

PERSPECTIVE

From evidence to delivery: an Implementation-Science blueprint for behavioural policy

Giuseppe Alessandro Veltri^{1,2} 

¹Center for Behavioural and Implementation Science, Yong Loo Lin School of Medicine, National University of Singapore, Singapore and ²Sociology and Social Research, Università di Trento, Trento, Italy
Email: gaveltri@nus.edu.sg

(Received 19 April 2025; revised 12 August 2025; accepted 15 August 2025)

Abstract

The notorious Rossi's 'Iron Law of Evaluation' – that the expected net impact of any large-scale social programme is zero – reminds us that expectations about policy interventions rarely survive real-world delivery. Behavioural Public Policy (BPP) faces many implementation challenges. Implementation Science (IS), which studies how evidence-based practices are adopted, delivered and sustained, offers BPP a toolkit for overcoming the knowledge–action gap. We show how IS frameworks like CFIR (Consolidated Framework for Implementation Research) and RE-AIM (Reach, Effectiveness, Adoption, Implementation, Maintenance) diagnose contextual barriers – leadership, workflow fit, resources – and supply metrics of fidelity, adoption, cost and sustainment. Next, we outline three hybrid trial types from IS that co-test policy impact and implementation: Type 1 emphasises behavioural effects while sampling implementation data; Type 2 balances both; Type 3 optimises implementation while tracking outcomes. Cluster-randomised and stepped-wedge roll-outs create feedback loops that enable mid-course adaptation and speed scale-up. Cases illustrate how spotting delivery slippage early averts costly failure; they reveal how early IS integration can turn isolated behavioural wins into scalable, system-wide transformations that genuinely endure long. We situate these recommendations within the literature on scalability and the 'voltage effect', clarifying how common drops from pilot to scale can be anticipated, diagnosed and mitigated using IS outcomes and process data.

Keywords: Behavioural Public Policy; evidence-based practice; hybrid trials; Implementation Science; intervention sustainability

Introduction

Almost forty years ago, the distinguished sociologist Peter H. Rossi (1987) reflected on the disappointing outcomes of policy evaluations from the 1970s and 1980s, formulating what he termed the 'Iron Law of Evaluation': 'the expected value of any

net impact assessment of any large-scale social program is zero'. Rossi's pessimistic assessment likely stems from inherent difficulties in consistently implementing social programs effectively, challenges equally relevant to contemporary Behavioural Public Policy (BPP) interventions. Indeed, despite significant advances and compelling evidence generated by behavioural science, the field remains confronted with persistent challenges in translating promising pilot results into sustained and widespread practice (Maier *et al.*, 2022; Chater and Loewenstein, 2023; Hallsworth, 2023; Varazzani and Hubble, 2025; Banerjee and John, 2025). Many pressing social, health and economic issues depend not merely on the existence of effective behavioural interventions but critically on the willingness and capacity of individuals, communities and organisations to consistently adopt beneficial behaviours. Nevertheless, the 'last mile' challenge, wherein effective innovations struggle to establish themselves in real-world settings, continues to limit the broader impact and sustainability of behavioural interventions (Fixsen *et al.*, 2005; Proctor *et al.*, 2011).

Implementation Science (IS) explicitly addresses this critical gap by focusing on how evidence-based practices can be systematically embedded into everyday operations. IS emerged precisely from recognising that interventions, even those demonstrably effective under controlled conditions, frequently falter when confronted with organisational inertia, limited resources, competing demands or inadequate stakeholder engagement (Damschroder *et al.*, 2009; Proctor *et al.*, 2009). By identifying and systematically addressing these adoption-influencing factors, IS provides structured frameworks to guide interventions from initial concept through sustainable, large-scale application. While contemporary BPP increasingly attends to organisational and contextual constraints, the distinctive contribution of IS lies in offering purpose-built frameworks (e.g., Consolidated Framework for Implementation Research [CFIR]; Reach, Effectiveness, Adoption, Implementation, Maintenance [RE-AIM]) and validated measures that render that attention systematic, comparable across sites, and decision-relevant; in this sense, IS places equal emphasis on structural and contextual factors – such as leadership commitment, workflow alignment, policy support and training – that are critical for ensuring practices endure beyond pilot phases (Glasgow *et al.*, 1999; Damschroder *et al.*, 2009).

