and reflective processes act in sequence, reinforcing each other; the latter, the parallel-competitive model of conflict resolution, suggests that a simultaneous nudge plus can be delivered at the same time prompting both types of brain processes to act together.

Operationalizing nudge plus

How can nudge plus be designed and administered? This depends on two factors: the timing of the delivery of the plus with the nudge and the combination strategy. The plus can be conceived by the policy-maker to be delivered before, after, or as part of the classic nudge, as either a one- or two-part device (see Table 1). The preferred order of the nudge and plus depends on the task, generating different treatment effects. Although nudge and plus can be separable, as in the two-part device, both elements are complementary in the functioning of nudge plus. The agent receiving the reflective plus switches from thinking fast to thinking slow in a way that helps responding to nudge. When stand-alone, plus reduces to simple think. While nudge involves any change in the choice architecture, it can prompt either an unconscious, reflexive action (a system 1 nudge) or a conscious, reflective action (system 2 nudge) but not both. A think involves a purely educative strategy that prompts deliberation. Nudge plus is a hybrid as it modifies the nudge by prompting both conscious and unconscious actions. It must have nudge as its fully functional and central unit, with the reflective device designed to enhance the reflection of the receipt of the nudge. While nudge plus promises greater autonomy and token transparency relative

Table 1. Some working examples of nudge plus.

		Timing of nudge plus	
		Simultaneous	Sequential
Type of nudge plus	One-part	GPS	Dual self-pledge device
	Two-part	A nudge (e.g., traffic lighting scheme) with an information signal	A nudge (e.g., default) either preceded or followed by a pledge

to nudge, each works differently. Some classes of plus may work by making the design and construct of the existing nudge more salient to the receiver, while others might allow the agent to reflect deeply on their own preferences. However, the change in effectiveness might be an ambiguous signal to the policy-maker in these latter instances. This remains a normative judgment, for what is considered best for the agent by the policy-maker might not be true for the agent.

An illustration is Sunstein's example of the global positioning system (GPS) which 'tells you how you can best get to your preferred destination, but it does not impose any sanction or costs if you refuse to do what it says' (Sunstein, 2015, p. 208). While such devices increase navigability, there remains the potential of reflection, acquired through prior experience and embedded in the agent's reliance on the device, and people have, at times, complained of a lack of precision and accuracy. Such glitches in GPS devices mean that, when used, they not only nudge but also direct an agent's conscious deliberative efforts to the choice environment in a way that they have to undertake some active decisions; for instance, the active choice of not using a GPS when one is familiar with a road system hints that the user has consciously thought about its use, possibly learning from past failures, and has not just been tricked into using the device heuristically once again. The conscious brain can be engaged even when relying on the automatic system. When these reflective features are improved, nudge can expand into a nudge plus: as with the GPS, it encourages the user to follow the map heuristically, who can choose to override the advice, a decision that is taken consciously through the experience of using the device in the past.

Also consider a dual self-pledge device to commit to a certain goal as an example of a one-part nudge plus. It differs from an ordinary commitment device in that it accounts for the dual self-nature of individuals by providing them with two different pledges: a current-scenario pledge where they commit to a short-run goal and a future-scenario pledge in which they envisage themselves in the long run. Take *Hinge*, a popular online dating application founded by Justin McLeod. The creators envisage it as a 'long-term relationship' app that is 'designed to be deleted'. In marketing their application, the creators encourage singles to first fulfill their short-term goal of finding a suitable romantic match. Having successfully realized this goal, Hinge encourages these users to delete the application as they embark on a romantic relationship with their suitable matches, a long-term goal.

Most fitness trackers prompt the user about their activity level on a daily or weekly basis. This prompt can be enhanced to include a reflective element by adding the option of setting up future fitness goals; for instance, the application could not

only prompt the user about their current active hours given their short-term commitment, but also engage them to think of their future goals (a weight goal or a particular physique), inducing reflection through the feed-in loop and helping them appraise their goals. Additional prompts could be built in to engage the user: for instance, often after a period of inactivity, a fitness app can either prompt the user to start a new activity or pop-up questions that assess reasons for missing out the daily goals, allowing users to update their short-term efforts if they were to stick to their long-term goal. By building in these mechanisms, the dual self-pledge device corrects an individual's present bias and induces reflection: first, committing to the future will require deliberation over willingness and capacities to achieve target; and, second, at the onset of a new period, the comparison between committed versus realized targets can feed back into deciding the next pledge, thus reinforcing this chain of reflection.

