*Behavioural Public Policy* (2025), **9**, 354–357
doi:10.1017/bpp.2024.23

CAMBRIDGE UNIVERSITY PRESS

RESPONSE

# The limits of self-reflection

Michael Hallsworth[1,2]

[1]The Behavioural Insights Team and [2]University of Pennsylvania, Philadelphia, PA, USA
Email: michael.hallsworth@bi.team

(Received 22 March 2024; accepted 25 March 2024)

I'm grateful to the authors for taking the time to read and respond to the manifesto. I have great respect for the contributions each of them has made to our field, and I'd like to offer some clarifications and reflections prompted by their thoughts.

The manifesto sets itself a difficult challenge. There is the baggage that comes with the word 'manifesto' – although I was thinking mainly of the 'manifesto for reproducible science', also published in *Nature Human Behaviour* (Munafò *et al.*, 2017). Then there is the issue of scope. If you want to comment on applied behavioural science as a whole, you must range fairly widely. At the same time, you need to draw the line somewhere: a manifesto is not a book. This situation invites the criticism that any particular point or perspective was not excluded.

There's some of this kind of criticism in Peter John's response, although I think his overall point is completely fair. I also find cultural theory a useful framework. In fact, I have a forthcoming article that also uses cultural theory to reconsider the debate between paternalist and anti-paternalist readings of behavioural public policy (Hallsworth, in press).

More specifically, I think something is promising in the idea of 'clumsy solutions', which recognise the need to maintain a balance between the four different perspectives (Verweij and Thompson, 2006). They recognise that the fatalist, hierarchical, egalitarian and individualist worldviews all add something, and successful policies do not let a single one dominate. I see these clumsy solutions as the policy equivalent of Hood's 'public management for all the seasons'. But, unlike Hood, their proponents see this instability as a strength, not a weakness.

John himself talks about the advantages of an 'all-seasonal approach'. I do feel that the manifesto invests a lot of time in how to move to a more enabling, distributed approach to applied behavioural science, but acknowledge that more could still be done.

Yet I think this point also raises a wider issue. John writes as if my sole focus was on behavioural public policy. In that vein, he ventures that the manifesto 'does not say much about the organisation of BPP, such as who decides these policies'. Varazzani and Hubble similarly say that the manifesto 'underplays the external factors shaping the policy landscapes within which BPP practitioners operate'.

However, I was very conscious when writing the manifesto that the applied behavioural science movement is not confined to the public sector and, therefore, the manifesto shouldn't be either. Moreover, as is well known, the public sector itself has become more complex and distributed into networks and communities of practice (Rhodes, 2006). The spin-out of The Behavioral Insights Team itself, as a private sector organisation partly owned by the UK government, is a good case in point.

Being aware of the sheer geographic and institutional diversity of the applied behavioural science field made me wary of generalising about how money and power function in it. I do not think it's true that applied behavioural science emerges solely from bureaucratic, hierarchical and political structures in the way that John posits. The players in applied behavioural science are not only 'bureaucracies and political principals'. They may be, for example, philanthropies, banks or integrated health systems. I try to present them this way:

> Figure 1 also indicates how responsibilities for implementing the proposals are allocated among four major groups in the behavioural science ecosystem: practitioners (individuals or teams who apply behavioural science findings in practical settings), the clients who commission these practitioners (for example, public or private sector organisations), academics working in the behavioural sciences (including disciplines such as anthropology, economics and sociology) and funders who support the work of these academics.

Figure 1 is the main diagram that brings the paper together. These players are not absent. But what I didn't feel confident of was making generalisations about the various situations and motivations of these funders and commissioners.

John, Varazzani and Hubble raise an interesting point, though. The manifesto does touch on the power structures of applied behavioural science at some points. The longer version of the report talks about how funders need to bear increased scrutiny and cannot see themselves as separate from the work they commission. The article discusses a future 'where policy or product designers act less like (choice) architects and more like facilitators, brokers and partnership builders'. These aspects gesture at a discussion of agency and institutional change, as some of the responders recommend.

Yet perhaps they don't go far enough. I felt that discussing broader changes to the structures of the public sector was just out of scope for an already long article. We can disagree on the merits of that choice. But, on deeper reflection, I think my decision also reflects a deeper set of incentives at play.

Rather than straddling hierarchical and egalitarian worlds, the manifesto is more trying to have both an academic and practitioner perspective. Achieving that dual perspective can be hard. Some of what John, Varazzani and Hubble recommend is the work of the 'oppositional' critic to use Edward Said's formulation (Said, 1975). This is someone who can look from the outside and has the freedom to tease apart ideology and existing power structures.

Instead, the manifesto starts from the inside, with practitioners, and tries to engender self-reflective practice. As I say, 'a main priority for behavioural scientists is to recognize the various ways that their own behaviour is being shaped by structural, institutional, environmental and cognitive factors'. This process goes to where

https://doi.org/10.1017/bpp.2024.23 Published online by Cambridge University Press

practitioners are and helps them to build something better. That is why I also tried to push on practical recommendations, since I've always found those to be harder to generate than theoretical inquisitions.

To my mind, this combination of critical awareness and practical proposals is productive. It means that you do not simply offer upgraded tools but also encourage changes to the scope and values of behavioural science.

