

# ATHENA

CRITICAL INQUIRIES IN LAW, PHILOSOPHY AND GLOBALIZATION


---

## When the Nudge Fails: The Limits of Behavioural Individualism and the Case for Meta-nudge

FLAVIO SCUDERI DI MICELI

*PhD Student in Human Rights, University of Palermo (Italy)*

✉ [flavio.scuderidimiceli@unipa.it](mailto:flavio.scuderidimiceli@unipa.it)

 <https://orcid.org/0009-0002-2883-0340>

### ABSTRACT

This paper explores the limits of behavioural public policy by critically addressing the debate between i-frame and s-frame approaches. It challenges the dichotomy that frames nudges as tools exclusively aimed at individual decision-making (i-frame), while structural reforms are seen as systemic interventions (s-frame). Drawing on Bobbio's functional taxonomy of legal measures and recent literature on meta-nudges and choice infrastructures, the paper argues that certain behavioural interventions — especially those targeting public officials or institutional processes — can produce systemic effects. Through the example of the “principle of trust” in the Italian Public Contracts Code, the analysis illustrates how regulatory nudges can transform administrative behaviour and reshape decision-making contexts. The aim is to develop a more integrated understanding of behavioural regulation, one that accounts for the interaction between interventions and the social, legal, and institutional environment in which they operate.

**Keywords:** nudge, normativity and behaviour, influence, behavioural public policy, meta-nudge

I wish to thank the anonymous reviewers for their insightful comments, which greatly improved the quality of the article. I am also grateful to Michele Ubertone and Giuseppe Rocchè for their careful editorial work and valuable suggestions.

ATHENA

**Volume 5.2/2025, pp. 71-106**

*Articles*

ISSN 2724-6299 (Online)

<https://doi.org/10.60923/issn.2724-6299/22516>



## **1. Bounded Rationality in Administrative Behaviour: An Introduction**

Traditional economic theories have long described rationality as an optimising choice process, in which individuals select the alternative that maximises their welfare, based on complete information and unlimited computational capacity. One of the first to challenge this paradigm, Herbert A. Simon (1955), introduced the concept of bounded rationality, according to which human decisions are influenced by cognitive limitations, partial information availability, and time pressures that prevent a comprehensive analysis of the available options. As a result, decision-makers adopt *satisficing* strategies, meaning they choose alternatives that are “good enough” rather than optimal. This theory applies particularly strongly to public organisations, which operate within complex regulatory and bureaucratic contexts where rationality cannot be understood as a process of absolute maximisation, but rather as a pragmatic management of the available informational and decisional resources (Simon 1997). Such structures develop routine procedures and standardised decision-making models to reduce uncertainty and ensure internal consistency within administrative processes. However, while these mechanisms may simplify choices, they can also generate rigidity and resistance to change.

In the field of public organisations, a recently established line of research is based on the assumption that the inefficiency or failure of public policies largely stems from the fallibility of human reasoning, and more specifically from behavioural inclinations. From this idea emerged the concept of choice architecture, that is, the context in which we all find ourselves when making decisions. Choice architecture carries the assumption that the context itself influences decisions, and that by altering this context, certain behaviours can be encouraged or discouraged. Since the nudge approach (Thaler and

Sunstein 2008) came to dominate academic debate, efforts have been focused on analysing and developing the components that make up the concept of a nudge, with choice architecture standing as its central pillar.

The public organisation thus becomes the primary decision-making context deserving attention. An administration is a structured system of rules, resources, and actors aimed at pursuing collective goals, within legal constraints and serving the public interest. Unlike private enterprises, which operate under economic efficiency and profit-maximisation criteria, public organisations must balance a plurality of goals – often conflicting – and manage a network of actors with heterogeneous interests. In this scenario, the limitations of human rationality are amplified: public decision-makers not only face fragmented and often contradictory information, but are also bound by regulatory, bureaucratic, and political constraints that affect their choices. As a result, administrations develop standardised procedures to simplify decision-making complexity, reducing uncertainty but also risking inefficiencies and institutional inertia (Crozier 1969). For this reason, if the potential of behavioural sciences applied to law is to be harnessed, public interventions cannot be limited to influencing the decisions of individual citizens, but must also address the internal functioning of institutions, creating mechanisms that foster more coherent and effective decisions. In this respect, one can observe a shift in the concept of rationality within organisations, marking an evolutionary stage in the theory of bounded rationality, where institutions not only operate under rational constraints but also strive to design systems that mitigate the impact of cognitive limitations<sup>1</sup>.

---

<sup>1</sup>To be more precise with terminology, one could refer, in Simonian terms, to “institutional rationality” (Simon 1957). Regarding the evolution of the concept of bounded rationality, consideration should be given to the theory of ecological rationality, which holds that rationality is not a matter of following a single universal rule but rather depends on the environment in which the decision is made. The nudge approach, in its foundational assumptions, aligns more closely with the “Biases and Heuristics school” (Kahneman 2012; Stanovich and West 2000). For the perspective of the “Fast and Frugal Heuristics school”, see (Gigerenzer 2000).

This article aims to examine how bounded rationality within administrative organisations influences decision-making processes, and how behavioural sciences can contribute to improving the quality of choices and public contexts. While the nudge approach initially focused primarily on individual decisions, there is now growing interest in strategies that alter rules and decision-making infrastructures, introducing the concept of the meta-nudge. The overarching objective is to demonstrate that the evolution of behavioural policies cannot overlook the cognitive and organisational limitations of the public decision-making process. Only through an integrated approach will it be possible to develop effective strategies to enhance institutional governance and the quality of administrative decisions.

## **2. The S-frame Critique**

The behavioural approach to public policy has frequently adopted an individual-centred model, according to which policy problems stem from cognitive limitations and behavioural biases of individuals and can be addressed through targeted interventions aimed at improving their decisions without altering the underlying rules (Chater and Loewenstein 2023). This paradigm developed from the notion that choice architecture can be modified through nudge strategies that influence human behaviour without imposing constraints, e.g. automatic defaults in pension schemes or nutritional labels to promote healthier eating choices (Thaler, Sunstein and Balz 2014). The effects of these interventions have often been limited and, in some cases, have even diverted attention away from more structural solutions, reinforcing the idea that individuals are solely responsible for their choices, and that political and economic failures can be resolved by acting upon individual preferences without changing systemic conditions.

In response to these considerations, a new perspective has emerged, known as the s-frame, which shifts the focus from improving individual choices to transforming the rules and institutions that shape the decision-making

context. This viewpoint begins with the assumption that many policy problems do not arise from individuals' cognitive errors, but from economic and regulatory structures that condition choices upstream (Connolly, Loewenstein and Chater 2024). From this perspective, truly effective interventions must aim to modify the structural conditions in which decisions are made, for example by regulating the nutritional quality of food products sold in supermarkets rather than merely providing information about their health effects (Chater and Loewenstein 2023,8). Chater and Loewenstein distinguish between types of interventions, defining:

- (1) i-frame (individual frame): an approach to public policy that focuses on individual behaviours and personal choices, aiming to modify them through tools such as nudges, incentives, or information. The central idea of the i-frame is that many social issues stem from cognitive errors or limitations in human rationality, and that these can be corrected without altering the underlying normative or institutional structures of the system.
- (2) s-frame (system frame): an approach that shifts the focus from individual choices to modifying the structures and rules that influence those choices. The s-frame is based on the belief that the primary causes of many social problems are not individual decision-making errors, but rather the institutional, regulatory, and economic contexts in which people make decisions. For this reason, s-frame solutions tend to intervene directly in policies of regulation, taxation, or structural planning to produce systemic change.