Recognising that Rossi's 'Iron Law of Evaluation' applies equally to behavioural interventions underscores the urgent need for a strategic integration between BPP and IS. Despite their evident complementarity, systematic integration between these fields remains relatively sparse. Many behavioural interventions still depend largely on small-scale demonstrations or localised experimentation, implicitly expecting initial successes to organically expand, an approach that needs to be revised (Almaatouq *et al.*, 2024; List, 2024). This perspective article argues for the intentional incorporation of IS strategies from the earliest stages of BPP intervention design – particularly through hybrid trial methodologies – to significantly enhance the likelihood of behavioural interventions achieving widespread and sustained adoption. By bridging these fields, IS can serve as a critical blueprint for uptake, empowering BPP interventions to overcome Rossi's pessimistic prediction and realise greater scalability and longevity within complex institutional landscapes.

A practical implication is to foreground why effects attenuate at scale – the ‘voltage effect’ – and link candidate drivers of attenuation to concrete IS metrics. Dilution of non-negotiable components maps to fidelity indicators; diminished demand at scale maps to reach and adoption; and operational drift over time maps to maintenance. Designing studies with these links pre-specified enables earlier diagnosis and course correction (Glasgow *et al.*, 1999; Proctor *et al.*, 2009; List, 2024).

Why IS matters for BPP

BPP interventions, whether they involve rearranging choice architectures or offering social norm feedback, are grounded in the notion that human behaviour is malleable and strongly shaped by environment and context (Michie *et al.*, 2011, 2014). Yet in practice, these contexts are not simply neutral backdrops; they are often highly structured settings (hospitals, municipal governments, workplaces) where entrenched cultures, resource constraints or political pressures can stifle innovation. IS supplies the tools to identify these systemic forces early in the design process. For instance, the CFIR details constructs like leadership engagement, available resources, organisational culture, and external policies, each of which can make or break a promising intervention (Damschroder *et al.*, 2009). By systematically assessing these factors at the outset, a BPP project can anticipate roadblocks and structure interventions that mesh with real-world constraints (Table 1).

IS also emphasises the concept of fidelity, defined as how faithfully an intervention is delivered according to its core components. Even if a BPP solution is theoretically sound, local adaptations or partial implementation can dilute essential elements, leading to weaker outcomes or null results. By measuring fidelity – along with other

Table 1. Key dimensions of Implementation Science and their relevance to BPP

IS dimension	Typical focus in IS	Application for BPP
Context and setting	Organisational readiness, leadership engagement, resource availability	Identifies whether a nudging intervention or incentive scheme aligns with local policies, budgets and administrative support
Fidelity and adaptation	Degree to which an intervention is implemented as intended vs local adaptations	Tracks how a BPP strategy (e.g., default options and social norm prompts) might be unintentionally diluted or modified during scaling
Stakeholder engagement	Inclusion of frontline staff, management, and beneficiaries in design & rollout	Ensures BPP solutions reflect the experiences of those tasked with applying new procedures or abiding by new rules
Sustainment	Processes to maintain and institutionalise an intervention beyond pilot phases	Examines how to embed BPP elements (e.g. new workflows or incentive structures) into ongoing organisational routines and policy frameworks
Implementation outcomes	Adoption, feasibility, acceptability, cost, maintenance	Provides metrics that gauge both short-term policy uptake and longer-term viability, supporting agile course corrections

implementation-related outcomes such as adoption, feasibility, acceptability and sustainability – IS offers an empirical feedback loop. This loop ensures that any observed effects (positive or negative) can be tied to how well the intervention was delivered, rather than leaving questions about whether the original protocol was substantially altered in the course of real-world deployment.

Moreover, IS underscores the significance of multi-stakeholder engagement. BPP innovations can benefit from the bottom-up perspectives of those who must implement the intervention daily (e.g., frontline staff, administrators and community leaders). If these individuals are sidelined during design and deployment, they may be less motivated to support the initiative, risking a superficial or reluctant adoption. IS frameworks like RE-AIM highlight the value of early and ongoing engagement with all relevant parties, ensuring that each stakeholder's readiness, resources, and concerns are integrated into policy rollout (Glasgow *et al.*, 1999). To be clear, many BPP teams already co-design with practitioners and communities; the value-add of IS lies in codifying these activities into replicable strategies and measures so that engagement is not episodic but built into the study design, data collection plan and decision rules.

Terminology and scope. We align terminology with canonical sources: 'classic nudges' are changes to choice architecture that alter behaviour predictably without restricting options or significantly changing incentives (Thaler and Sunstein, 2008); 'nudge plus' couples a nudge with a prompt for deliberation or reflection (Banerjee and John, 2024); and 'boosts' aim to build decision competences (e.g., risk literacy) that persist and can generalise (Hertwig *et al.*, 2025). This shared vocabulary helps map intervention classes to IS supports and outcomes.