Yet I also understand how this approach can be seen as limited. It fits neatly in the technocratic view of applied behavioural science as an approach that attempts to use evidence to find better ways of achieving goals. That technocracy can encourage a sphere of 'experts speaking to experts', while fundamental issues of politics and power get played out in parallel elsewhere.

I have written before about the dangers of a purely technical approach to policy-making (Hallsworth, 2011). I don't think the manifesto eschews political issues entirely. But I can see that, in order to grapple with the fundamental ideologies, power structures and resource allocations that shape important aspects of applied behavioural science, you would need a more wide-ranging set of critical tools than self-reflective pragmatism. I welcome someone else doing that. I suspect that the two modes would have clashed badly in a single article.

I agree with Malte Dold's core recommendation:

> Behavioural science can and should do more to study socially and culturally embedded processes of goal formation and the type of institutional conditions, including the *material resources* and *civic freedoms* needed for people to form their goals in a competent and self-determined way.

He says that, while I'm aware of this line of thinking, it could be pursued more thoroughly. I think he's right: there is a wider space to help people to 'become aware of the situational, social, and cultural influences on their goal formation processes'. Having said that, I also think we need to be sensitive to how much time and resources people say they want to spend achieving this goal.

Varazzani and Hubble's SINS offer a welcome range of ways to take things further. They cover the need for a broader range of methodologies, organisational change strategy and leadership. I look forward to seeing the OECD's proposals on 'fully embedding behavioural science insights and methods into policy making practices'. I admit that there's a need to be more specific and detailed than the manifesto was in this area.

I do want to clarify one potential confusion in Varazzani and Hubble's article. I didn't propose resource rationality as the definite solution to the gap in theory. Instead, I included it as an example of something closer to a solution, to make the possibilities real to the reader. And I think it does a decent job of being both explanatory and predictive. I take their call for causal models as being similar to the plea for 'strong theories' mentioned in the manifesto. I don't think these calls are wrong; I just think we could spend a lot of time waiting for them to be answered.

When you are writing about the future, there's the risk it arrives quickly. The recent advances in artificial intelligence (AI), particularly large language models, were made public while the manifesto was in peer review. Obviously, they create both opportunities and risks.

https://doi.org/10.1017/bpp.2024.23 Published online by Cambridge University Press

Most obviously, AI can help deliver some of the ambitions in the manifesto. Stuart Mills, Samuel Costa and Cass Sunstein argue convincingly that AI could help us to 'see the system' better and identify promising leverage points (Mills *et al.*, 2023). Perhaps the most exciting opportunities come in the democratisation of behavioural science. Most obviously, an AI-enabled application could scan the extant behavioural science literature and provide a tailored guide for you to achieve your personal goals. This capacity could be transformative for the proposal to 'be humble, explore, and enable'.

On the other hand, the risks raised by AI seem like a sharper and more troubling version of the concerns raised in the 'data science for equity' proposal. AI-based behavioural interventions make the existing concerns around privacy, data collection and acceptability even more pressing (Hagendorff, 2022). Most obviously, they also supercharge the concerns about the sophisticated targeting of people in order to harm them (Bar-Gill *et al.*, 2023).

Since we can rarely predict the future, maybe the best we can hope for is to offer a guide for interpreting it when it arrives. In that sense, I feel like the manifesto passes the test in terms of AI. The more difficult part, as the responders indicate, is to identify how to shape this future ourselves, as behavioural science practitioners and observers. The manifesto's reception so far has left me hopeful that it can play a part in that task – but the hard work is making the journey, not drawing the map.

## References

Bar-Gill, O., C. R. Sunstein and I. Talgam-Cohen (2023), 'Algorithmic harm in consumer markets', *Journal of Legal Analysis*, **15**(1): 1–47.

Hagendorff, T. (2022), 'Blind spots in AI ethics', *AI and Ethics*, **2**(4): 851–867.

Hallsworth, M. (2011), *Policy Making in the Real World*. London: Institute for Government.

Hallsworth, M. (in press), 'Paternalism and behavioral public policy: how do practices diverge from academic debates?', *Social Philosophy and Policy*.

Mills, S., S. Costa and C. R. Sunstein (2023), 'AI, behavioural science, and consumer welfare', *Journal of Consumer Policy*, **46**(3): 387–400.

Munafò, M. R., B. Nosek, D. V. Bishop, K. S. Button, C. D. Chambers, N. Percie du Sert, U. Simonsohn, E. J. Wagenmakers, J. J. Ware and J. Ioannidis (2017), 'A manifesto for reproducible science', *Nature Human Behaviour*, **1**(1): 1–9.

Rhodes, R. A. W. (2006), 'Policy Networks', in M. Moran, M. Rein and R. E. Goodwin (eds.). *The Oxford Handbook of Public Policy*, Oxford: Oxford University Press, 425–447.

Said, E. (1975), *The World, The Text and the Critic*, Cambridge, MA: Harvard University Press.

Verweij, M. and M. Thompson (eds) (2006), *Clumsy Solutions for a Complex World: Governance, Politics and Plural Perceptions*, London: Springer.

**Cite this article:** Hallsworth, M. (2025), 'The limits of self-reflection', *Behavioural Public Policy* **9**, 354–357. https://doi.org/10.1017/bpp.2024.23