

When Nudges Backfire: A Distinction Between Spillovers and Misfires

Christos Themistocleous¹ [0000-0001-9074-7711] and Evangelos Karapanos¹ [0000-0001-5910-4996]

¹ Persuasive Tech Lab, Cyprus University of Technology, Cyprus
c.themistocleous@cut.ac.cy

Abstract. The concept of nudging (libertarian paternalism) has been eagerly adopted in the design of digital behavioural interventions, yet discussion on adverse effects of nudges has been limited. In an attempt to enhance resolution, this theoretical paper conceptually differentiates two types of adverse effects, namely, *spillovers* and *misfires*. We propose that contrary to behavioural spillovers that occur as a consequence of a targeted behaviour, the causality of misfires is primarily attributed to the nudge type. In doing so, we employ visual representations of the behavioural trajectories that demonstrate intended and unintended effects of nudges to support choice architects in blueprinting intervention processes. We differentiate between spillovers and misfires by summarizing key empirical research that seeks to explain the causes of each and the contextual nature that distinguishes them. We conclude by drawing implications for the design of technology-mediated nudges as well as the role of its *timing*.

Keywords: Nudge Theory; Behavioural Spillover; Nudge Misfire; Behavioural Pathway; Libertarian Paternalism

1 Introduction

Behavioural sciences propose a series of intervention tools that aim to alter courses of human action to more beneficial alternatives. Such tools are wide in range and their nature is dependent on the discipline from which they emerged. The use of incentives or taxes are widely accepted propositions that stem from economics [1]; Boosting and the enhancement of human competencies in decision-making serves as a viable proposition by developmental psychologists [2], while Nudging and the use of interventions that primarily make use of human biases, originate from behavioural economics [3].

Focusing on the latter, Nudge theory, founded on Tversky's and Kahneman's [4] heuristics and biases program, is only seemingly reaching its maturity stage via a series of meta-analyses employed to assess the tool's competency [5-7] in a variety of contextual applications [8-9]. Counterintuitive results, nevertheless, make the nudge's effectiveness inconclusive at best. For example, in a meta-analysis by Mertens et al. [10], nudges were found to be effective across a range of contextual applications, yet follow-

up examinations by Maier et al. [11] refuted these findings on the basis that publication bias was not accounted for. As such, when results were adjusted using RoBMA (Robust Bayesian Meta-Analysis), favourable nudge effects eclipsed. These examples demonstrate a growing pattern in the literature that focuses on the tool’s effectiveness (positive utility) and ineffectiveness (neutral utility), yet there are fewer examinations on the unintended adverse effects (negative utility) caused by failed nudges.

Sunstein’s [12] acknowledgment of failed nudges led to distinguishing between ineffective and counterproductive nudges, yet other than reactance, little is offered on the causality of the latter type and by extension on ways to mitigate it. In addressing counterproductive interventions, research on negative behavioural spillovers serves as the primary source for insights. Dolan and Galizzi [13] categorized factors like moral licensing and ego depletion, as possible predictors of spillovers. Their conceptualisation was based on how a targeted behaviour 1 can lead to an unintended behaviour 2, with the main focus fixed on behaviour 2. The premise of such conceptualisations is that spillovers occur after an individual has entered into the intended behavioural trajectory. For example, a social comparison nudge, that provides data to the user comparative to the performance of other similar users, can push an individual to start exercising (behaviour 1) yet at the end of the exercise cycle, lead the user to consume a cake (behaviour 2) demonstrating a negative spillover.

We argue that adverse effects can also occur prior to entering into behaviour 1 and as a consequence of a rejected nudge. This signifies a difference to adverse effects attributable to spillovers. For example, an individual might reject a social comparison nudge to exercise in the form of “Out of 1000 similar users you are ranked 990th for your calorie-burning performance”. Due to this negative social comparison that reminds the individual of their overly negative performance, the individual does not engage in exercising, and as a consequence, her self-esteem drops [14], ultimately leading to cake consumption. Here, behaviour 2 was triggered because of the nudge type and not as a consequence of behaviour 1.

In this theoretical paper we conceptualise the above as a *Nudge Misfire* and argue that contrary to nudge spillovers that can be examined and mitigated through the psychological lenses proposed by relevant spillover research [13, 15], misfires and their adverse effects are primarily caused by the type of nudge employed as well as its timing [16]. We focus on the definition of Nudges as proposed by Thaler and Sunstein [3], namely, libertarian paternalism, and how the tool is informed by research on heuristics and decision-making processes. As such nudges that induce, for example, social comparisons, sequence effects, framing effects and defaults are relevant examples of the tool’s application. Our focus has less to do with issues pertaining to autonomy restrictions, ethical or philosophical issues of nudges. Instead, we focus more on dimensions where nudges generate anomalous and inconsistent behaviours compared to the intentions of the choice architect that don’t just position individuals in neutral situations but in worse-off ones.