The application of the dual self-pledge device can be conveniently delivered in a variety of online contexts. The trigger of the reflective plus can be activated whenever a person views their monthly e-statement, for example. The arrangement can be extended to other user-friendly service domains as well; most network providers, for instance, give customers the option to control their spending beyond their chosen plan by specifying a threshold limit such that any spending beyond the limit automatically terminates additional services, unless reinstated voluntarily. This lock-in contract may be thought as the parallel to the current-self pledge in the previous examples. Now, add on to this a trigger that accounts for the temporality of usage, where the customer is also prompted to set up a future threshold consumption limit that they want to achieve over a period of time to reduce their bills. Similar to the current limit, the future self-threshold allowance can be revisited by the user and adjusted over time. When the customer logs in to view their billing statement, they are prompted with their performance history, including, but not limited to, suggested tips to stick with their goals. Another extension is to limit one's digital screen time usage.

A combination of tools, such as a two-part device, may also be seen as a nudge plus. A nudge, for instance, when combined and delivered with active triggers that increase the salience of utilizing information, can be classified to be a simultaneous nudge plus. These information signals can relate either to the construct of the nudge or to the choice environment in which citizens are functioning. Consider the traffic-lighting scheme as an example of a nudge. While the construct of a labeling scheme like the traffic-lighting nudge rests on the tenet that the agent is subtly reminded of red lights meaning stop, green lights meaning go, and amber ones meaning at one's own risk, such that they make the healthier and safer lifestyle choices automatically, agents, with strong antecedent preferences or working in defiance of ecological rationality,[1] might miss the visual cue, thereby rendering the nudge ineffective. In such cases, adding an information trigger that explicitly explains what the color coding means, which is the plus, would initiate an agent's reflection. Interestingly, however, increasing efficacy is not the only overarching objective of the nudge plus; for example, the given design can also be extended to other nudges

---

[1]The theory of ecological rationality specifies norms of rational action as an adaptive interaction of human cognition and the external environment in which humans work.

to increase transparency; for example, a default can be administered with an information signal that increases the salience of opting-out to the receivers. Jachimowicz *et al.* (2019) show that the efficacy of defaults hinges mainly on endorsement or endowment effects, potential threats to consumer agency. Similarly, the heterogeneity in the uptake of a labeling scheme varies with the covertness of the nudge, the effective ones being the ones that work heuristically (Galizzi, 2012), compromising consumer sovereignty. Whether the addition of a reflective plus enhances the effectiveness of such a nudge remains to be empirically validated, but it definitely makes such a covert scheme epistemically transparent and restores the individual's agency and autonomy. It is important to note that the additional information provision does not fall within the remit of the classic nudge. As nudge taps into biases by making existing information more accessible, the information signal as the plus provides additional knowledge that induces deliberation, much like the short-term boost or educative nudge (Damgaard & Nielsen, 2018).

Similarly, providing a choice to commit before or after a nudge could have different implications. Consider the opt-out default nudge once again. A default setting is usually taken up by an agent due to the cognitive easing it comes with; for instance, choosing from a set menu is often easier for an individual who dreads a large menu with many options. Providing the agent with the choice to commit to a healthier diet, such as a pledge to Veganuary, before the default menu is presented makes the uptake of the default more salient to the agent. However, if the sequence of this nudge plus is reverted such that the set menu is presented first, and just before ordering the agent is asked to commit to a healthier diet, the choice thereafter would be governed by self-reflection, and in essence should be more transparent to the agent, even though the treatment effect of the nudge plus might be different relative to its nudge counterpart. Through these exemplars, summarized in Table 1, it is clear that nudge plus respects the ability of individuals to decide for themselves to granting autonomy; it also makes the design of the instrument transparent.