IS can – and should be calibrated to the depth of reflection each behavioural technique demands. For classic nudges (Thaler and Sunstein, 2008), which rely on subtle choice architecture changes (e.g., defaults and salience cues), IS prioritises low-friction fidelity checklists, rapid audit-and-feedback cycles and the tracking of contextual moderators (leadership buy-in, workflow fit) that might silently erode the nudge's potency at scale. Where designers add an element of deliberation – so-called nudge plus (Banerjee and John, 2024), in which recipients are prompted to pause and reflect – IS highlights additional implementation ingredients: staff training to facilitate reflection, materials that scaffold self-explanation, and pragmatic measures of dose received (did participants actually engage in the reflective step?). Finally, boosts aim to build enduring decision competences (risk literacy, future cost visualisation) rather than steer one-off choices (Hertwig *et al.*, 2025). Here, IS shifts emphasis toward capacity-building strategies (e.g., train-the-trainer models and refresher sessions), longitudinal sustainment metrics, and equity audits that verify whether competence gains are maintained across heterogeneous groups. Across all three classes, IS offers common scaffolds – CFIR diagnostics for context mapping, RE-AIM indicators for reach and maintenance – but applies them with differing intensity: lightweight and rapid for nudges, integrated with reflective content for nudge plus, and resource-intensive, educational, and longitudinal for boosts. Embedding these calibrated IS supports from the outset increases the likelihood that each intervention type not only works in principle but also survives the messy realities of routine delivery. Importantly, even 'light-touch' nudges often require substantial enabling infrastructure – IT changes to defaults, legal review, staff training and data governance – to achieve and sustain effect, so their simplicity at the

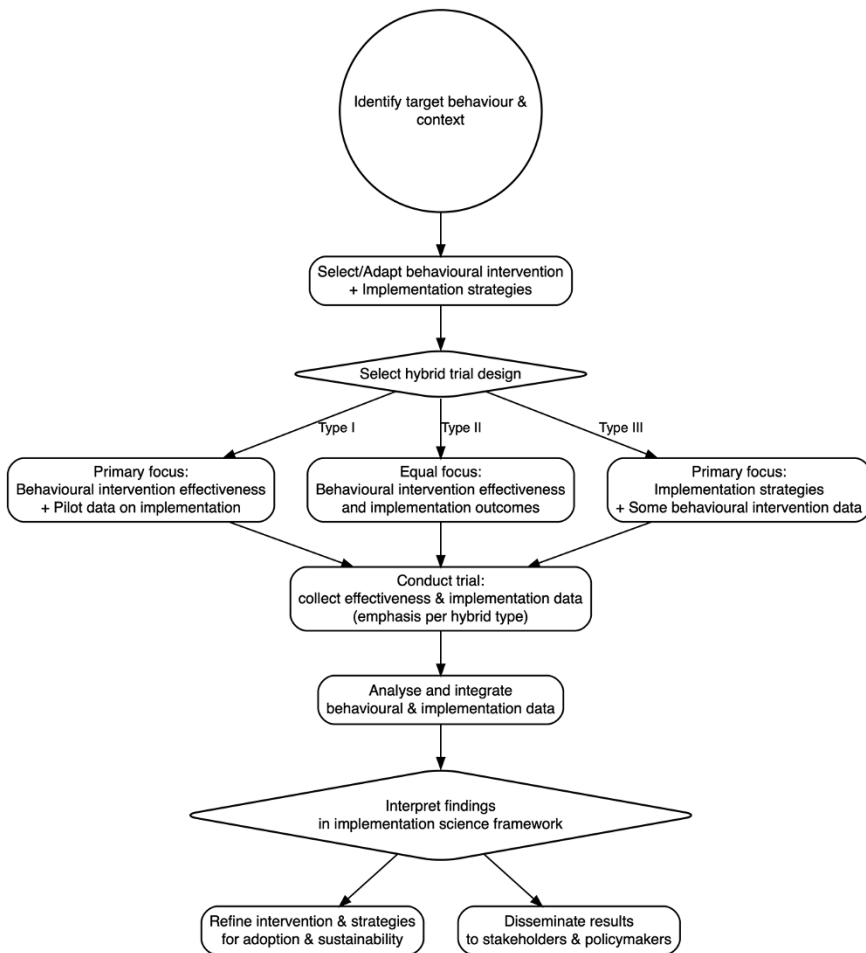


Figure 1. Conceptual depiction of how hybrid trial designs integrate effectiveness and implementation research. Illustrative Type 2 pathway: a municipal energy default intervention compares two implementation strategies – (A) IT auto-switch plus staff training versus (B) IT auto-switch plus targeted public messaging – while concurrently tracking fidelity (default configuration delivered as specified), adoption (households enrolled), cost (IT time, staff hours, media spend), and maintenance (retention at 6–12 months).