One of the main arguments in favour of the s-frame is that the most complex social problems cannot be solved through minor behavioural adjustments, as their causes are often rooted in political and administrative dynamics that require structural change. For instance, while an i-frame intervention to combat climate change might involve providing consumers with information about their carbon emissions or introducing incentives for the use of renewable energy, an s-frame approach would entail the adoption

of carbon taxation policies or stricter regulations on industrial emissions (Chater and Loewenstein 2024). This systemic approach has the advantage of addressing the root causes of the problem, rather than merely mitigating its surface-level manifestations.

Talking about “individual” and “systemic” frames may suggest that systemic effects are simply collective or large-scale effects. That is not what is meant here. “Systemic” does not refer to how many people are affected, but to where the intervention acts in the causal chain of decision-making. An intervention is systemic when it alters the rules, routines, expectations, or infrastructures that organise choices upstream, even if it targets only a limited group of institutional actors rather than the entire population.

The adoption of an s-frame approach may encounter significant political and economic resistance, as it entails deep changes to existing structures and may clash with the entrenched interests of corporations and powerful groups. Sunstein (2022; 2023) and Thaler (2023), on the other hand, have argued that the s-frame risks overlooking the role of individual action and the capacity of i-frame interventions to produce incremental changes that, over time, may contribute to broader transformations. There is also the issue of political feasibility: while nudges are often accepted because they do not impose significant costs on citizens or businesses, s-frame interventions – such as new regulations or taxes – can generate opposition and require greater political consensus to be implemented. The clash between the i-frame and the s-frame is not merely about which policy lever should be pulled first. Rather, it appears to concern the allocation of responsibility for policy failure. I-frame solutions are often promoted as structural substitutes for deeper regulatory reform: by framing social problems as matters of individual choice and self-control, they implicitly shift accountability from institutions and market structures to the individual citizen.

Another criticism that can be raised concerns the risk that the dichotomy between i-frame and s-frame may be overly rigid and simplistic. In many cases, the most effective interventions result from a combination of both

approaches, in which structural changes are complemented by behavioural interventions to support their acceptance and effectiveness. For instance, in the case of social security, an s-frame model based on mandatory pension contributions can be strengthened by i-frame interventions that help individuals better understand their saving needs and manage their financial resources more consciously. Ultimately, the distinction between i-frame and s-frame, while not uncontroversial, should not be entirely dismissed. While early interventions focused on individual corrections, the growing awareness of the limitations of this approach is prompting increased attention to the normative and institutional structures that shape collective choices. This does not mean that i-frames should be abandoned, but rather that they should be integrated into a broader framework, one that seeks not only to improve individual decisions, but also to transform the context in which such decisions are made. The ultimate aim should be to develop policies that not only correct individual behaviours but make virtuous choices the natural outcome of a well-designed system.

The s-frame critique is valuable in drawing greater attention to the effects of interventions on layered contexts, and in shifting perspective on how interventions targeting human behaviour are structured. Indeed, it appears that the general interest of behavioural scientists in individual choice architecture tends to focus predominantly on nudges that work – and work well – such as default rules (Johnson 2022). However, by focusing too heavily on individualised, tailored aspects centred around the person, the true efficacy of the nudge may fail to materialise (Starke and Willemsen 2024). From another angle, the effectiveness of i-frames has been questioned by empirical studies in two key areas: their large-scale impact and their ability to address complex problems. In a study by Della Vigna and Linos (2022), it was found that nudges implemented by governmental “Nudge Units” demonstrated modest average efficacy, with an increase in the take-up (i.e. adoption of the desired behaviour) of just 1.4%. This figure is significantly lower than the average effect of 8.7% observed in published academic studies.

Such a discrepancy suggests that the effectiveness seen in academic settings may not automatically translate when nudges are designed and implemented by government agencies. However, when it comes to complex issues, it has been observed that the actual impact of the i-frame may be modest, even when the intervention appears highly effective on the surface. A recent study on a default rule influencing choices toward renewable energy found that over 85% of Swiss households and 75% of Swiss businesses, despite the higher cost, preferred the green energy option (Liebe, Gewinner and Diekmann 2021). Nevertheless, the final impact of this intervention, despite the high proportion of individuals affected, proved to be minor. The energy system does not respond by instantly producing more green energy for the newly enrolled consumers, and furthermore, such an intervention could not be applied universally as there would not be enough green energy to supply. Thus, while i-frames may work, they often prove to be insufficiently incisive regarding the broader objective that the intervention aims to achieve. This triggers a reinforcement loop in the decision-maker, as the measurement of seemingly positive results may be perceived as effective—even if only modestly so. This belief contributes to a diminished focus on potential systemic interventions and, consequently, to the reallocation of fewer human and financial resources in that direction, thereby reinforcing the maintenance of the status quo (Andreas and Jabakhanji 2023).

### **3. Towards a Systemic Vision of Behavioural Intervention**

We can interpret the distinction between i-frame and s-frame as one that depends on the context to which a given measure is addressed. If we denote with I the intervention (any intentional action – normative, administrative, or behavioural – aimed at modifying the behaviour of one or more individuals) that alters the context, with B the behaviour of the subject (or group of subjects) intentionally targeted by the intervention, and with C the pre-existing context prior to the intervention (i.e., the set of causal factors



determining the subject's behaviour), we can express the relationship as follows:  $B = f(I, C)$ .

In this function,  $I$  represent all relevant aspects of the social control measure that directly or indirectly modify the context  $C$  and, consequently, the behaviour. The term  $I$  is to be read broadly. It includes not only the formal design of the intervention, but also the strategic intentions, implementation choices and operational constraints of the policymaker or choice architect who introduces it. For our purposes, the choice architect is not treated as an independent variable: what matters is not who designs the intervention, but how the intervention, as implemented, interacts with the context to produce – or fail to produce – behavioural change.

In our analysis, we focus on a single type of actor – the public decision-maker – whom we assume to be constant. The s-frame critique appears to have the insight that the context of a specific choice,  $C$ , is composed of two main components, distinct yet interconnected: an individual component  $C_i$ , representing the situational and immediate conditions that directly affect individuals' choices and behaviours with respect to that specific decision (including elements such as choice architecture, local incentives, and contingent environmental factors); and a systemic component  $C_{sy}$ , which includes the normative, economic, and institutional structures that define the broader framework within which the choice occurs. This latter component provides the rules, constraints, and long-term stability that influence collective behaviour and determine the available options.

It is important to emphasise that the context  $C$  is always relative to a specific choice or set of choices. Every causal factor that affects a decision can be classified as belonging to  $C_i$  or  $C_{sy}$ , depending on its scope and nature. Immediate situational factors that act upon the individual at the time of decision-making fall under  $C_i$  (e.g., the order in which options are presented, the scarcity of time in which to decide, or the colour or design of a button that encourages a particular choice). In contrast, structural and collective factors that regulate the broader framework and restrict or expand the decision-

making possibilities are part of  $C_{sy}$  (e.g., the absence or presence of electric vehicle charging infrastructure, or a streamlined procedure for authorising renewable energy plant construction). On an aggregate level, a single measure may influence multiple contexts –  $C_1, C_2, \dots, C_n$  – each related to a particular choice, thus generating effects across a set of similar decisions. Human beings naturally tend to conform to the status quo (Kahneman, Knetsch, and Thaler 1991). For this reason, we can say that even when  $C_i$  is influenced by an effective intervention  $I$ , it remains constrained by  $C_{sy}$ . Returning to the function:  $B = f(I, C_{sy}, C_i)$ .