The role plus plays, and the outcomes it aims to achieve, depends on the kind of reflection it seeks to deliver. In the dual-process view, type 2 processes involve reflection. When this reflection is embedded into the nudge as the plus, it leads to an experiential learning environment. This means that when a nudge plus is taken up by an individual, they reflect on account of the plus and learn from such an experience, which ultimately leads to a behavior change. If this learning experience is conducive to the individual and aligns with his/her personal goals, it strengthens the effect of the classic nudge. Reflection, true to its origin as seeing 'one's reflection in the mirror', has often been used to relate to self-reflection. An intellectual, for instance, is a mind that watches itself (cf., Camus, 1963). However, contrary to the common belief, reflection could personify different meanings (Bortolotti, 2011; Van Seggelen-Damen *et al.*, 2017); for instance, one could reflect on choices, beliefs, thoughts, or feelings, or on the available alternatives, or even on the structural assumptions behind a construct. Reflection involves thinking about something.

However, is the experience of reflection always conscious? Reflection could be consciously experienced, for instance, when a person deliberately thinks of options and successfully rerepresents it upon introspection, or unconsciously experienced in that that the experience of reflection cannot be introspected and rerepresented (see

Kastrup, 2017). However, irrespective of whether this experience of reflection is conscious or unconscious, depending on the awareness of the individual, it leads to the same experiential learning environment. Even more, it could be a solitary act when one reflects on one's own choices, or as part of the herd; for instance, one could introspect and evaluate their preferences either individually or through the process of a group discussion. Often, the most familiar area in which reflection occurs in groups is when debriefing takes place. As Boud writes, 'debriefing occurs when participants in a learning activity, be it a simulation, workshop experience or other event, are led through a session in which they relive parts of their experience in a supportive environment and draw conclusions from it' (Boud, 1985, p. 15). However, whether self- or group-led, conscious or unconscious, reflection involves a transformation of perspectives.

What does perspective transformation entail? Scholars recognize different stages in reflection that lead to the generation of new perspectives; Schon (1991), for instance, identifies three different stages in reflection: think, criticize, and act accordingly. Atkins and Murphy (1993) summarize this to involve an initial discomfort/dissonance from a certain stimulus, followed by critical analysis of one's feelings and thereafter acting in accordance with them. This critical reflection could either come as a knee-jerk reaction, for instance, a sudden heart disorder can lead to reflexive changes in lifestyle, or it can come in transition, for instance, borderline changes in blood sugar levels can make someone conscious of their lifestyle habits and encourage small lifestyle changes. Yet, certain conditions must be fulfilled, these being typified as necessary[2] and sufficient,[3] to engage in the process of perspective transformation. This distinguishes reflection from just letting thoughts emerge, which is clarified in the following definition:

> Reflection is the act of thinking and re-evaluating prior actions, choice constructs, or available alternatives and search rules, triggered by conscious or unconscious experiences, resulting in the uptake of new perspectives.

The plus, when taken up by the agent, implies self-reflection in relation to long-term preferences, for instance, when asked to commit after a default, or to deliberate on the design of the tool at hand that prompts behavioral change, or with the dual self-pledge cards. Whatever might be the reflective process, the agent learns in the environment and behaves accordingly with a new reactance. The environment is transparency enhancing, autonomy inducing, and/or effectiveness enhancing. Nudge plus has an effect through both the nudge and its plus; but to reinforce a persistent change, it is important to decide whether the plus comes before, after, or with the nudge depending on the context in hand. The application of the nudge plus, however, is not just limited to policy-makers.

---

[2]Being conscientious and goal-oriented helps to critically analyze feelings in response to experiences; it is necessary to initiate the process of perspective transformation; for instance, an agent facing an initial trigger, but lacking conscientious and goal-orientation, will fail to habitually engage in critically evaluating their feelings.

[3]Being motivated is sufficient to engage with the reflective plus. Motivation is key to transforming intentions into behavior, leading to perspective transformation.