user interface should not be conflated with ease of organisational delivery (cf. CFIR constructs of resources, leadership engagement and workflow fit; Damschroder *et al.*, 2009).

A practical tension concerns fidelity versus adaptation. We distinguish core components – non-negotiable elements tied to the causal theory – from an adaptable periphery that can be tailored to local context without eroding mechanism integrity (Carroll *et al.*, 2007). Using adaptation logs and reporting frameworks such as FRAME (Stirman *et al.*, 2013), teams can document what was modified, why, and with what

Table 2. Key characteristics of hybrid trial types and their potential applications in Behavioural Public Policy

Hybrid type	Primary focus	Secondary focus	Potential use in BPP
Type 1	Efficacy/effectiveness: Demonstrate that the intervention works (e.g., does it increase vaccination rates?).	Preliminary data on implementation outcomes: Collect early indicators of feasibility, fidelity, and adoption.	Useful when the evidence base for the behavioural intervention is still emerging but there is interest in early insights into barriers/facilitators of real-world uptake.
Type 2	Dual focus on both effectiveness and implementation: Assess how well the intervention works while concurrently testing implementation strategies.	Balanced approach to both ‘does it work?’ and ‘how can it be delivered effectively?’.	Ideal for BPP when policymakers want comprehensive insights, including contextual factors (e.g., organisational culture, leadership engagement) that shape behaviour change in practice.
Type 3	Implementation: Determine the best strategy for embedding the intervention into usual care or real-world systems.	Continued measurement of effectiveness outcomes to confirm ongoing impact.	Optimal when the intervention’s efficacy is already well-established, but scalable, context-specific solutions are needed for routine operation (e.g., varying local incentives and training protocols).

consequences. In BPP, this means specifying which elements of a default change are fixed (e.g., enrolment logic) and which can vary (message channel, timing), and then measuring both fidelity and adaptations to explain outcomes.

Figure 2 summarises these choices in a practical decision tree that links intervention class and delivery footprint to IS intensity and a matching hybrid design.

Hybrid trial methodologies as a convergence strategy

One especially powerful strategy for integrating IS and BPP is the use of ‘hybrid’ trial designs (Curran *et al.*, 2012) in one unified study (see Figure 1). Unlike traditional research approaches that first conduct a standalone effectiveness trial and then launch a separate implementation trial, hybrid designs systematically evaluate both the impact of an intervention and the processes underpinning its delivery in one unified study. By gathering data on both the primary policy outcomes of interest (e.g., increases in vaccination uptake and reductions in energy consumption) and key implementation factors (e.g., fidelity, adoption, cost, organisational readiness and stakeholder engagement), hybrid trials can identify in real time whether a disappointing result arises from limitations in the policy itself or from suboptimal implementation strategies. Hybrid trial methodologies are frequently categorised into three main types, see Table 2 and Figure 2, each striking a different balance between testing effectiveness and examining implementation (Curran *et al.*, 2012). Type 1 prioritises assessing the policy’s effectiveness while collecting preliminary data on key implementation outcomes. Type 2 conducts a more equal exploration of both the intervention’s effectiveness and the