Chater and Loewenstein argue that when an intervention focuses solely on the i-frame, the resulting behavioural change ( $B$ ) may generate a substitution effect that reduces the pressure for systemic reform. In such cases,  $C_{sy}$  remains unchanged in an i-frame-only analysis, since the system itself is not altered in any way. The result is that the impact of such interventions is often weakened, even if a small, visible improvement in behaviour occurs. According to Chater and Loewenstein, the fundamental problem with i-frames is that they can create the illusion of change while leaving the structural architecture intact. If a nudge merely alters individual choices without addressing all components of the decision-making context, the change will be marginal and potentially counterproductive, as it diverts attention away from deeper solutions. For example, if the government introduces nutritional labelling to combat obesity (an i-frame intervention), individuals may be encouraged to make healthier choices. However, this measure does not change the broader food system, which may continue to promote the sale of highly processed and unhealthy foods. If labelling is perceived as a sufficient solution, it may weaken support for more robust s-frame measures, such as a sugar tax or stricter regulations on junk food advertising.

In the analysis proposed by the two authors, the role of behavioural influence and the partial effectiveness of nudges is acknowledged. However, the possibility of influencing the system itself through nudges, as an

alternative to classic command and control interventions, is underestimated. It is certainly useful to adopt a different interpretative lens (Madva, Brownstein and Kelly 2023) when analysing intervention types, but it seems overly hasty to classify the nudge solely as a tool capable of influencing only one component of the context, without generating what they consider to be systemic change. Isn't a nudge that uses a default rule to improve pension plan choices an intervention that alters the pension system?

Adopting a functional model of intervention allows behaviour to be described as the outcome of the interaction between the intervention and the context in which it operates. In this view, the i-frame / s-frame distinction is not an ontological division between different policy instruments, but an indication of which component of the context an intervention primarily modifies: the immediate, situational choice environment ( $C_i$ ) or the broader institutional and normative structures that organise decision-making ( $C_{sy}$ ). The familiar assumption in the i-frame / s-frame debate that nudges “belong” to the i-frame, while command-and-control measures “belong” to the s-frame, is therefore misguided. What matters is not the label attached to the tool, but which part of the decision-making environment it durably alters. Downplaying the rigidity of the distinction – as this paper proposes to do – makes it possible to recognise that nudges, under certain conditions, may also have systemic effects, especially when they persistently modify decision-making structures, for example by introducing generalised defaults, standardising procedures, or stabilising new social norms. In this sense, an intervention can be both a nudge and systemic. We will refer to these as “meta-nudges”.

At the same time, structural interventions that disregard behavioural dynamics may prove ineffective or counterproductive. The key lies in developing an integrated vision that considers the interaction between interventions and context, and that can guide the design of policies capable of influencing the deeper dynamics of collective behaviour. To overcome this rigidity of categories, one can envisage an innovative tool that is capable of

acting on the system without resorting to incentives or direct coercion. This type of intervention could therefore address the s-frame critique by using behavioural influence to modify both components of the context identified above. A possible solution for coordinating the two perspectives will be addressed later. It is not the case that only command and control interventions influence  $C_{sy}$  and only nudges affect  $C_i$ . Both types of interventions concern both types of contexts. A taxonomy of social control interventions offered by N. Bobbio may be useful in illustrating this point.

#### **4. Bobbio's Taxonomy of Policy Interventions**

When discussing interventions by public administration, it is important to acknowledge a significant evolution in the role of law and the state. There has been a fundamental shift from a state primarily oriented towards repressing undesirable behaviour to one that, alongside this function, actively promotes socially desirable behaviour (Bobbio, 2007). This transformation is closely linked to the emergence and development of the welfare state, which, in response to new social and economic demands, no longer confines itself to protecting certain interests through the repression of deviant conduct but also aims to encourage innovative and economically beneficial behaviours. Within this context, the traditional technique of negative sanction, used to discourage undesirable conduct, has increasingly been complemented by the positive sanction, such as rewards and incentives, intended to actively promote desired behaviours. This shift deeply affects the functional conception of law, transforming it from a mere instrument of social control into a means of both control and social guidance.

According to Norberto Bobbio, instruments of social regulation can be divided into direct and indirect measures (*Ibidem*, 46). Direct measures attempt to obtain the desired behaviour – or prevent the undesired one – by acting directly upon the behaviour itself, such as using physical force by the police. Included in this category are measures of control and surveillance,

which are primarily negative in nature and aimed at preventing undesirable actions from occurring. Indirect measures, in contrast, aim to influence behaviour by acting on the motivations or conditions underlying the behaviour. Sanctions, as well as facilitations and hindering measures, fall within this category. Facilitation refers to the array of mechanisms through which an organised social group exercises control over its members by promoting behaviours in a desired direction, making their enactment easier or less difficult. Hindering is its direct opposite.

It should be noted that social control measures exist on a continuum, and it is often difficult to identify clear-cut boundaries between categories. To clarify their differences, Bobbio distinguishes three levels according to the causal intensity an intervention seeks to exert on behaviour. These levels can be ordered in terms of how forcefully the intervention acts upon the individual. At the highest level of intensity are (1) constriction or preclusion measures, which aim to bring about the desired behaviour or prevent the undesired one by making it unavoidable or impossible. These include direct measures, such as the use of public force to prevent an action. The objective is to render a particular behaviour necessary (in the case of positive direct measures) or impossible (in the case of negative ones). The second level is occupied by (2) retribution or reparation measures. These are applied after the behaviour has occurred and aim either to attach pleasant consequences to the desired behaviour (rewards), unpleasant consequences to the undesired behaviour (punishments or negative sanctions), or to restore the order disturbed by the behaviour (restorative measures or compensation). Only the latter are considered sanctions in the narrow sense. Finally, there are (3) facilitation and hindering measures, which aim to favour the adoption of desired conduct or discourage undesired conduct. Although they exert a lower level of influence, these measures occupy an intermediate position: like (1) they target the behaviour itself; and like (2), they are indirect in nature, relying on psychological rather than physical pressure.

Traditionally, public administrations have exercised authority primarily through the first two levels of social control, which we may simplify under the term command and control measures, as they necessarily alter costs or impose constraints. In contrast, the distinction between facilitation and hindering appears to mirror the one between nudge and sludge<sup>2</sup> (Sunstein 2021). Indeed, the nudge may be considered a particular form of facilitation, distinguished by its reliance on cognitive mechanisms. Thaler and Sunstein define a nudge as: “any aspect of choice architecture that alters behavior in a predictable way without forbidding options or significantly changing economic incentives” (Thaler and Sunstein 2021,6). More precisely, a nudge influences the behaviour of “Humans” – individuals who deviate from the assumptions of neoclassical rationality – even though it would be ignored by “Econs”, the rational agents in economic models. Both nudges and facilitation aim to encourage desirable behaviours without imposing obligations or prohibitions. They share a preventive nature, acting before or during a behaviour, rather than as a consequence of it. Their relatively low intensity ensures that freedom of choice is largely preserved: no alternatives are eliminated, but the desired choice is made easier or more attractive.