While most nudge proponents justify nudging on the presumption that agents are 'cognitive cripples' who suffer from myopia and a lack of self-regulation (Edwards, 1983, p. 508), the essence of nudge plus lies in overriding this assumption by allowing individuals to reflect and do what is best as judged for by themselves. As such, the nudge plus can be administered either by an external agency, like a policy-maker as with a regulator or third-party provider, or by the individual. In the more classic nudge dialogue, the former approach involves the plus being delivered by the policy-maker who also delivers the nudge for the agent; for instance, energy/water regulators can engage in providing dual self-pledge devices to residential consumers. It can also be delivered as a two-part device by a regulator and a third-party provider, unrelated to one another; for instance, an external pledge like Veganuary, delivered in the UK by a registered charity, could significantly increase reflection for an agent who volunteers for the pledge. When visiting a store that has food items labeled as mandated by the regulator or a restaurant that provides with a default menu, the agent is able to reflect and update priors through perspective transformation initiated by the pledge made earlier. Nudge plus can also be self-administered as the agent devises strategies to self-nudge as a means to strengthen self-control. Individual agents can function as 'citizen choice architects' by 'not only learning the trick [but] also some insight into the psychological mechanisms behind it' (Reijula & Hertwig, 2019, p. 24).

To sum up, nudge plus embeds reflective strategies into the classic nudge, either as a single device that prompts reflection besides capitalizing on one's heuristics or as a two-part device that involves a combination of tools. Nudges can be upgraded to a nudge plus by embedding an active trigger of reflection in them. Most already include a passive reflective component, even the default, as the dialogue between the policy-maker and the citizen is not a one-off, but occurs over time, perhaps over the whole lifespan of a citizen, such as a pension default that is introduced at the start of someone's career but is also explained at a later stage.

The mechanistic scheme

Having set out an account of cognition that supports the effective use of nudge plus and provided some examples, the next step is to outline the mechanistic design. Let us consider a search strategy. To find an optimal strategy, individuals rely on their characteristic set that is broadly composed of agent-specific traits and some ancillary conditions. In a given setting, these agent-specific traits include dispositional factors that affect an individual's behavior; for example, these refer to one's grit and commitment to goals or their inherent preferences of risk, time, and information, while the ancillary conditions can be best thought of as 'a feature of the choice environment that may affect behaviour but is not taken as relevant to a social planner's evaluation' (Bernheim & Rangel, 2004, p. 55). The ancillary conditions, in turn, are composed of contextual information cues and the typical choice construct (involving situational factors) in which the agent is functioning. Drawing on this characteristic set, the agent devises simple search rules to narrow down the alternatives of the choice set along with their properties, which leads to a final choice.

Given this search, the behavioral instruments work by engaging with different elements of the characteristic set of the individual. Nudge, for instance, operates by co-opting the biases of an individual and changing the choice construct only, such that all other

attributes, including the set of alternatives and their properties, remain unchanged; for example, when consumers choose a restaurant, they might see a menu with labeling that influences their meal choice. Or they can be served a smaller portion size. Contrarily, regulatory policies involve changing the set of alternatives or its properties; for instance, a ban reduces the set of alternatives, or a price alteration changes the properties set. Nudge plus could work differently based on what role the plus plays; for instance, the plus can induce reflection on one's dispositional factors as with a dual self-pledge device, or it can use contextual information in order to issue alternative commands to help someone get to their destination, as explained in our example of combining the GPS with AI technology assistants to signal better navigability.

Consider the traffic lighting scheme which is combined with a pledge: while the nudge facilitates a change in the situational factors and hence facilitates the uptake of a specific dish by a visual cue, a pledge on the menu to eat sustainably redirects consumers' attention to reflect strongly on their traits (e.g., one possibility is by making the agent a sophisticated hyperbolic time discounter). If these consumers have other goals in place, for instance, a fitness regime they have signed up for, the reflection on the dishes can entail other parallel considerations as well. It might well happen that consumers decide to resist and blocks such a change (e.g., take a cheat day just because she has been asked to reflect) which is still autonomy enhancing and open, but less effective compared with the classic nudge of traffic lights alone.