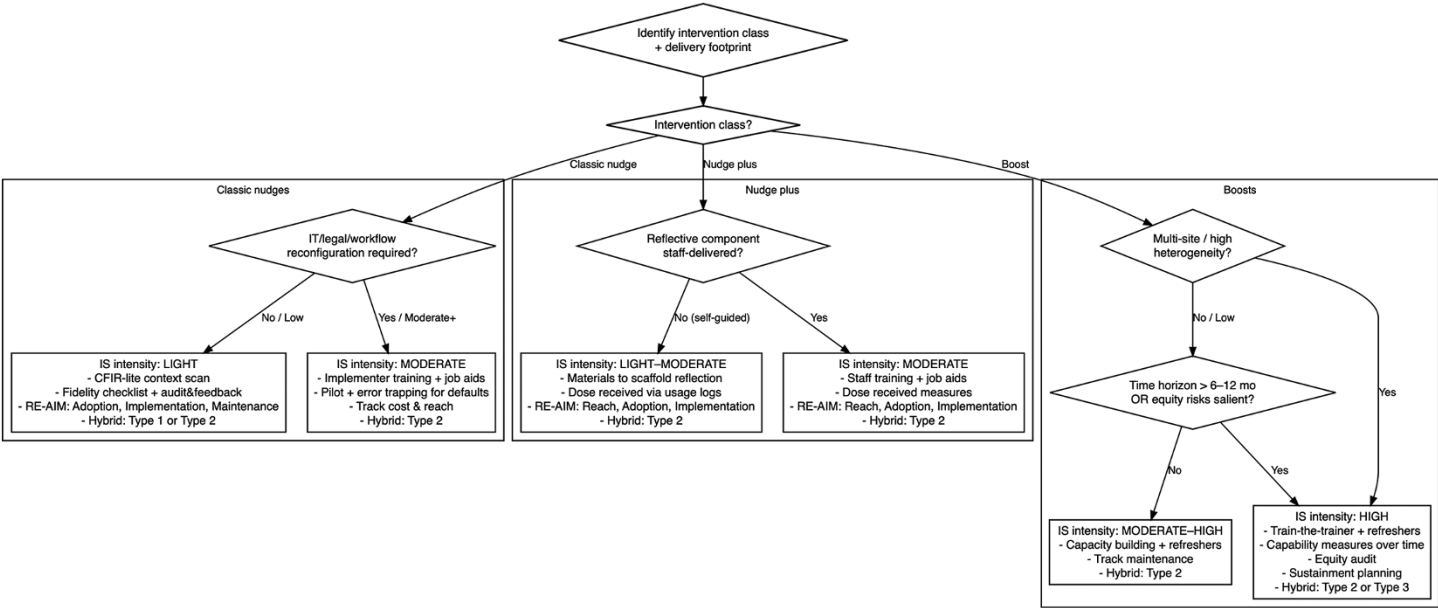


Figure 2. Decision tree for calibrating Implementation Science (IS) supports by intervention type and delivery footprint. The branches map intervention class (classic nudges, nudge plus, boosts) and delivery considerations (e.g., IT/legal/workflow footprint; reflective component; heterogeneity, time horizon, equity risk) to a recommended IS intensity and a suitable hybrid trial type.

implementation strategies, allowing researchers to draw robust conclusions about how well the intervention works in real-world settings and why. Type 3 flips the emphasis toward implementation, rigorously investigating how best to embed the intervention in practice while still tracking its impact on the desired outcomes. In BPP, Type 2 designs are well-suited to experimentally compare allowable adaptations (e.g., alternative message channels or training intensity) while tracking fidelity, cost, and equity, thereby turning the fidelity–adaptation tension into a testable design feature rather than an afterthought.

In practice, hybrid trials employ a range of research designs and data-collection methods to capture the nuance of both policy effects and implementation quality. For example, cluster-randomised (Hemming and Taljaard, 2023) or stepped-wedge designs (Barker *et al.*, 2016) are often used when interventions are rolled out in multiple sites or over multiple time points, enabling researchers to compare different implementation strategies in real-world settings. These approaches are particularly advantageous for BPP because they account for local variations in leadership support, resource availability, staff training, and incentives – factors that can dramatically alter the success or failure of an intervention. Data collection typically involves standardised behavioural metrics (e.g., administrative data on service uptake or consumption), fidelity checklists (documenting whether and how various components of the intervention are delivered), cost analyses (capturing direct and indirect expenses associated with implementation), and stakeholder feedback (gauging buy-in and perceived feasibility). To manage this diversity of data, hybrid trials increasingly utilise multilevel models and related techniques to parse how context, fidelity, organisational culture, and leadership engagement mediate or moderate policy outcomes.

Another advantage of hybrid methodologies is their adaptive, iterative capacity. Because outcomes and implementation processes are tracked concurrently, teams can quickly detect and diagnose issues such as poor alignment with local workflows, insufficient staff training, or a mismatch in incentive structures. Rather than waiting until the trial concludes to make mid-course corrections, they can refine training protocols, adjust incentives, or address leadership engagement challenges as soon as they surface. This ensures that any observed effect (or lack thereof) is not merely a function of inadequate implementation. Continuous stakeholder involvement – from front-line staff to community representatives – helps refine the intervention to better match the evolving needs of the target population. In this way, hybrid trials can accelerate the transition from pilot projects to broader policy rollouts, avoiding scenarios where a policy that succeeded in a controlled demonstration fails to maintain its effectiveness in day-to-day operation.