Facilitation appears to be a broader category than nudge, including interventions that do not necessarily rely on behavioural economics and are often more transparent in their objectives. While some facilitation measures simply aim to remove practical barriers or reduce administrative complexity, others may entail modifications to the economic incentives involved, such as reduced costs, expedited procedures, or material benefits. Nudges, by contrast, are defined precisely by the absence of such changes, they do not significantly alter the payoff structure or impose material constraints. This makes nudging a specific subset of facilitation, characterised not only by its

---

<sup>2</sup> The term sludge refers to excessive or unnecessary frictions in decision-making processes – such as paperwork, delays, confusing language, or complex procedures – that hinder people from achieving their goals or accessing services. Introduced by Sunstein (2021), sludge is the negative counterpart to nudges: while nudges aim to facilitate behaviour, sludge impedes it, often unintentionally or through bureaucratic inertia.

cognitive focus but also by the fact that it preserves the existing incentive structure. The distinction is therefore twofold: facilitation may or may not act on incentives, whereas nudges never do. Compared to facilitation, a nudge includes an additional behavioural condition (Congiu and Moscati 2022), it exploits cognitive limits, biases, routines, and habits in both individual and collective decision-making. It does so by embedding these factors into the design of the decision-making environment, what is known as choice architecture. Importantly, a nudge operates independently of prohibiting or adding rationally relevant options, of changing incentives (in terms of time, effort, social or economic sanctions, etc.) and of providing factual information or rational argumentation. What characterises this type of indirect measure is its intrinsic link to the intentional attempt to influence the judgement, the choice or the behaviour of people in a predictable way. In other words, the effect (i.e., the predictable change in behaviour) is a function of the intervention (the nudge), which is itself defined by this intentional attempt (Hansen 2016).

Interventions are always designed and implemented regarding the final effect they are meant to produce. Facilitation typically involves removing practical obstacles or simplifying procedures to make a desired behaviour more accessible, without directly intervening in the decision-making process. Examples include streamlining bureaucratic steps, offering support services, or improving access to information. In Bobbio's terms, these are forms of facilitation: they remove or reduce external barriers that make a desired course of action difficult, costly, or time-consuming. Nudging can be understood as a specific mode of facilitation. Rather than intervening on external obstacles, it works by leveraging predictable psychological mechanisms – for instance by setting defaults, structuring the presentation of options, or reframing outcomes – so as to guide choice without coercion. Both approaches aim to increase the likelihood of certain behaviours, but they act on different fronts. Classical facilitation modifies the external conditions of action, whereas nudging modifies the decision-making context as it is

internally perceived by the individual. For this reason, nudging should not be treated as an alternative to facilitation but as one of its behavioural sub-forms, an indirect measure of social control that operates psychologically rather than materially. The distinction is subtle (and perhaps mainly methodological) but it is analytically valuable. It allows us to treat these indirect techniques within a single family of interventions, while still distinguishing how behavioural change is produced through regulation and clarifying the intentions of those who seek to bring about such change. Bobbio's taxonomy is particularly useful because it shifts the focus from the formal nature of interventions to their functional role in shaping behaviour. Rather than distinguishing measures based on whether they target individuals or institutions, Bobbio classifies them by how they operate, directly or indirectly, and with what intensity. This approach reveals that interventions like facilitation can act not only on individual behaviour but also within institutional contexts, depending on how and where they are applied. In this light, the distinction between i-frame and s-frame does not align with a fixed boundary between "behavioural" and "systemic" tools. What matters is the causal structure of the intervention and its position within the decision-making chain. Bobbio's functional perspective thus helps to overcome the assumption that only traditional rules influence systemic contexts, while behavioural tools are limited to individual effects.

## 5. What Makes an Intervention Successful or Unsuccessful

Having established the relevant categories for command-and-control and nudge measures, it is essential to distinguish when these interventions are truly efficacious, namely when the public decision-maker's intention to change a behaviour aligns with the actual effect that is produced in reality. All three categories of intervention may be successful or not in influencing either  $C_i$  or  $C_{sy}$ . Interventions in category (3) may be successful or unsuccessful regardless of whether they impose costs on individuals, and



therefore regardless of whether they qualify as nudges. At least four scenarios must be considered when evaluating the success or failure of an intervention: (A) the intervention generates only the desired behaviour, e.g. a local authority introduces a progressive tariff for waste collection based on the amount produced; as a result, citizens reduce their waste and increase recycling rates; (B) the intervention produces no behavioural change, e.g. the government launches an anti-smoking awareness campaign via TV ads and posters, yet the habits of smokers remain unchanged; (C) the intervention produces the desired behaviour but also causes an undesirable one, e.g. to reduce traffic congestion, a city introduces a congestion charge for access to the city centre, leading to decreased traffic during peak hours but increased pollution in suburban areas (D) the intervention generates only undesirable behaviours – for example, to reduce sugar consumption, the government heavily taxes fizzy drinks; yet, instead of reducing sugar intake, consumers switch to equally unhealthy but untaxed alternatives, such as industrial fruit juices.

Drawing on Tuzet (2016), efficacy can be understood in two distinct ways: in a legal-technical sense, as the ability of an intervention to produce – or at least be able to produce – certain effects; and in a more philosophical sense, as the extent to which the intervention realises the aims for which it was devised. The second definition allows us to further distinguish efficacy from effectiveness, the latter referring to the degree to which an intervention is actually observed or complied with in practice<sup>3</sup>. These two concepts are independent: one may exist without the other. This means that an intervention can be effective (in terms of implementation) but not efficacious (in achieving its goals), or vice versa. A common belief is that nudges offer public administrations quick, visible results at relatively low political and administrative cost. But this does not mean that they are necessarily

---

<sup>3</sup> In this regard, Tuzet refers specifically to regulatory interventions, that is to legal norms; however, this classification is well suited to broader reflections on the design of any direct or indirect measure.

efficacious in addressing the underlying policy problem. A nudge can display high levels of observed compliance (high effectiveness) and still fail to tackle the structural causes of the issue it targets (low efficacy). It is therefore worth examining how design flaws in an intervention might compromise either its efficacy or its effectiveness. To visually represent the causal relationship between policies and behaviours, we can illustrate an intervention (I) – whether command-and-control or a nudge – with a solid line when the behaviour (a) directly influences behaviour (b). This model corresponds to scenario (A), where both efficacy and effectiveness are beyond doubt. At this stage, it is important to make a preliminary observation: in terms of causal relationships, nudge-based measures and command-and-control measures are equivalent. However, at times, the outcome produced may differ from the one originally intended. Even though the initial intention is to generate a particular effect through the intervention, that specific outcome does not occur, and instead, another effect is produced. This relationship can be represented with a dashed line. This model aptly describes scenario (D), in which only undesirable effects are generated, making the intervention both ineffective and inefficacious. The behaviours initiating the causal chain may or may not qualify as nudges, depending on whether the effect on the next node in the chain is achieved by altering the cost for the individual performing the relevant behaviour.