This work contributes on three fronts. First, we differentiate between nudge spillovers and misfires by summarizing key empirical research that seek to explain the causes of these adverse effects, such as the role of moral licensing and ego-depletion in spillovers, and the contextual nature of misfires. We propose behavioural pathways (visual

representations of behavioural trajectories following a nudge) as a design support tool that enables choice architects in identifying intended and unintended effects of nudges. Secondly, we use Dolan and Galizzi's [13] framework to expand on the causes of negative spillovers. We extend implications by proposing ways of mitigating each of the five main causes of negative spillovers. Finally, we analyse the new category of nudge misfires and provide examples of nudge types that can lead to these adverse behavioural trajectories.

2 Defining Nudge Spillovers and Misfires

Spillovers have received the attention of behavioural science and psychology literature [17] yet their application and empirical investigation specifically in Nudge theory is relatively limited. A behavioural spillover is defined as the effect an intervention that targets behaviour 1 has towards a behaviour 2 [15, 18–20]. Specifically, Dolan and Galizzi [13] base their spillover categorization on “(...) two different behaviours that take place sequentially: behaviour 1 is followed by behaviour 2”. Following this logic, a nudge spillover occurs when a nudge that aims to address a behaviour also alters another, normally unaccounted, behaviour by the choice architect. Here we note that not all spillovers signify negative effects. For instance, the effects of physical activity nudges may extend to other health behaviours, such as adopting healthier nutrition (i.e., positive consistency, [21]). Our focus, nonetheless, is on spillovers that generate adverse effects on individuals - for example nudges that aimed at reducing household energy consumption through switching incandescent light bulbs to LED ones, yet lead to increased consumption, due to a follow-up change in the occupants' behaviour (e.g., leaving the lights on when away).

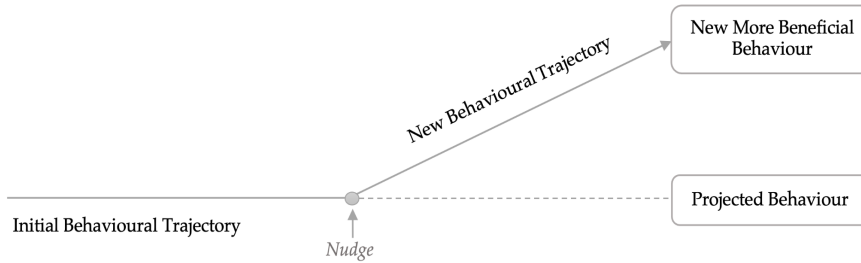
Nudge misfires, contrary to spillovers, lead to unintended adverse effects before the adoption of the intended behaviour. For example, a social comparison nudge that positions an individual in the lowest quartile of exercise time might not only discourage her from exercising but can also induce a worse-off psychological state that leads to calorie consumption. In the latter case two things occurred, first the nudge was ignored in the sense that the individual did not engage in exercising, and second, a worse-off position was accidentally instigated to the nudged individual via calorie consumption. Behavioural interventions that backfire is a topic touched upon by both HCI [22] and behavioural research [15] yet a distinction, at the procedural level, of adverse behavioural effects to our knowledge remains underexplored.

3 Differentiating between Nudge Spillovers and Misfires

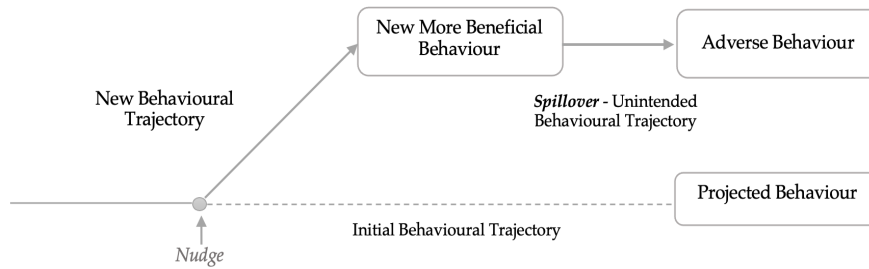
Contrary to Spillovers, Nudge misfires occur when the individual rejects the nudged trajectory yet ends in a worse-off situation compared to the absence of that nudge. Conceptually, prior research positions the recommended term of Misfires as a type of Spillover [13] yet we argue that the two have adequate differences to justify a new taxonomy. We base this on two main premises. First, Spillovers are a consequence of the nudged behaviour while misfires are a consequence of the nudge type itself. Secondly,