Overall, by weaving together direct measures of policy impact with a rigorous exploration of the contexts and processes that shape real-world delivery, hybrid trials maximise both internal and external validity. They not only clarify ‘what works’ in terms of behaviour change, but also illuminate the ‘why’ and ‘how’ behind successful (or unsuccessful) implementation. This holistic perspective underscores the importance of systematically integrating IS and BPP in a single methodological framework – one that is well-suited to dynamic policy arenas where flexibility, scalability, and contextual fit are paramount.

The ASPIRE (Beidas *et al.*, 2021) and 3HP Options trials (Kadota *et al.*, 2020) provide examples of Type 3 hybrid studies that leverage behavioural science to design and test implementation strategies for established clinical interventions. The ASPIRE trial focused on implementing the S.A.F.E. Firearm safety programme in paediatric primary care by comparing a behavioural science-informed nudge strategy (an Electronic Health Record prompt) against nudge plus practice facilitation, with clinician fidelity as the primary outcome. The 3HP Options trial aimed to optimise delivery of a 12-week tuberculosis preventive therapy for people living with HIV in Uganda. It compared facilitated Directly Observed Therapy (DOT), facilitated Self-Administered Therapy (SAT), and Informed Patient Choice, using COM-B and the Behaviour Change Wheel to design strategies that address capability, opportunity, and motivation. These cases highlight distinct applications of behavioural science – ASPIRE’s targeted nudge versus 3HP’s comprehensive, theory-driven approach – both within a Type 3 framework prioritising implementation effectiveness.

Finally, hybrid trials complement rather than displace theory-rich approaches such as realist evaluation and theory-based evaluation. Realist logic models (mechanism–context – outcome configurations) help specify where and for whom mechanisms should fire (Pawson and Tilley, 1997), while theory-based evaluation clarifies testable causal pathways linking components to outcomes (Weiss, 1997). Embedding their propositions within hybrid designs strengthens interpretation of both effectiveness and implementation findings.

Key challenges and implications

Despite the apparent synergy, merging IS with BPP is not without complications. One challenge lies in reconciling different terminologies and disciplinary norms. BPP practitioners may use terms like ‘choice architecture’ or ‘boosts’, while IS researchers focus on ‘outer setting’, ‘implementation strategy’ or ‘fidelity measures’. Clarity and alignment can be cultivated through early project discussions and shared training opportunities, ensuring that each side understands the other’s lexicon and goals. Empirically, efforts that surface and align assumptions early – via structured context assessments and strategy specification – tend to avoid later drift and rework (Damschroder *et al.*, 2009; Powell *et al.*, 2015).

Resource constraints also loom large. Hybrid trials demand thorough planning, mixed methods data collection, and often longer timelines to capture sustainability. BPP projects, particularly those within government settings, may face tighter budget cycles or political pressures for quick results. To mitigate these pressures, it is advisable to concentrate on a handful of critical implementation metrics – like fidelity, acceptability and cost – rather than attempting to measure every dimension of IS theory in a single project. Using digital platforms for data collection and focusing on feasible, pragmatic outcome measures can also lessen participant burden. Evidence from service-delivery settings shows that pragmatic fidelity tools, coupled with light-touch audit-and-feedback, can improve uptake while containing measurement burden (Carroll *et al.*, 2007; Powell *et al.*, 2015).

Another issue is the inherently context-specific nature of BPP. A social norm campaign that works in one region may fail elsewhere due to cultural variations, differing

levels of trust in public institutions, or other structural disparities. IS frameworks offer systematic ways to document contextual differences, such as leadership involvement or resource availability, but BPP practitioners need to remain open to adaptation and iteration. Hybrid trials, by design, facilitate ongoing evaluation and provide evidence on which modifications are beneficial without drifting from core components of the intervention. Synthesis work on the diffusion and scale-up of complex interventions underscores that local history, professional norms, and inter-organisational networks often shape outcomes as much as formal design features (Greenhalgh *et al.*, 2004).