A further observation on efficacy is necessary. Consider the case in which a government sends personalised letters to taxpayers containing messages such as: “90% of your fellow citizens pay their taxes on time. If you are part of the minority who don’t, you may be fined”. The aim is to encourage individuals to declare their income correctly. The intervention is effective, in that it leads some people to regularise their tax status. However, other citizens, worried about potential consequences, turn to intermediaries (e.g., accountants, consultants), who often continue to suggest evasive strategies, undermining accurate tax calculation. Although the behaviour of some individuals improves, the overall effect of the intervention is inefficacious,

because the broader context continues to offer structural incentives for tax evasion. In this case, the intervention is observed and applied (demonstrating effectiveness), and the intention to modify behaviour aligns with the outcome. Graphically, this would correspond to scenario (C): the intervention is effective to some extent but also generates a directly related undesirable behaviour. If we assess efficacy in terms of the immediate behavioural change achieved, the intervention might appear efficacious. However, if efficacy is judged in relation to the ultimate aim of the intervention – i.e., addressing the systemic problem it was designed to resolve – then it fails to achieve that purpose. Thus, we have two evaluative criteria: efficacy in influencing behaviour and efficacy in achieving the intended target. When the latter is lacking, the intervention cannot be deemed truly efficacious and must be considered a failure. We may now consider scenario (B), in which there is no causal link between fact (a) and fact (b). In such cases, no behavioural change occurs as a result of the intervention. Since the intended outcome is neither enacted nor observed, the intervention is ineffective.

Representation of Causal Relationships Between Policies and Behaviors

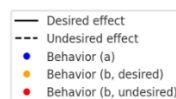
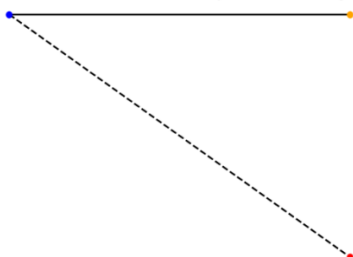
Case (A): Effectiveness and efficacy

Case (B): Ineffectiveness but ultimate efficacy



Case (C): Effective but partial efficacy

Case (D): Undesired effect



Let us suppose, for example, that in an effort to reduce cronyism in the appointment of senior public officials, the government introduces an online portal that publishes in real time the names and CVs of candidates selected for high-level positions. The idea is that increased transparency will discourage non-meritocratic practices. However, the intervention is largely ignored, as entrenched networks of favouritism persist and public officials continue to appoint individuals close to their personal or political circles, regardless of media exposure. Despite this, the heightened attention from the public and the media, caused by the introduction of the policy, increases social and political pressure on the issue, eventually compelling public decision-makers to introduce stricter evaluation criteria for appointments. Thus, even though the original intervention did not function as intended, it still contributed to a positive change aligned with its initial objectives. Cases like this may be rare, but they are not unthinkable. The considerations on efficacy outlined above apply here as well: the intervention is ineffective, yet ultimately efficacious, as it achieves its intended purpose through non-causal means. Nonetheless, the intervention fails, as its implementation was not observed, and the efficacy arose from external factors not inherent in the intervention itself. In scenarios (B) and (C) just discussed, we encounter a common flaw in the design of the intervention. In the first case, the systemic component of the context cancels out the potential individual shift prompted by the nudge. In the second case – such as the role of intermediaries in tax compliance – failure to intervene in the broader institutional conditions results in the emergence of alternative behaviours that neutralise the desired change or even produce unintended effects. In other words, that part of the context ( $C_{sy}$ ) comprising legal, economic, institutional or social structures remains the same and alters the intended behavioural effect of the intervention.

This taxonomy is relevant not merely for distinguishing between different forms of failure, but because it reveals how both behavioural and normative interventions may fall short when they neglect the systemic component of the context. Nudges without regulatory backing often fail to generate lasting

change, as in the case of personalised tax reminders that do not modify institutional incentives. Conversely, formal regulations that overlook behavioural dynamics (such as complex administrative reforms implemented without adequate support mechanisms) may also prove ineffective. These examples reinforce the need for integrated interventions that address both individual behaviour and the structural conditions under which it occurs, a strategy that the concept of meta-nudge aims to promote.

## **6. Meta-nudge: Possible Counterexamples to the S-frame Critique?**

Across the extensive literature on nudges, numerous authors have proposed taxonomies aimed at classifying interventions based on various criteria: for example, according to the aspect of reality they seek to modify (Münscher, Vetter and Scheuerle 2016), such as information, structure, or decision assistance; the cognitive processes involved (Luo, Li, Soman and Zhao 2023), such as attention, memory, or intrinsic/extrinsic motivation; the medium through which behaviour is influenced, such as digital nudges (Valta and Maier 2025); or even the degree to which individual autonomy is impacted (Baldwin 2014). While these classifications offer useful tools for application, they often fall short of explaining why a particular intervention is efficacious, which causal processes are activated, and which actors are most strongly affected by them. It is possible to move beyond these approaches by analysing nudges through the lens of the causal relationship they establish with the target behaviour. This perspective enables a distinction between two different types of nudges: on one hand, direct nudges, which act immediately on the individual choice; on the other, meta-nudges, which influence the broader context by acting upon intermediate agents (e.g., public officials, financial or commercial intermediaries) who in turn shape the behaviour of others through

mechanisms such as enforcement, normative expectations, or subsequent interventions (Dimant and Shalvi 2022)<sup>4</sup>.

The meta-nudge thus emerges as a form of systemic intervention. It does not aim merely to change individual behaviour, but rather inserts itself into a broader causal chain, acting upon the regulatory and operational conditions that influence the choices of multiple actors simultaneously. It does so by targeting those figures who are in a position to deliberately influence the actions of others. This makes it a particularly powerful tool for public administration, where institutional leverage can be exercised across several levels, generating lasting and scalable effects on collective behavioural dynamics. However, it is important to clarify that the concept of meta-nudge should not be conflated with the broader domain of administrative law or with general measures aimed at regulating the behaviour of public officials. Not all norms targeting intermediaries qualify as meta-nudges. What characterises a meta-nudge is not the normative status of the rule, but its behavioural structure. The intervention must be intentionally designed to influence the decision-making heuristics, perceptions, or framing processes of intermediaries, rather than simply prescribing duties or procedures. In this sense, meta-nudge occupies a specific space within the wider set of systemic interventions, defined by its cognitive and indirect mode of influence. By exploiting the automatic processes of human reasoning, the nudge seeks to create conditions in which the choices of an individual spontaneously align with those intended by the choice architect, without any formal rule requiring such alignment.

This approach evokes the form of governmentalist power described by Foucault as governmentality, i.e. the capacity to create and maintain circumstances in which the actions of a multiplicity of actors naturally

---

<sup>4</sup> I will use the term “meta-nudge” to refer to the concept as defined by Dimant and Shalvi (2022). However, the term has also been employed in a different sense, to describe an intervention designed to prompt critical reflection on the effects of another nudge, thereby safeguarding individual autonomy (Gelfand 2023). That said, “counter-nudge” would arguably be a more appropriate term for this latter type of intervention.

converge – without coercion – towards a broader, deemed-optimal order (Brigaglia, 2019). As with the norm-nudges<sup>5</sup> analysed by Bicchieri and Dimant (2022), the effect of a meta-nudge is not always immediate or linear. Its success depends on the distribution of social expectations, the role played by the actors influenced, and their ability either to amplify or dilute the initial intention of the policymaker. Yet, precisely because of its cascading structure, the meta-nudge represents a theoretical evolution of the traditional nudge, one geared towards the systemic design of behavioural change. If nudges can be linked to a form of governmentalist power, then the meta-nudge appears to embody another modality of influence identified by Foucault, namely a de-subjectified form of power that lies somewhere between intentional influence (power in the strict sense) and unintentional, in other words the anonymous power. In this case, the intervention no longer directly governs behaviour of people, instead operates diffusely and across networks, through repeated actions that follow an originally designed scheme and continue to generate the intended effects, even in the absence of a subject who maintains conscious control over the original intention (Brigaglia 2019,127).