there is a differentiation in the stage within the behavioural process where each adverse effect occurs. A spillover leads to a new behaviour 1 as a consequence of the nudged behaviour 2. Contrary to that, a misfire occurs when the individual rejects the nudge yet the dismissed nudge leads to a new worse-off behaviour. This can be attributed to either the nudge type and/or its timing [16]. In essence, the difference between *spillovers* and *misfires* can be summarised as follows: The adverse behaviour for the spillovers occurs after accepting the nudged behaviour, while the adverse behaviour of misfires occurs after rejecting the nudge and as a consequence of it. In essence, the latter triggers a new behaviour trajectory by *missing* the behaviour it was intended to target and instigating a new negative one. Figure 1 depicts the *behavioural pathways* of each type. The visual representations demonstrate the intended and unintended behavioural changes as a result of the nudge.

1a. An intended Nudge



1b. A Nudge Spillover



1c. A Nudge Misfire:

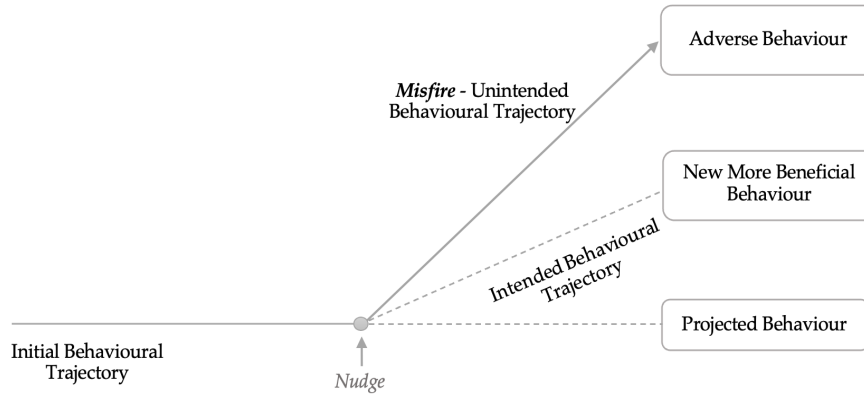


Fig. 1. Behavioural pathways for an intended Nudge (1a), a Nudge Spillover (1b) and a Nudge Misfire (1c).

4 Nudge Spillovers: Contributing Factors

Research on spillovers identifies a series of factors that create it. As negative spillovers are the present paper’s focus point, we bring attention to Dolan and Galizzi [13] categorisation of permitting spillovers where inducing a positive behaviour 1 leads to a negative behaviour 2. Their categorisation focused on five identified factors that contribute to these spillovers, namely: Moral Licensing, Ego Depletion, Reverse foot-in-door, Rest-on-Laurels and Single-Action bias. The objective of this section is to reflect on each type and propose ways of mitigation for choice architects and intervention designers.

4.1 Moral licensing

Moral licensing is a psychological phenomenon where individuals feel justified in engaging in behaviours that might otherwise conflict with their moral standards after they have previously acted in a morally positive way [23, 44]. Good deeds for example serve as "licenses" that allow the person to feel they have earned the right to act less virtuously. Merritt et al. [24] mention that moral licensing occurs when "... individuals can derive confidence from their past moral behaviour, such that an impeccable track record increases their propensity to engage in otherwise suspect actions."

Consider the previously discussed LED lights example. Moral licensing suggests that the good deed of energy saving by switching to LED lamps can justify a relaxed and *deserved* attitude of not checking whether lights are unnecessarily on. Moral licensing has seen a series of empirical investigations in different contexts. For example, a prompt that led to a charity donation justified the purchase of a luxury good (designer jeans) over a necessity for an individual [25]. In a similar fashion and in the context of dieting, Fishback and Dhar [26] identify how exercising can provide licensing for consuming high-calorie chocolate over a healthy fruit alternative. Embedded in moral

licensing is self-perception theory which suggests that individuals infer their own attitudes and beliefs by observing their behavior, especially when internal cues are weak or ambiguous [27]. As such after a good deed, individuals feel that their moral self-image is sufficiently strong. This can reduce the pressure to act morally in subsequent situations, as they feel they've earned the right for a moral deviation.