Finally, long-term sustainability demands embedding changes into policy or organisational structures. While some BPP approaches – like nudges – are low-cost and minimally invasive, they may quickly lose impact if supporting systems (training, incentives, feedback loops) are not in place. IS underscores the role of institutional champions, policy alignment, and ongoing accountability to anchor new practices. Policymakers can be encouraged to adopt formal mechanisms – such as integration into routine funding lines or performance measures – that make the intervention a standard part of how agencies and organisations operate. In addition, implementation rarely unfolds in a political vacuum. Power dynamics, shifting agendas, and policy windows shape feasibility and timing. Simple power-mapping, the cultivation of champions, and sensitivity to agenda-setting processes can accelerate adoption and sustainment (Kingdon, 2014; Cairney, 2016).

Conclusions

Although the central aim here is to show how IS can strengthen BPP, the synergy can flow in the opposite direction as well (Hodson *et al.*, 2024). Incorporating behaviourally informed principles into IS practice can help address common challenges like stakeholder resistance, cognitive biases within leadership teams, or inertia in bureaucratic systems. For example, defaults and friction can be used to simplify new workflows for implementers and remove rarely used but error-prone options; timely prompts, commitment devices, and salient feedback can support leaders' regular review of implementation dashboards and follow-through on agreed actions (Thaler and Sunstein, 2008; Hodson *et al.*, 2024). Moving forward, more BPP programmes could be designed with IS principles baked in from the start: identifying feasible fidelity measures, clarifying roles and responsibilities, and building adaptive protocols that allow for site-specific tailoring. Policy teams might also consider stepped-wedge trial designs that roll out an intervention incrementally across jurisdictions, tracking both outcomes and implementation processes at each step to refine the approach. Such designs, already familiar in IS, would enhance the evidence base for BPP interventions by revealing how organisational and community variables moderate a policy's effectiveness. Ultimately, by merging IS's systemic lens with BPP's nuanced understanding of behaviour, policymakers will be better equipped to mount evidence-informed interventions that do not merely spark short-term improvements but also create lasting transformations. Hybrid trials, iterative feedback loops, fidelity checks and a stronger emphasis on multi-level stakeholder engagement can make the difference between an intervention that flourishes beyond its initial grant period and one that quietly

fades. Embracing these combined methods and perspectives is the next logical step for advancing the science and practice of BPP. Framing studies explicitly around likely voltage-loss mechanisms, and mapping them to IS outcomes, can further increase the odds that effects persist at scale (List, 2024).

Acknowledgements. The author is grateful to Nick Sevdalis, Ioannis Bakoulis and Ivo Vlaev for the insightful discussions about the interplay between behavioural and implementation sciences.

References

- Almaatouq, A., T. L. Griffiths, J. W. Suchow, M. E. Whiting, J. Evans and D. J. Watts (2024), 'Beyond playing 20 questions with nature: integrative experiment design in the social and behavioral sciences', *Behavioral and Brain Sciences*, **47**: e33.
- Banerjee, S. and P. John (2024), 'Nudge plus: incorporating reflection into behavioral public policy', *Behavioural Public Policy*, **8**(1): 69–84.
- Banerjee, S. and P. John (2025), Behavioral public policy: past, present, & future. *Policy and Society*. Retrieved 7 August 2025, from <https://dx.doi.org/10.1093/polsoc/puaf012>
- Barker, D., P. McElduff, C. D'Este and M. Campbell (2016), 'Stepped wedge cluster randomised trials: a review of the statistical methodology used and available', *BMC Medical Research Methodology*, **16**: 1–19.
- Beidas, R. S., B. K. Ahmedani, K. A. Linn, S. C. Marcus, C. Johnson, M. Maye *et al.* (2021), 'Study protocol for a type III hybrid effectiveness–implementation trial of strategies to implement firearm safety promotion as a universal suicide prevention strategy in pediatric primary care', *Implementation Science*, **16**: 1–16.
- Cairney, P. (2016), *The Politics of Evidence-Based Policy Making*, London: Palgrave Macmillan.
- Carroll, C., M. Patterson, S. Wood, A. Booth, J. Rick and S. Balain (2007), 'A conceptual framework for implementation fidelity', *Implementation Science*, **2**: 40.
- Chater, N. and G. Loewenstein (2023), 'The i-frame and the s-frame: how focusing on individual-level solutions has led behavioral public policy astray', *Behavioral and Brain Sciences*, **46**: e147.
- Curran, G. M., M. Bauer, B. Mittman, J. M. Pyne and C. Stetler (2012), 'Effectiveness– implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact', *Medical Care*, **50**(3): 217–226.
- Damschroder, L. J., D. C. Aron, R. E. Keith, S. R. Kirsh, J. A. Alexander and J. C. Lowery (2009), 'Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science', *Implementation Science*, **4**: 50.
- Fixsen, D. L., S. F. Naoom, K. A. Blase, R. M. Friedman and R. Wallace (2005), *Implementation Research: A Synthesis of the Literature*, Tampa, FL: University of South Florida, National Implementation Research Network.
- Glasgow, R. E., T. M. Vogt and S. M. Boles (1999), 'Evaluating the public health impact of health promotion interventions: the RE-AIM framework', *American Journal of Public Health*, **89**(9): 1322–1327.
- Greenhalgh, T., G. Robert, F. Macfarlane, P. Bate and O. Kyriakidou (2004), 'Diffusion of innovations in service organizations: systematic review and recommendations', *The Milbank Quarterly*, **82**(4): 581–629.
- Hallsworth, M. (2023), 'A manifesto for applying behavioural science', *Nature Human Behaviour*, **7**(3): 310–322.
- Hemming, K. and M. Taljaard (2023), 'Key considerations for designing, conducting and analysing a cluster randomized trial', *International Journal of Epidemiology*, **52**(5): 1648–1658.
- Hertwig, R., S. Michie, R. West and S. Reicher (2025), 'Moving from nudging to boosting: empowering behaviour change to address global challenges', *Behavioural Public Policy*, 1–12.
- Hodson, N., B. J. Powell, P. Nilsen and R. S. Beidas (2024), 'How can a behavioral economics lens contribute to implementation science?', *Implementation Science*, **19**(1): 33.
- Kadota, J. L., A. Musinguzi, J. Nabunje, F. Welishe, J. L. Ssemata, O. Bishop *et al.* (2020), 'Protocol for the 3HP Options trial: a hybrid Type 3 implementation–effectiveness randomized trial of delivery strategies for short-course tuberculosis preventive therapy among people living with HIV in Uganda', *Implementation Science*, **15**: 1–12.