This gives rise to a faceless dynamic of power, in which the initial intervention by the public decision-maker is internalised by other actors – either intermediaries or secondary agents – who then replicate, amplify or institutionalise that influence, often unintentionally. From this perspective, the meta-nudge is no longer merely a technical instrument but becomes the trigger of an impersonal process that fills the space of decision-making with power, without exercising it directly. Although the concept of choice architecture has played a crucial role in shaping the behavioural public policy approach to individual decisions, the growing interest in systemic and multi-level interventions now clearly demonstrates the need to broaden its scope. Interventions limited to specific decision-making moments, however well designed, risk failure when they come up against structural barriers,

---

<sup>5</sup> This type of nudge leverages social norms, i.e., what individuals perceive to be common or desirable behaviour, to steer individual choices.

regulatory misalignments or perverse incentives. Factors that a single choice architecture, by its very nature, is not equipped to address. For this reason, it has been proposed that the notion of choice architecture be complemented – or, in some cases, replaced – by the concept of choice infrastructure (Schmidt 2024). Unlike the former, which focuses on micro-environments of decision-making and targeted interventions on individual behaviour, choice infrastructure refers to the broader systemic conditions, institutional structures, operational processes, and functioning rules that support (or hinder) the efficacy of such interventions over time. It is the invisible yet essential technical framework of the system<sup>6</sup>, the web of causal factors that determines whether a behavioural policy succeeds or fails. Within the context of public administration, this paradigm shift proves particularly valuable for at least three reasons: (1) Decisions do not occur in isolated contexts but within complex and hierarchically organised environments, where citizens, civil servants, policymakers, and intermediaries all interact; (2) institutional dynamics are often slow-moving, distributed and governed by both formal and informal norms, rendering ineffective any approach that fails to account for these systemic constraints; (3) public interventions often aim to influence not only individual but also collective and recurring behaviours, with effects that are spread across multiple categories of actors. In such a context, introducing the concept of choice infrastructure enables the design not only of individual decision points but also of stable and reproducible conditions that facilitate the alignment of desired behaviours with the broader aims of public policy.

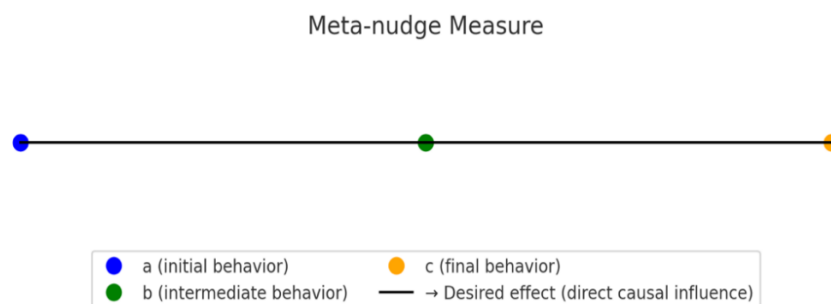
This shift in approach is what makes an effective meta-nudge possible: it acts not on a single move, but on the rules of the game themselves, indirectly influencing a wide array of agents and creating a structured pathway for behavioural change. Whereas the “classic” nudge is characterised by a direct

---

<sup>6</sup>On this point, E. Duflo (2017) offers a particularly fitting metaphor: if choice architecture is like the visible furnishing of a house, then choice infrastructure is the “plumbing” that ensures water flows to the right tap at the right time.



relationship between the intervention and a change in individual behaviour, the meta-nudge introduces a mediated and cascading causal structure, involving multiple actors and decision-making levels. The intention of the public decision-maker – who may be both a legislative or administrative actor – is not aimed directly at the final behaviour, but rather at the design or alteration of the decision-making environment of other subjects, who in turn influence the final behaviour of third persons. The public decision-maker implements a meta-nudge in order to influence an intermediate actor (such as a functionary, manager or organisation). This intermediary then acts as a new choice architect, generating a secondary intervention or reshaping the choice structure for a second subject, who is the ultimate target of the desired behavioural change. Finally, the behaviour of the end-user changes not as a direct response to the original intervention, but because of the action of an intermediary, which was shaped by the meta-nudge.



This causal chain highlights that the meta-nudge is not merely an enhancement of a classical nudge, but rather a form of second-order design, one that acts upon the capacity of other actors to exert behavioural influence. It is here that the systemic dimension becomes evident: the intervention is de-personalised, aiming to generate automatic influence within the intermediary subject. The effect of the intervention propagates through institutional relationships, roles, and implementation mechanisms operating across multiple levels. This means that the original intention tends to give rise to a subsequent intention, one that, even if not identical, is nonetheless oriented

towards producing a convergent final outcome. The concept of meta-nudge is therefore central to the theoretical claim advanced in this paper. By showing that behavioural interventions can operate on the systemic context (not only on individual decision points) meta-nudges directly challenge the idea that nudging is inherently limited to i-frame strategies. They demonstrate that nudging can take on a structural dimension, and thus contribute meaningfully to institutional and policy change. This confirms the central thesis: that nudges, when designed with attention to causal depth and intermediary dynamics, can fully inhabit the s-frame.

### *6.1 Meta-nudge in Administrative Law: The Principle of Trust in the Italian Public Contracts Code*

The introduction of the principle of trust in the Public Contracts Code (Legislative Decree no. 36/2023)<sup>7</sup> represents an example of a regulatory meta-nudge, that is, a systemic intervention that does not act directly upon the behaviour of citizens or businesses, but rather modifies the conduct of intermediate actors, in this case public officials, who in turn shape the decisions of other involved parties. This kind of intervention, which focuses more on transforming institutional contexts than on influencing individuals, goes beyond both the traditional command-and-control model and the limitations of individual nudges, opening the door to broader reflection on systemic governance.

According to article 2 of Legislative Decree 36/2023, trust becomes one of the founding principles of administrative action in matters of public procurement, alongside those of outcome-orientation, legality, market access, and good faith. This legislative choice marks a decisive shift in paradigm

---

<sup>7</sup> The text of the first two paragraphs under examination is reproduced here: “The allocation and exercise of power in the field of public procurement is based on the principle of mutual trust in the lawful, transparent, and proper conduct of the administration, its officials, and economic operators. The principle of trust encourages and enhances the initiative and decision-making autonomy of public officials, with particular regard to assessments and choices concerning the acquisition and execution of services, in accordance with the principle of achieving results.” (my translation).

compared to the previous code, which had been characterised by a defensive and formalistic framework, rooted in a presumption of distrust towards public officials and economic operators. The introduction of the principle of trust is not merely declarative; it aims to bring about a profound cultural and organisational transformation within the administrative apparatus (Carloni 2024). It is part of a logic aimed at re-functionalising administrative discretion, no longer seen as a grey area to be contained, but rather as an operational space to be preserved through accountability and alignment with public interest objectives. The norm is grounded in a conception of trust that tightly links autonomy and responsibility. Those applying the norm should, on the one hand, have the faculty to assess whether the “conditions of trust” exist in order to depart from mechanical rule application, and on the other, be subject to a form of oversight that prevents arbitrariness and abuse, however without automatically resorting to sanctions that would undermine the very purpose of the rule (Ursi 2024).