Nonetheless, 'no behaviour sits in vacuum', since at any given time an individual is involved in multiple tasks (Dolan & Galizzi, 2015, p. 1). Engaging in a process of perspective transformation through the reflective plus and adopting a given sustainable behavior can also spill over to other behavioral domains as there could be different mechanisms through which such a spillover can manifest itself; for instance, upon perspective transformation, reflection could lead to an increase in the salience of one's personal goals. This, in turn, could encourage citizens to be consistent with their new goal orientation: they generate new dispositions strengthening a broad domain of related behaviors, for example, proenvironmental behavior (Bem, 1972), induce greater compliance in future, or generate identity effects (Lacasse, 2016). In a recent field experiment, Lacasse (2017) demonstrates that participants, who were randomly assigned to adopt a new proenvironmental behavior for three weeks, experience a sense of heightened environmental responsibility, culminating in the purchase of organic produce and support for proenvironmental policies. However, since the experience of reflection through the plus can be unconscious, the resulting behavioral spillover can also happen unconsciously. For instance, Nash *et al.* (2019) use qualitative semistructured interviews to show that around half their sample respondents, on self-reflection, unconsciously engaged in behavioral spillovers. This setup creates the following testable implications:

H1: Nudge plus empowers agents and increases their autonomy compared with classic nudges only.

H2: Nudge plus leads to a persistent behavioral change compared with the classic nudge conditional on the fulfilment of the necessary and sufficient conditions for one to engage in the process of perspective transformation.

H3: The change in an individual's direct behavior induced by the nudge plus causes spillovers to other prosocial behaviors.

Hypothesis 1 is justified by the arguments provided earlier. Nudge plus respects the cognitive abilities of agents and provides token transparency and so a greater autonomy for the agent. Hypothesis 2, the availability of the reflection, leads to a transformation of perspectives that then induces a persistent one-off change, unlike a classic nudge, whose effect may wear off when withdrawn if a habit is not created. This hinges on the fulfillment of the necessary (conscientiousness and goal orientation) and sufficient (motivation) conditions. Nudge does not translate into permanent behavioral changes because agents are simply responding to the modified external choice environment and are not updating their beliefs. Hypothesis 3 depends on the success of the plus in altering the direct behavior under consideration. As someone learns to reflect on dietary choices for environmental considerations, they simultaneously adopt or alter other related proenvironmental behaviors. The transformation of perspectives can have rippling effects on related behaviors.

Conclusion

We have outlined a modification of the toolkit of behavioral public policy called nudge plus. It is based on John and Stoker's (2019) idea that encouraging an element of reflection as part the delivery of nudge enhances outcomes because it gives an opportunity for citizens to own the process and thereby to commit and invest in it. As well as offering greater efficacy, nudge plus is desirable on its own terms by offering greater autonomy.

We have set out the conceptual foundations for nudge plus. By reviewing theories of dual processes, we show that the pure dependence on dual processes implied by a classic nudge is not sustainable, at least not in all domains. As a result, we claim that nudge plus is based on a plausible account of cognition. We then elaborated the potential for nudge plus in behavioral public policy, giving examples and ideas for researchers and practitioners to test and adopt. Although there have been considerable advances in the scope and range of behavioral interventions in recent years (see Benartzi *et al.*, 2017), we have identified a new range of interventions where nudge could be enhanced with the addition of the plus.

There remains much work to do, such as to examine the welfare implications of increasing autonomy, and to check whether nudge plus subtly manipulates individuals as the sponsor may be designing reflection to lead to an understated but preferred choice. Alternatively, nudge plus might be a sustainable route to other kinds of participation, which could address more general collective action problems, such as the prevention of climate change, by providing a link between citizen action on public policy issues and bottom-up movements for social and political action. By encouraging reflection, people may become more aware of wider political issues. Even with these intellectual challenges, the way forward is more tests of nudge plus.

References

Atkins, S. and K. Murphy (1993), 'Reflection: A review of the literature', *Journal of Advanced Nursing*, **18** (8): 1188–92.