Handling spillovers caused by Moral Licensing: Empirical findings pertaining to moral licensing are especially valuable to nudge designers. Moral licensing is accentuated primarily in altruistic behaviours such as donations and other acts of philanthropy [25, 28]. Choice architects specifically in these domains need to be aware that individuals might indulge a nudge for a donation yet moral licensing might hinder them from more meaningful acts of charity. Default nudges with a recommended amount for a donation might be less relevant to tackle such a spillover effect. In handling this, assigning an *identity* to the individual in the form of framing and informative system 2 nudges that aim to instigate a sense of responsibility to philanthropic acts instead of being perceived as acts of generosity can mitigate this, as the latter could give *license* for other unsustainable hedonic behaviours.

4.2 Ego Depletion

Ego depletion suggests that self-control or willpower draws from a limited pool of mental resources that can be depleted with use [29]. When individuals exert significant self-control in one task, their ability to exercise self-control in subsequent tasks may be reduced, leading to diminished performance or decision-making capacity [30]. Dolan and Galizzi [13] acknowledge an association between ego depletion and spillovers especially when an ego depleting behaviour precedes one that requires willpower and cognitive energy to resist. They argue that exercising which is both a physically and mentally tiring task, can lower the individual's willpower to resist to a cake with equal (if not more) amount of calories to the ones burned by the activity. Applied to nudges, from figure 1b, an exercise-reminder nudge might trigger a new beneficial behaviour trajectory, yet a subsequent behaviour, that requires willpower to resist, might render the individual susceptible to it. If the second behaviour decreases the utility achieved by the nudge, we have a negative nudge spillover attributable to ego depletion.

Ego depletion has seen applications from sports [31, 32] and how tiring mental tasks affect performance, to how food choices are made following cognitively demanding tasks [33]. Nudges designed to reduce smoking renders that person's affective dimension superior over her cognitive one. This is due to the cognitive scarcity caused by the smoking reduction task. It is thus quite common for nudged smokers trying to quit, to demonstrate reduced willpower and thus lean more towards affect decisions than ones requiring cognitive effort.

Handling spillovers caused by Ego Depletion: Practical implications of ego depletion as a cause of nudge spillovers relate to the nature of behaviours affected by ego depletion. Nudging for exercising or for other willpower-demanding tasks, like quitting smoking, can lead to unintended spillovers like increasing sugar intake, eating more unhealthy food or buying affect-laden products. Due to the inability to predict the exact area of the spillover we propose an alternative way. Designers need to be aware of

demanding behaviours that they aim to alter and perform a *willpower-demanding task audit*. As a first step, mapping the targeted behaviours will allow the avoidance of nudges that simultaneously require significant effort from the individual, for example, exercising and quitting smoking at the same time. This dual targeting can maximise the chance of ego-depleted states and by extension susceptibility to affect decisions that follow. Step two and after mapping demanding tasks, designers can utilise the right *timing* for each nudge, allowing chronological space between nudges that target different willpower-demanding behaviours, one at a time.

4.3 Single-Action Bias

Single-action bias is a cognitive phenomenon where individuals take a single action in response to a perceived problem and then fail to take further necessary steps due to the emotional relief felt by the initial action [34]. Originally studied in the context of risk perception and decision-making, it has been particularly noted in environmental behaviour, health interventions, and financial decision-making [35]. When individuals face an issue, such as climate change or health risks, taking a single action can provide psychological relief, reducing the motivation for additional efforts. For instance, a person concerned about energy consumption might switch to LED bulbs but neglect further actions like reducing air travel or supporting impactful policy changes. Similar patterns emerge in health behaviour where a single preventive action, such as taking a vitamin supplement, can lead individuals to feel less urgency to adopt more impactful habits like regular exercise or a balanced diet.

Handling spillovers caused by Single-Action Bias: The implications of Single Action Bias are linked to how the initial action, on an issue important to the individual, leads to emotional relief. If individuals feel emotional consolation due to the initial action, nudge interventions aiming for long-term and systemic changes may be undermined. As such, one could mitigate this by employing strategies that encourage sustained action, for example, through adjusting nudges and their frequency in a way that (i) emphasizes one's incomplete efforts, (ii) provides continuous feedback, or (iii) structures commitments that promote follow-through strategies. For instance, reframing a nudge from a "thank you for recycling" to a "Great start! Here's what you can do next", can position the initial action as a starting point to a process. With respect to timing and frequency of nudge provision, we note that one-off nudges arguably exacerbate single-action bias, and should be complemented by tailored-to-the-user well-timed nudge reminders to facilitate follow-up actions.