From a behavioural perspective, the principle of trust functions as a meta-nudge because it reshapes the expectations, perceptions, and internal constraints of public functionaries and ultimately those of the general public. It re-legitimises the exercise of discretion by reducing the freezing effect caused by what has been referred to as “fear of signing”, i.e. the tendency toward decision paralysis due to fear of future repercussions. The functionary is encouraged to assume an active, diligent, and outcome-oriented role, grounded in a logic of substantive rather than merely procedural accountability. In this way, the administration is no longer simply a constraint on economic activity but becomes an active facilitator for the economic operators with whom it interacts. The legislative intervention does not directly modify the behaviour of citizens but rather reforms the internal operational logic of public administrations, acting upon the intermediate nodes of the public decision-making chain. It is within these figures – functionaries, managers and procedural officers – that the true systemic efficacy of the principle lies.

This cognitive shift in the behaviour of public functionaries is intended to produce, in cascade, a transformation in the behaviour of all actors involved in the administrative procedure. The perception of a more reliable and less punitive system, more oriented towards cooperation than surveillance, encourages greater compliance, reduces litigation, and fosters a higher degree of spontaneous adherence to the rules. The multiplier effect of the principle of trust is realised insofar as it acts upon a class of actors (functionaries) who are capable of structuring the decision-making context of others (economic operators), according to a logic of institutional propagation of influence. This process is all the more efficacious when it is supported by organisational coherence, behavioural training, and reinforcement measures that reward virtuous conduct. It thus represents a form of systemic influence, in which trust operates as a cognitive simplification within decision-making. Its effect on the context is twofold. On the one hand, the principle acts as a transformative factor for the administrative choice infrastructure ( $C_{sy}$ ), redefining implicit rules and organisational routines, making the system more receptive to change and more outcome oriented. On the other hand, it generates a new form of diffuse normativity, nourished by coherence between regulatory intentions and expected behaviours, giving rise to a networked form of governmentality (Foucault 2004) that does not impose but rather guides, it does not prescribe but instead structures.

This framework fosters distributed institutional learning, where the improvement of processes occurs not only through regulatory means but also through the reinforcement of practices and collective expectations. What is particularly significant about this type of intervention is that it operates without sanctions or formal coercion. Rather than imposing duties or threatening consequences, it seeks to influence the behaviour of public officials by redefining the institutional framework within which their decisions take place. This non-authoritative character is one of its strengths: by avoiding the rigidity of command-and-control mechanisms, it allows for greater adaptability and endogenous appropriation within administrative

practice. As noted in the Italian legal literature, such interventions challenge the traditional conception of administrative power as the exercise of formal legal authority and instead promote a vision of power as intentional influence over decision-making contexts (Zito 2021). In this view, the effectiveness of behavioural governance does not derive from enforceability, but from its ability to embed desirable orientations into the very environment in which decisions are made.

The norm does not alter the costs faced by agents. Its aim is to promote the presumption that the government trusts public officials, thereby reducing the irrational “fear of signing” that affects many public functionaries. At the same time, it seeks to foster among citizens the social norm that officials may legitimately exercise discretion, without modifying the existing legal boundaries of such discretion. Whether interventions of this kind can produce systemic effects remains an empirical question. Conceptually, however, there is no reason to believe they cannot.

This case shows how S-frame interventions can be made through nudging. I-frame and S-frame interventions are not necessarily associated with a specific mode of influence. While the first focuses only on modifying individual behaviours ( $C_i$ ) and the latter on changing normative structures ( $C_{sy}$ ) the trust-based meta-nudge operates on both levels ( $C$ ), activating a circuit of influence between system and behaviour, between rules and practices, between norms and action. The outcome is an administration that perceives itself – and is perceived – not merely as an executor of procedures, but as a relational actor, capable of fostering trust-based and generative decision environments. There is a risk of oversimplifying the discourse by assigning *a priori* superiority to one type of intervention over another, without considering the concrete conditions of design and implementation. The interpretation of the principle of trust as a form of meta-nudge rests on its specific mode of operation. It does not prescribe behaviour, alter legal entitlements, or impose sanctions, but reorients the interpretive and decisional context within which public officials act. What qualifies it as a meta-nudge is

not the presence of regulatory content per se, but the absence of coercion and the presence of a systemic mechanism of indirect behavioural influence. Rather than governing end-users directly, the intervention targets an intermediary class (functionaries) whose practices shape the conditions under which others make decisions. The legislative norm functions as a behavioural lever within institutional practice, aiming to transform defaults, routines, and expectations at multiple levels of administration. This meets the definition of a meta-nudge insofar as it triggers a cascading effect through the reconfiguration of a choice infrastructure (rather than a micro-level choice architecture), generating systemic outcomes through non-coercive, intention-driven influence. While grounded in legal form, its operative logic is behavioural, not prescriptive. For this reason, it can be understood not as classic regulation, but as a paradigmatic case of a norm that governs through trust and systemic propagation rather than command and control.

## 7. Final Considerations

The claim of this work has been to critically reformulate the conditions under which behavioural interventions in public policy can be considered efficacious, moving beyond static categories and conceptual dichotomies that too often impoverish the analysis. It has sought to question the rigid division between i-frame interventions (targeting individuals) and s-frame interventions (targeting systems), a distinction that tends to assume behavioural tools are inherently incapable of producing structural effects. The critique levelled against the nudge, in its classical formulation, has been valuable in highlighting the risks of depoliticization and systemic inefficacy. However, it implicitly assumes that the form of an intervention determines its capacity to bring about change.

Within this perspective, nudges would be confined to minor behavioural corrections, whereas only hard rules and coercive interventions would be capable of transforming the context. The argumentative path proposed here

aims to show that such a view is inadequate. An intervention succeeds or fails depending on how it is designed, not on the label it carries. The true key lies in the causal structure of the intervention: which behaviours it aims to modify, through which actors, in what context, and with what expected (or unintended) effects. Every regulatory measure – whether legal, incentive-based, or behavioural – can produce systemic effects, provided it is incorporated within a coherent strategy that considers the cognitive, organisational, and institutional constraints faced by the actors involved.

In this sense, the meta-nudge represents both a theoretical and operational proposal capable of overcoming the s-frame critique without abandoning the behavioural approach. By acting on intermediate actors who hold the power to shape the choices of others, the meta-nudge initiates a causal chain that transforms not only behaviours but also the underlying organisational and cultural logics. The example of the principle of trust in the Italian public contracts code clearly illustrates this point: it is a norm that, without imposing, has success in recalibrating administrative discretion, activating the accountability of the functionaries, and fostering cooperation, while reducing bureaucratic distortions. Starting from this example, one can argue that even a nudge, if properly designed, can generate systemic change, and that every intervention should be assessed not by its formal category, but by its causal and relational function.

Thus, nudges can alter social norms and produce systemic effects. It is a mistake to assume that nudges can only ever constitute i-frame interventions. This distinction is crucial: the boundary between i-frame and s-frame does not align with the distinction between “weak” and “strong” tools but rather depends on the strategic objective and the causal network that the intervention is capable of activating. Nudges, when conceived to act upstream of decision-making conditions, can reshape implicit rules, organisational practices, and widely shared social norms. It should be emphasised, that social control measures are not monolithic or mutually exclusive, but form parts of a *continuum*. Coercion, incentives, facilitation and persuasion are all

techniques that follow different but complementary logics. Their efficacy often depends on their capacity to be integrated. In complex and differentiated contexts, a single tool is unlikely to produce lasting outcomes. On the contrary, it is through the simultaneous and well-calibrated adoption of multiple instruments – adapted to the type of actor, the decision-making moment, and the constraint to be addressed – that robust, adaptive, and genuinely transformative public policies can be built. Eventually, the contemporary challenge is not to choose between nudge or rule, or individual or system, but rather to rethink intervention design as intentional action across a multi-level causal chain. Only in this way can we move beyond reductive classifications and build a model of public governance able to generate trust, efficacy and systemic impact.