Baldwin, R. (2014), 'From regulation to behaviour change: Giving nudge the third degree', *The Modern Law Review*, **77**(6): 831–57.

Bem, D. J. (1972), 'Self-perception theory', *Advances in Experimental Social Psychology*, **6**: 1–62.

Benartzi, S., J. Beshears, K. L. Milkman, C. R. Sunstein, R. H. Thaler, M. Shankar, W. Tucker-Ray, W. J. Congdon and S. Galing (2017), 'Should governments invest more in nudging?', *Psychological Science*, **28**(8): 1041–55.

Bernheim, B. D. and A. Rangel (2004), 'Addiction and cue-triggered decision processes', *American Economic Review*, **94**(5): 1558–1590.

Blattman, C., J. C. Jamison and M. Sheridan (2017), 'Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in Liberia', *American Economic Review*, **107**(4): 1165–1206.

Bohl, V. and W. van den Bos (2012), 'Toward an integrative account of social cognition: Marrying theory of mind and interactionism to study the interplay of Type 1 and Type 2 processes', *Frontiers In Human Neuroscience*, **6**: 274.

Bortolotti, L. (2011), 'Does reflection lead to wise choices?', *Philosophical Explorations*, **14**(3): 297–313.

Boud, D. (1985), *Reflection: Turning experience into learning*, London: Kogan Page.

Bovens, L. (2009), 'The ethics of nudge', in M. J. Hansson and T. Grüne-Yanoff (eds), *Preference change: Approaches from philosophy, economics and psychology*, Berlin: Springer, Theory and Decision Library A, 207–20.

Bradt, J. (2019), 'Comparing the effects of behaviorally informed interventions on flood insurance demand: An experimental analysis of "boosts" and "nudges"', *Behavioural Public Policy*, 1–31. doi: 10.1017/bpp.2019.31.

Bruns, H., E. Kantorowicz-Reznichenko, K. Klement, M. L. Jonsson and B. Rahali (2018), 'Can nudges be transparent and yet effective?', *Journal of Economic Psychology*, **65**(April): 41–59.

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

Camus, A. (1963), *Carnets, 1935–1942* (1st edn). London: H. Hamilton.

Damgaard, M. T. and H. Nielsen (2018), 'Nudging in education', *Economics of Education Review*, **64**(C): 313–42.

De Neys, W. (2012), 'Bias and conflict: A case for logical intuitions', *Perspectives on Psychological Science*, **7** (1): 28–38.

De Neys, W. (2014), 'Conflict detection, dual processes, and logical intuitions: Some clarifications', *Thinking & Reasoning*, **20**(2): 169–187.

De Neys, W. and G. Pennycook (2019), 'Logic, fast and slow: Advances in dual-process theorizing', *Current Directions in Psychological Science*, **28**(5): 503–509.

Dolan, P. and M. M. Galizzi (2015), 'Like ripples on a pond: Behavioral spillovers and their implications for research and policy', *Journal of Economic Psychology*, **47**: 1–16.

Edwards, W. (1983), 'Human cognitive capabilities, representativeness, and ground rules for research', In P. Humphreys, O. Svenson, and A. Vári (eds), *Analysing and aiding decision processes*, Advances in psychology, Vol. **14**, Amsterdam, North-Holland, 507–13.

Engelen, B., A. Thomas, A. Archer and N. van de Ven (2018), 'Exemplars and nudges: Combining two strategies for moral education', *Journal of Moral Education*, **47**(3): 346–65.

Epstein, S. (1994), 'Integration of the cognitive and the psychodynamic unconscious', *American Psychologist*, **49**(8): 409–24.

Evans, J. (2007), 'On the resolution of conflict in dual process theories of reasoning', *Thinking & Reasoning*, **13**(4): 321–39.

Evans, J. S. B. T. (2010), *Thinking twice: Two minds in one brain*, New York: Oxford University Press.

Evans, J. and K. E. Stanovich (2013), 'Dual-process theories of higher cognition advancing the debate', *Perspectives on Psychological Science*, **8**(3): 223–41.