4.4 Rest-on-Laurels effect

The Rest-on-Laurels effect occurs when individuals reduce further effort after an initial success, believing they have already done enough. Unlike single-action bias, which is driven by emotional relief after taking a single step, Rest-on-Laurels stems from self-satisfaction and a sense of accomplishment. Amir and Ariely [36] argue that individuals break an initial goal into smaller, more manageable sub-goals. The achievement of these subgoals can lead to a sense of success resulting to complacency which gives way

to behavioural stagnation hindering the completion of the initial *main* goal. The achievement of these sub-goals hinders an individual from the achievement of all sub-goals that constitute the main goal [37]. This effect is particularly evident in personal habits and social behaviour. For example, someone who sets a starting goal to exercise for a month, does so, yet the feeling of success to this goal can lead to him neglecting further workouts. The sub-goal achievement, hinders him from achieving the main goal which is about developing an exercise routine. Over time, this mindset can lead to declining motivation, missed opportunities for growth, and a false perception of long-term impact.

Handling spillovers caused by Rest-on-Laurels: To counteract this effect, interventions should frame achievements as steps in a process, encourage new goal-setting and provide progress-reminding nudges that mitigate this bias. Specifically, designers can capitalise on the latter and utilise nudges in the form of *visual progress indicators* that have two main dimensions: First congratulatory notes are offered to retain motivation by acknowledging subgoal achievements and secondly highlighting that such achievements constitute a part of a larger process and, by extension, that these efforts are incomplete without continuation to the next pillar of the process. For example, fitness apps can use streak tracking and/or progressive goal-setting to ensure users don't stop after achieving a few initial, short-term workout goals.

4.5 Reverse Foot-in-the-Door

The reverse foot-in-the-door spillover occurs when an individual, after completing a large or demanding action, becomes less likely to engage in smaller, follow-up actions. This contrasts with the classic foot-in-the-door effect, where a small initial action increases the likelihood of further engagement, (which is a positive spillover; [38]). Reverse foot-in-the-door is driven by effort-based exhaustion or a feeling of having already contributed enough [39]. In environmental behaviour, individuals who make a major commitment, such as installing solar panels, may neglect smaller sustainable actions like conserving daily energy [18]. Similarly, in health, someone who completes an intense workout may feel justified in skipping minor habits like taking the stairs. In charitable giving, a large one-time donation can reduce motivation for future smaller contributions.

Handling Reverse Foot-in-the-Door: The consequences of reverse foot-in-the-door spillover include behavioural stagnation and reduced long-term engagement. To counteract this, nudges should, (i) frame small actions as enhancing prior efforts, (ii) provide explanations that emphasize cumulative impact, and (iii) encourage ongoing goal-setting to sustain behaviour over time. Table 1 provides a summarisation of the mitigation strategies for each of the main five identified causes of negative spillovers.

Table 1. Mitigation strategies for the main five causes of negative spillovers.

Causes of Negative Spillovers	Mitigation Strategies	Description
-------------------------------	-----------------------	-------------

Moral Licensing	<i>Framing Nudges</i>	Framing <i>e.g. Philanthropic acts framed as responsibilities and not solely as acts of generosity</i>
	<i>Informative Nudges</i>	Informative <i>e.g. explanations as to the implications of systematic giving, which philanthropic acts can have the most impact, to whom and why</i>
Ego Depletion	<i>Willpower-demanding Task Audit</i>	Step 1. Identification of willpower-demanding tasks <i>e.g. quit smoking; start exercising</i>
		Step 2. Focused targeting <i>e.g. targeting a single willpower-demanding task at a time</i>
Single-Action Bias	<i>Progress-Reporting Nudges</i>	Visual Progress Bars <i>e.g. use of both dynamic and static types</i>
	<i>Framing Nudges</i>	Framing <i>Framing the initial action as a part of a larger process</i> <i>e.g. Instead of “thank you for recycling” frame as “Great start! Here’s what you can do next”</i>
Rest-on-Laurels	<i>Progress-Reporting Nudges</i>	Visual Progress Bars <i>e.g. both dynamic and static types.</i>
	<i>Reminder Nudges</i>	Reminders <i>e.g. reminding users of their remaining sub-goals; reminding them of their current stage in the process.</i>
Reverse Foot-in-the Door	<i>Progress-Reporting Nudges</i>	Visual Progress Bars <i>e.g. both dynamic and static types.</i>
	<i>Informative Nudges</i>	Informative <i>e.g. Explain the importance of cumulative impact while encouraging on-going goal setting.</i>

5 Nudge Misfires: Contributing Factors

When it comes to nudge misfires we argue that their causality is highly dependent on the nudge type used. Due to the case-specific nature of misfires, we review social comparison nudges as an example to analyse the nudge misfire category.