While this article has focused on the structural, functional and causal dimensions of behavioural interventions – particularly in relation to public governance and the concept of meta-nudge – it is important to acknowledge that such interventions, especially when implemented by public authorities, raise important questions concerning democratic legitimacy and individual autonomy. The debate on libertarian paternalism has long highlighted the potential opacity of behavioural tools and their capacity to influence choices without conscious awareness or deliberative endorsement (Hansen, Jespersen 2013; Baldwin 2014). Although these normative issues fall outside the primary scope of this paper, they remain crucial for evaluating the legitimacy – and not merely the efficacy – of meta-nudging strategies in institutional contexts. Some authors have pointed out that libertarian paternalism, even when well-intentioned, may give rise to subtle forms of asymmetry between those who design the choice architecture and those who are subjected to it. This imbalance, based on expertise, cognitive access, and institutional authority, can result in a condition of “supervised freedom,” where individuals formally retain autonomy but are structurally nudged towards preferred behaviours without fully transparent justification (Galletti, Vida 2018). Future research should therefore complement the functional analysis



of behavioural interventions with a reflection on their normative implications, especially in terms of how they interact with democratic processes and public reasoning, without assuming that such tools are inherently manipulative or illegitimate.

## References

- Baldwin R. (2014). From regulation to behaviour change: giving nudge the third degree, *The Modern Law Review*, vol. 77, n. 6.
- Bicchieri C. and Dimant E. (2022). Nudging with care: the risks and benefits of social information, in *Public Choice*, n. 191.
- Bobbio N. (2007). *Dalla struttura alla funzione. Nuovi studi di teoria del diritto* (Laterza).
- Brigaglia M. (2019). *Potere. Una rilettura di Michel Foucault* (Editoriale Scientifica).
- Carloni E. (2024). Verso il paradigma fiduciario? Il principio di fiducia nel nuovo Codice dei contratti e le sue implicazioni, in *Diritto pubblico*, n. 1.
- Chater N. and Loewenstein G. (2023). The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray, in *Behavioral and Brain Sciences*, n. 46.
- Congiu L. and Moscati I. (2022). A review of nudges: Definitions, justifications, effectiveness, in *Journal of Economic Surveys*, n. 36.
- Connolly D. J., Loewenstein G. and Chater N. (2024). An s-frame agenda for behavioral public policy research, in *Behavioural Public Policy*, first view.
- Crozier M. (1969). *Il fenomeno burocratico* (Etas Kompas).
- DellaVigna S. and Linos E. (2022). RCTs to Scale: Comprehensive Evidence from Two Nudge Units, in *Econometrica*, vol 90, n. 1.
- Dimant E. and Shalvi S. (2022). Meta-nudging honesty: Past, present, and future of the research frontier, in *Current Opinion in Psychology*, n. 47.
- Duflo E. (2017). The economist as plumber, in *American Economic Review*, vol. 107, n. 5.

- Foucault M. (2004). *Sécurité, territoire, population. Cours au Collège de France (1977-1978)* (Seuil-Gallimard).
- Gelfand S. D. (2023). Nudging, Bullshitting, and the Meta-Nudge, in *Cambridge Quarterly of Healthcare Ethics*, vol. 32, n. 1.
- Gigerenzer G. (2000). *Adaptive Thinking: Rationality in the Real World* (Oxford University Press).
- Galletti G. and Vida S. (2018). *Libertà vigilata. Una critica del paternalismo libertario* (IF Press).
- Hansen P. G. (2016). The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove?, in *European Journal of Risk Regulation*, vol. 7, n. 1.
- Hansen P. G. and Jespersen A. M. (2013). Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy, in *European Journal of Risk Regulation*, vol. 4, n. 1.
- Johnson E. J. (2022). *The elements of choice: Why the way we decide matters* (Simon and Schuster).
- Kahneman D. (2012). *Thinking, fast and slow* (Penguin Books).
- Kahneman D., Knetsch J. L. and Thaler R. H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias, in *Journal of Economic Perspectives*, vol. 5, n. 1.
- Liebe U., Gewinner J. and Diekmann A. (2021). Behavioral interventions and climate change: A meta-analysis, in *Ecological Economics*, n. 185.
- Luo Y., Li A., Soman D. and Zhao J. (2023). A meta-analytic cognitive framework of nudge and sludge, in *Royal Society Open Science*, vol. 10, n. 11.
- Madva A., Brownstein M. and Kelly D. (2023). It's always both: Changing individuals requires changing systems and changing systems requires changing individuals, in *Behavioral and Brain Sciences*, n. 46.

- Münscher R., Vetter M. and Scheuerle T. (2016). A review and taxonomy of choice architecture techniques, in *Journal of Behavioral Decision Making*, vol. 29, n. 5.
- Schmidt R. (2024). A model for choice infrastructure: looking beyond choice architecture in Behavioral Public Policy, in *Behavioural Public Policy*, n. 8.
- Simon H. A. (1955). A Behavioral Model of Rational Choice, in *The Quarterly Journal of Economics*, vol. 69, n. 1.
- Simon H. A. (1957). *Models of Man: Social and Rational* (Wiley).
- Simon H. A. (1997). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations* (Simon and Schuster).
- Stanovich K. E. and West R. F. (2000). Individual Differences in Reasoning: Implications for the Rationality Debate?, in *Behavioral and Brain Sciences*, vol. 23, n. 5.
- Starke A. D. and Willemsen M. C. (2024). Psychologically Informed Design of Energy Recommender Systems, in B. Ferwerda, M. Graus, P. Germanakos and M. Tkalčič (eds.), *A Human-Centered Perspective of Intelligent Personalized Environments and Systems* (Springer).
- Sunstein C. R. (2021). *Sludge: What Stops Us from Getting Things Done and What to Do About It* (MIT Press).
- Sunstein C. R. (2022). The rhetoric of reaction redux., in *Behavioural Public Policy*, vol. 7, n. 3.
- Sunstein C. R. (2023). Conspiracy theory, in *Behavioral and Brain Sciences*, n. 46.
- Thaler R. H. (2023). Nudging is being framed, in *Behavioral and Brain Sciences*, n. 46.
- Thaler R. H. and Sunstein C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Penguin).
- Thaler R. H. and Sunstein C. R. (2021). *Nudge: The Final Edition* (Yale University Press).

- Thaler R. H., Sunstein C. R. and Balz, J. P. (2014). Choice architecture, in E. Shafir (ed.), *The Behavioral Foundations of Public Policy* (Princeton University Press).
- Tuzet G. (2016). Effettività, efficacia, efficienza, in *Materiali per una storia della cultura giuridica*, n. 1.
- Ursi R. (2024). La ‘trappola della fiducia’ nel Codice dei contratti pubblici, in *Bilancio comunità persona*, n. 1.
- Valta M. and Maier C. (2025). Digital Nudging: A Systematic Literature Review, Taxonomy, and Future Research Directions, in *SIGMIS Database*, vol. 56, n. 1.
- Zito A. (2021). *La nudge regulation nella teoria giuridica dell’agire amministrativo. Presupposti e limiti del suo utilizzo da parte delle pubbliche amministrazioni* (Editoriale Scientifica).