Frankish, K. (2010), 'Dual-process and dual-system theories of reasoning', *Philosophy Compass*, **5**(10): 914–26.

Galizzi, M. M. (2012), 'Label, nudge or tax? A review of health policies for risky behaviours', *Journal of Public Health Research*, **1**(1): 14–21.

Gawronski, B. and L. A. Creighton (2013), 'Dual process theories', in D. E. Carlston (ed.), *Oxford library of psychology. The Oxford handbook of social cognition*. USA: Oxford University Press, 282–312.

Glod, W. (2015), 'How nudges often fail to treat people according to their own preferences', *Social Theory and Practice*, **41**(4): 599–617.

Grayot, J. D. (2020), 'Dual process theories in behavioral economics and neuroeconomics: A critical review', *Review of Philosophy and Psychology*, **11**(1): 105–36.

Gronchi, G. and F. Giovannelli (2018), 'Dual process theory of thought and default mode network: A possible neural foundation of fast thinking', *Frontiers in Psychology*, **9**: 1237.

Hansen, P. G. and A. M. Jespersen (2013), 'Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy', *European Journal of Risk Regulation*, **4**(1): 3–28.

Heller, S. B., A. K. Shah, J. Guryan, J. Ludwig, S. Mullainathan and H. A. Pollack (2017), 'Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago', *The Quarterly Journal of Economics*, **132**(1): 1–54.

Hertwig, R. (2017), 'When to consider boosting: Some rules for policy-makers', *Behavioural Public Policy*, **1**(2): 143–61.

Hertwig, R. and T. Grüne-Yanoff (2017), 'Nudging and boosting: Steering or empowering good decisions', *Perspectives on Psychological Science*, **12**(6): 973–86.

Jachimowicz, J., S. Duncan, E. Weber and E. Johnson (2019), 'When and why defaults influence decisions: A meta-analysis of default effects', *Behavioural Public Policy*, **3**(2): 159–186.

Jacoby, L. L. (1991), 'A process dissociation framework: Separating automatic from intentional uses of memory', *Journal of Memory and Language*, **30**(5): 513–41.

John, P. and G. Stoker (2019), 'Rethinking the role of experts and expertise in behavioural public policy', *Policy & Politics*, **47**(2): 209–25.

John, P., S. Cotterill, A. Moseley, L. Richardson, G. Smith, G. Stoker and C. Wales (2019), *Nudge, nudge, think, think: Experimenting with ways to change citizen behaviour*, Manchester: Manchester University Press.

Kahneman, D. (2012), *Thinking, Fast and Slow*, London: Penguin.

Kahneman, D. and S. Frederick (2002), 'Representativeness revisited: Attribute substitution in intuitive judgment', in T. Gilovich, D. Griffin, and D. Kahneman (eds), *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York, NY, USA: Cambridge University Press, 49–81.

Karlan, D., T. Piyush and S. Welch (2019), 'Behavioral economics and donor nudges: Impulse or deliberation? (SSIR)'.

Kastrup, B. (2017), 'There is an "unconscious," but it may well be conscious', *Europe's Journal of Psychology*, **13**(3): 559–72.

Lacasse, K. (2016), 'Don't be satisfied, identify! Strengthening positive spillover by connecting pro-environmental behaviors to an "environmentalist" label', *Journal of Environmental Psychology*, **48**: 149–158.

Lacasse, K. (2017), 'Can't hurt, might help: Examining the spillover effects from purposefully adopting a new pro-environmental behavior', *Environment and Behavior*, **51**(3): 259–287.

Lades, L. K. and L. Delaney (2019), 'Nudge FORGOOD', *Behavioural Public Policy*, 1–20. doi: 10.1017/bpp.2019.53.

Lindsay, D. S. and L. L. Jacoby (1994), 'Stroop process dissociations: The relationship between facilitation and interference', *Journal of Experimental Psychology: Human Perception and Performance*, **20**(2): 219–34.

Loewenstein, G., C. Bryce, D. Hagmann and S. Rajpal (2015), 'Warning: You are about to be nudged', *Behavioral Science and Policy*, **1**(1): 35–42.