Social comparison nudges are based on Social comparison theory and the idea that individuals evaluate their own behaviors, attitudes, or abilities by comparing themselves to those of others. Social comparison nudges are prominently used in persuasive technology and beyond (see [40–42]). For example, a social comparison nudge can be used to promote exercising through comparisons. An individual’s exercising performance can be contrasted to a pool of others with similar demographic characteristics, generating a descriptive statistic reflective of their performance such as: “You are currently ranked 90th compared out of 100 others in the same age and sex group in terms of weekly exercise” (see [43]). Meta-analytical data show that comparisons can have negative effects dependent on the context of application and the behaviour they are designed to alter [44], including low self-esteem and depressing states [14]. As a result, individuals might reject the exercise prompt, but also engage in adverse health behaviours due to the adverse psychological effects of this nudge type.

Some ways to mitigate misfires for social comparison nudges include the nudge type and its timing. Consider the example of Gouveia et al. [43] who developed and field-trialed a physical activity promotion nudge that was comparing one’s daily steps to that of others’ sharing a similar daily goal. Every time users would check their watch, they would see two rings, representing their current and others’ average performance on a 5-min resolution. They found that when users checked their watch, they were quicker to initiate a new physical activity if they were not too far behind or ahead of others (+/- 500 steps).

While users check their watch on their volition, subtle prompts, being haptic, auditory or textual, could serve as attentional cues to engage with physical activity feedback. The *timing* technique would suggest increasing the frequency of those attentional cues when individuals’ performance is similar to that of others. The *nudge type* technique would suggest considering what information to provide and how to frame it, in order to minimize the likelihood of misfires. For instance, and while one could question the ethics of the technique, Gouveia et al. [43] and others [39-40] identified that artificially lowering the performance of others, or changing how it is portrayed, can communicate an opportunity for the user to catch up. Colusso et al. [46] empirically assessed its efficacy in the context of game design and found that it increased gamer performance, but only for experienced gamers, who were motivated by comparisons that deemed them closer to the leader’s performance.

Furthermore, social norm nudges and the often-reported *boomerang effect*, can further lead to misfires. The boomerang effect occurs where a person who acts prosocially is led to abandon his previous behaviour when the nudge informs him that the majority does not support his prosocial behaviour. Techniques such as a “*norm from the top*” [47] which provides information based on the most altruistic people in the population, can offer an effective solution. One should, however, consider that misfires can have very different causes, which, most of the time, one cannot anticipate. To effectively identify

and tackle misfires, intervention designers should extensively pilot early prototypes of their nudge and carefully observe and qualitatively inquire into the likely misfires and their causes while abiding to ethical considerations.

6 Conclusions

This theoretical paper aids designers in proactively predicting potential backfires in their nudge design within a given context and reactively uncovering the causes of a nudge with adverse effects. In doing so we distinguish between two main nudge issues; Spillovers and Misfires. More importantly, we provide a clearer distinction the two, enhancing previous conceptualisations that positioned them in a single category. We argue that misfires lead to adverse behaviours because of the intervention itself (e.g. the type of the nudge) while spillovers lead to adverse situations because of both the behaviour induced by the intervention and the intervention type. This allows designers to more accurately blueprint behavioural processes and anticipate points of vulnerability in the choice architecture.

We attempt to bring forth possible explanations to this new taxonomy. We elaborate on a number of causes for Nudge spillovers, such as ego depletion, moral licensing, single-action bias, rest-on-laurels effect, and the reverse foot-in-the-door effect, while for nudge misfires we suggest considering the theory behind each nudge type as a more effective tool of analysis. The latter is especially important as nudges assessed to be ineffective do not mean that they can not be harmful.