- Kingdon, J. W. (2014), *Agendas, Alternatives, and Public Policies*, 2nd edn, Updated. Harlow: Pearson Education Limited.
- List, J. A. (2024), 'Optimally generate policy-based evidence before scaling', *Nature*, **626**(7999): 491–499.
- Maier, M., F. Barto's, T. Stanley, D. R. Shanks, A. J. L. Harris and E.-J. Wagenmakers (2022), 'No evidence for nudging after adjusting for publication bias', *Proceedings of the National Academy of Sciences*, **119**(31): e2200300119.
- Michie, S., L. Atkins and R. West (2014), *The Behaviour Change Wheel: A Guide to Designing Interventions*, London: Silverback Publishing.
- Michie, S., M. M. van Stralen and R. West (2011), 'The Behaviour Change Wheel: a new method for characterising and designing behaviour change interventions', *Implementation Science*, **6**: 42.
- Pawson, R. and N. Tilley (1997), *Realistic Evaluation*, London: Sage.
- Powell, B. J., T. J. Waltz, M. J. Chinman, L. J. Damschroder, J. L. Smith, M. M. Matthieu *et al.* (2015), 'A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project', *Implementation Science*, **10**: 21.
- Proctor, E. K., J. Landsverk, G. Aarons, D. Chambers, C. Glisson and B. Mittman (2011), 'Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges', *Administration and Policy in Mental Health and Mental Health Services Research*, **38**(1): 24–34.
- Proctor, E., H. Silmere, R. Raghavan, P. Hovmand, G. Aarons, A. Bunger *et al.* (2009), 'Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda', *Administration and Policy in Mental Health and Mental Health Services Research*, **36**(2): 65–76.
- Rossi, P. (1987), 'The iron law of evaluation and other metallic rules', *Research in Social Problems and Public Policy*, **4**(1): 3–20.
- Stirman, S. W., C. J. Miller, K. Toder and A. Calloway (2013), 'The FRAME: a framework for reporting adaptations and modifications to evidence-based interventions', *Implementation Science*, **8**: 65.
- Thaler, R. H. and C. R. Sunstein (2008), *Nudge: Improving Decisions about Health, Wealth, and Happiness*, New Haven, CT: Yale University Press.
- Varazzani, C., and C. Hubble (2025), 'Four sins in behavioural public policy', *Behavioural Public Policy*, **9**, 345–353. <https://doi.org/10.1017/bpp.2024.22>.
- Weiss, C. H. (1997), 'Theory-based evaluation: past, present, and future', *New Directions for Evaluation*, **76**: 41–55.