Lurquin, J. H. and A. Miyake (2017), 'Challenges to ego-depletion research go beyond the replication crisis: A need for tackling the conceptual crisis', *Frontiers in Psychology*, **8**: 568.

Marteau, T. M., D. Ogilvie, M. Roland, M. Suhrcke and M. P. Kelly (2011), 'Judging nudging: Can nudging improve population health?', *BMJ*, **342**: d228.

Mühlböck, M., F. Kalleitner, N. Steiber and B. Kittel (2020), 'Information, reflection, and successful job search: A nudging experiment'. Retrieved from SSRN: https://ssrn.com/abstract=3576740.

Nash, N., L. Whitmarsh, S. Capstick, J. Thøgersen, V. Gouveia, R. D. C. Rodrigues Araújo, M. K. Harder, X. Wang and Y. Liu (2019), 'Reflecting on behavioral spillover in context: How do behavioral motivations and awareness catalyze other environmentally responsible actions in Brazil, China, and Denmark?', *Frontiers in Psychology*, **10**: 788.

Oliver, A. (2017), *The origins of behavioural public policy*, Cambridge, UK: Cambridge University Press.

Pennycook, G. (2018), 'A Perspective on the theoretical foundation of dual-process models', in W. De Neys (ed.), *Current issues in thinking and reasoning. Dual process theory 2.0*. Oxon and New York: Routledge/ Taylor & Francis Group, 5–27.

Reijula, S. and R. Hertwig 2019), 'Self-nudging and the citizen choice architect'. Preprint SocArXiv.

Schon, D. A. (1991), *The reflective practitioner: How professionals think in action*. Aldershot: Ashgate.

Stanovich, K. E. (1999), *Who is rational?: Studies of individual differences in reasoning*. Milton: Psychology Press.

Stanovich, K. E. and R. F. West (2000), 'Individual differences in reasoning: Implications for the rationality debate?', *Behavioral and Brain Sciences*, **23**(5): 645–65.

Stutzer, A., L. Goette and M. Zehnder (2011), 'Active decisions and prosocial behaviour: A field experiment on blood donation', *The Economic Journal*, **121**(556): F476–93.

Sunstein, C. R. (2015), 'Nudges do not undermine human agency', *Journal of Consumer Policy*, **38**(3): 207–10.

Sunstein, C. R. (2016a), 'People prefer System 2 nudges (kind of)'. SSRN Scholarly Paper ID 2731868. Rochester, NY: Social Science Research Network. Retrieved from: https://papers.ssrn.com/abstract=2731868.

Sunstein, C. R. (2016b), *The ethics of influence: Government in the age of behavioral science*. Cambridge: Cambridge University Press.

Sunstein, C. R. and L. A. Resich (2019), 'Educative nudges and noneducative nudges', in C. Sunstein and L. Reich (eds), *Trusting nudges toward a bill of rights for nudging*. Abingdon, Oxon: Taylor & Francis Group, 95–118.

Thaler, R. H. and H. M. Shefrin (1981), 'An economic theory of self-control', *Journal of Political Economy*, **89**(2): 392–406.

Thaler, R. H. and C. R. Sunstein (2009), *Nudge: Improving decisions about health, wealth and happiness*, London: Penguin.

Van Seggelen-Damen, I. C. M., R. Van Hezewijk, A. S. Helsdingen and I. G. J. H. Wopereis (2017), 'Reflection: A socratic approach', *Theory & Psychology*, **27**(6): 793–814.

Varga, A. L. and K. Hamburger (2014), 'Beyond Type 1 vs. Type 2 processing: The tri-dimensional way', *Frontiers in Psychology*, **5**: 993.

Visintin, E. P., J.-M. Bonvin, F. Varone, F. Butera, M. Lovey and E. Rosenstein (2020), 'Can street-level bureaucrats be nudged to increase effectiveness in welfare policy?', *Policy and Politics*. doi: 10.1332/030557320X15955051687823.

Yeung, K. (2012), ''Nudge as fudge', *The Modern Law Review*, **75**(1): 122–48.