References

1. Wall, J., Mhurchu, C.N., Blakely, T., Rodgers, A., Wilton, J.: Effectiveness of monetary incentives in modifying dietary behavior: A review of randomized, controlled trials. *Nutr. Rev.* 64, 518–531 (2006).
2. Kurvers, R.H., Herzog, S.M., Hertwig, R., Krause, J., Carney, P.A., Bogart, A., Argenziano, G., Zalaudek, I., Wolf, M.: Boosting medical diagnostics by pooling independent judgments. In: *Proceedings of the National Academy of Sciences*, 113(31). pp. 8777–8782 (2016).
3. Thaler, R.H., Sunstein, C.R.: *Nudge: Improving Decisions About Wealth, and Happiness*. Penguin Publications (2008).
4. Tversky, A., Kahneman, D.: Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science* (80-). 185, 1124–1131 (1974).
5. Caraban, A., Karapanos, E., Gonçalves, D., Campos, R.: 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In: *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019)*, May 4–9, 2019, Glasgow, Scotland, UK.
6. Broers, V.J., Breucker, C., Van den Broucke, S., Luminet, O.: A systematic review and meta-analysis of the effectiveness of nudging to increase fruit and vegetable choice. *Eur. J. Public Health.* 27, 912–920 (2017).
7. Arno, A., Thomas, S.: The efficacy of nudge theory strategies in influencing adult dietary behavior: a systematic review and meta-analysis. *BMC Public Health.* 16, 1–11 (2016).

8. Loewenstein, G., Sunstein, C.R., Golman, R.: Disclosure: Psychology changes everything. *Annu. Rev. Econom.* 6, 391–419 (2014).
9. Themistocleous, C., Smith, A., Wagner, C.: The ethical dilemma of implicit vs explicit data collection: Examining the factors that influence the voluntary disclosure of information by consumers to commercial organizations. In: *IEEE International Symposium on Ethics in Science, Technology and Engineering* (2014).
10. Mertens, S., Herberz, M., Hahnel, U.J., Brosch, T.: The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. In: *Proceedings of the National Academy of Sciences* (2022).
11. Maier, M., Bartoš, F., Stanley, T.D., Shanks, D.R., Harris, A.J., Wagenmakers, E.J.: No evidence for nudging after adjusting for publication bias. *Proc. Natl. Acad. Sci.* 119, (2022).
12. Sunstein, C.R.: Nudges that fail. *Behav. public policy.* 1, 4–25 (2017).
13. Dolan, O., Galizzi, M.: Like ripples on a pond: Behavioural spillovers and their implications for research and policy. *J. Econ. Psychol.* 47, 1–16 (2015).
14. White, J.B., Langer, E.J., Yariv, L., Welch, J.C.: Frequent social comparisons and destructive emotions and behaviors: The dark side of social comparisons. *J. Adult Dev.* 13, 36–44 (2006).
15. Galizzi, M.M., Whitmarsh, L.: How to measure behavioral spillovers: A methodological review and checklist. *Front. Psychol.* 10, 342 (2019). <https://doi.org/https://doi.org/10.3389/fpsyg.2019.00342>.
16. Themistocleous, C., Karapanos, E.: Precision nudging: The future of behavioural change. In: Christou, P. (ed.) *AI in Social Research*. p. Forthcoming. CABI Publishing (2025).
17. Lanzini, P., Thøgersen, J.: Behavioural spillover in the environmental domain: An intervention study. *J. Environ. Psychol.* 40, 381–390 (2014).
18. Truelove, H.B., Carrico, A.R., Weber, E.U., Raimi, K.T., Vandenbergh, M.P.: Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. *Glob. Environ. Chang.* 29, 127–138 (2014).
19. Nafziger, J.: Spillover effects of nudges. *Econ. Lett.* 190, 109086 (2020). <https://doi.org/https://doi.org/10.1016/j.econlet.2020.109086>.
20. Thøgersen, J.: Spillover processes in the development of a sustainable consumption pattern. *J. Econ. Psychol.* 20, 53–81 (1999).
21. Fleig, L., Lippke, S., Pomp, S., Schwarzer, R.: Intervention effects of exercise self-regulation on physical exercise and eating fruits and vegetables: A longitudinal study in orthopedic and cardiac rehabilitation. *Prev. Med. (Baltim.)* 53, 182–187 (2011).
22. Stibe, A., Cugelman, B.: Persuasive backfiring: When behavior change interventions trigger unintended negative outcomes. In: *International Conference on Persuasive Technology*. pp. 65–77. Charm: Springer International Publishing (2016).
23. Lasarov, W., Hoffmann, S.: Social moral licensing. *J. Bus. Ethics.* 165, 45–66 (2020).
24. Merritt, A.C., Daniel, A.E., Benoit, M.: Moral self-licensing: When being good frees us to be bad. *Soc. Personal. Psychol. Compass.* 4, 344–357 (2010).
25. Khan, U., Dhar, R.: Licensing effect in consumer choice. *J. Mark. Res.* 43, 259–266 (2006).
26. Fishbach, A., Dhar, R.: Goals as excuses or guides: The liberating effect of perceived goal progress on choice. *J. Consum. Res.* 32, 370–377 (2005).
27. Bem, D.J.: Self-perception theory. *Adv. Exp. Soc. Psychol.* 6, (1972).
28. Clot, S., Della Giusta, M., Jewell, S.: Once good, always good? Testing nudge’s spillovers on pro-environmental behavior. *Environ. Behav.* 54, 655–669 (2022).
29. Baumeister, R.E., Bratslavsky, E., Muraven, M., Tice, D.M.: Ego depletion: Is the active self a limited resource? *J. Pers. Soc. Psychol.* 74, 1252–1265 (1998).

30. Muraven, M., Baumeister, R.F.: Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychol. Bull.* 126, 247 (2000).
31. Englert, C., Bertrams, A., Furley, P., Oudejans, R.R.: Is ego depletion associated with increased distractibility? Results from a basketball free throw task. *Psychol. Sport Exerc.* 18, 26–31 (2015).
32. Englert, C., Wolff, W.: Ego depletion and persistent performance in a cycling task. *Int. J. Sport Psychol.* 46, 137–151 (2015).
33. Shiv, B., Fedorikhin, A.: The interplay of affect and cognition in consumer decision making. *J. Consum. Res.* 26, 278–292 (1999).
34. Weber, E.U.: Perception and expectation of climate change. *Psychol. Perspect. to Environ. ethical issues Manag.* 314–341 (1997).
35. Choudhary, G., Dutt, V.: Analyzing single-action bias in dynamic climate change environments: insights from feedback and probability. *Humanit. Soc. Sci. Commun.* 11, 1–15 (2024).
36. Amir, O., Ariely, D.: Resting on laurels: The effects of discrete progress markers as subgoals on task performance and preferences. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 1158–1171 (2008).
37. Kruglanski, A.W., Shah, J.Y., Fishbach, A., Friedman, R., Chun, W., Sleeth-Keppler, D.: A theory of goal-systems. In: Zanna, M.P. (ed.) *Advances in experimental social psychology*. pp. 331–376. New York: Academic Press (2002).
38. Beaman, A.L., Cole, C.M., Preston, M., Klentz, B., Steblay, N.M.: Fifteen years of foot-in-the door research: a meta-analysis. *Personal. Soc. Psychol. Bull.* 9, 181–196 (1983).
39. Guadagno, R.E., Asher, T., Demaine, L.J., Cialdini, R.B.: When saying yes leads to saying no: Preference for consistency and the reverse foot-in-the-door effect. *Personal. Soc. Psychol. Bull.* 27, 859–867 (2001).
40. Dijkstra, P., Gibbons, F.X., Bruunk, A.P.: Social comparison theory. In: Maddux, J.E. and Tangney, J.P. (eds.) *Social Psychological Foundations of Clinical Psychology*. pp. 195–211. The Guildford Press (2010).
41. Suls, J., Martin, R., Wheeler, L.: Social comparison: Why, with whom, and with what effect? *Curr. Dir. Psychol. Sci.* 11, 159–163 (2002).
42. Kankane, S., DiRusso, C., Buckley, C.: Can we nudge users toward better password management? an initial study. In: *Extended abstracts of the 2018 CHI conference on human factors in computing systems*. pp. 1–6 (2018).
43. Gouveia, R., Pereira, F., Karapanos, E., Munson, S.A., Hassenzahl, M.: Exploring the design space of glanceable feedback for physical activity trackers. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. pp. 144–155 (2016).
44. Gerber, J.P., Wheeler, L., Suls, J.: A social comparison theory meta-analysis 60+ years on. *Psychol. Bull.* 144, 177 (2018).
45. Adar, E., Tan, D.S., Teevan, J.: Benevolent deception in human computer interaction. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. pp. 1863–1872 (2013).
46. Colusso, L., Hsieh, G., Munson, S.A.: Designing closeness to increase gamers’ performance. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. pp. 3020–3024 (2016).
47. Rouillé, A.: Norm from the top: a social norm nudge to promote low-practiced behaviors without boomerang effect. *Socioecon. Challenges.* 8, 123–142 (2024).