nature human behaviour

Perspective

https://doi.org/10.1038/s41562-023-01555-3

A manifesto for applying behavioural science

Received: 26 September 2022

Accepted: 10 February 2023

Published online: 20 March 2023

Check for updates

Michael Hallsworth

Recent years have seen a rapid increase in the use of behavioural science to address the priorities of public and private sector actors. There is now a vibrant ecosystem of practitioners, teams and academics building on each other's findings across the globe. Their focus on robust evaluation means we know that this work has had an impact on important issues such as antimicrobial resistance, educational attainment and climate change. However, several critiques have also emerged; taken together, they suggest that applied behavioural science needs to evolve further over its next decade. This manifesto for the future of applied behavioural science looks at the challenges facing the field and sets out ten proposals to address them. Meeting these challenges will mean that behavioural science is better equipped to help to build policies, products and services on stronger empirical foundations—and thereby address the world's crucial challenges.

There has been "a remarkable increase in behavioural studies and interventions in public policy on a global scale" over the past 15 years[1]. This growth has been built on developments taking place over many preceding decades. One was the increasing empirical evidence of the importance of non-conscious drivers of behaviour. While psychologists have studied these drivers since at least as far back as the work of William James and Wilhelm Wundt in the nineteenth century, they received renewed attention from the research agenda that showed how "heuristics and biases" influence judgement and decision-making[2]. These and other studies led many psychologists to converge on dual-process theories of behaviour that proposed that rapid, intuitive and non-conscious cognitive processes sit alongside deliberative, reflective and self-aware ones[3].

These theories challenged explanations that foregrounded the role of conscious attitudes, motivations and intentions in determining actions[4]. One result was the creation of the field of behavioural economics, which developed new explanations for why observed behaviour diverged from existing economic models[5]. For example, the concept of "mental accounting" showed how people assign money to certain purposes and—contrary to standard economic theory—are reluctant to repurpose those sums, even when they might benefit from doing so[6].

Behavioural economics may represent only one strand of applied behavioural science, but it has attracted substantial attention. By the mid-2000s, these advances had an increasingly receptive audience among some governments and policymakers[7]. The publication of the book *Nudge* in 2008 responded to this demand by using the evidence mentioned earlier to create practical policy solutions (Box 1)[8]. Then, in 2010, the UK government set up its Behavioural Insights Team[9]. The creation of the Behavioural Insights Team is notable because it became "a paradigmatic example for the translation of behavioural insights into public policy" that acted as "a blueprint for the establishment of similar units elsewhere"[10–12]. Similar initiatives were adopted by many public sector bodies at the local, national and supra-national levels and by private companies large and small[1,11,13,14]. The Organisation for Economic Development and Cooperation has labelled this creation of more than 200 dedicated public entities a "paradigm shift"[15] that shows that applied behavioural science has "taken root in many ways across many countries around the world and across a wide range of sectors and policy areas"[16].

This history is necessarily selective; it does not attempt to cover the full range of work in the behavioural sciences. Rather, my focus is on the main ways that approaches often grouped under the term 'behavioural insights' have been applied to practical issues in the public and private sectors over the past 15 years[17] (see Box 1 for definitions of these and other terms). These approaches have been adopted in both developed and developing economies, and their precise forms of implementation have varied from context to context[18]. However, a crucial point to emphasize is that they have gone far beyond the self-imposed limits of nudges, even if that label is still used (often unhelpfully) as a blanket term. Instead, a broader agenda has emerged that explores how behavioural science can be integrated into core public and private sector activities such as regulation, taxation, strategy and operations. This broader agenda is reflected in the creation of research programmes on "behavioural public policy"[19] or "behavioural public administration"[20].

Proponents of these approaches can point to improved outcomes in many areas, including health[21], education[22], sustainability[23] and criminal justice[24]. Yet criticisms have emerged alongside these successes. For example, there is an ongoing debate about how publication bias

Behavioural Insights Team, Brooklyn, NY, USA. e-mail: michael.hallsworth@bi.team

may have inflated the published effect sizes of nudge interventions[25,26]. Other criticisms target the goals, assumptions and techniques associated with recent applications of behavioural science (Box 2).

This Perspective attempts to respond to these criticisms by setting out an agenda to ensure that applied behavioural science can fulfil its potential in the coming decades. It does so by offering ten proposals, as summarized in Table 1. These proposals fall into three categories: scope (the range and scale of issues to which behavioural science is applied), methods (the techniques and resources that behavioural science deploys) and values (the principles, ideals and standards of conduct that behavioural scientists adopt). These proposals are the product of a non-systematic review of relevant literature and my experience of applying behavioural science. They are not an attempt to represent expert consensus; they aim to provoke debate as well as agreement.

Figure 1 shows how each proposal aims to address one or more of the criticisms set out in Box 2. 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 organizations), academics working in the behavioural sciences (including disciplines such as anthropology, economics and sociology) and funders who support the work of these academics. These groups constitute the 'we' referred to in the rest of the paper, which summarizes a full-length, in-depth report available at www.bi.team.

Scope

Use behavioural science as a lens

The early phase of the behavioural insights movement was marked by scepticism about whether effects obtained in laboratories would translate to real-world settings[27]. In response, practitioners developed standard approaches that could demonstrate a clear causal link between an intervention and an outcome[28]. In practice, these approaches directed attention towards how the design of specific aspects of a policy, product or service influences discrete behaviours by actors who are considered mostly in isolation[29].

These standard approaches are strong and have produced valuable results in many contexts around the world[20,30]. However, in the aggregate, they have also fostered a perspective centred on the metaphor of behavioural science as a specialist tool. This view mostly limits behavioural science to the role of fixing concrete aspects of predetermined interventions rather than aiding the consideration of broader policy goals[31].

Over time, this view has created a self-reinforcing perception that only certain kinds of tasks are suitable for behavioural scientists[29]. Opportunities, skills and ambitions have been constricted as a result; a rebalancing is needed. Behavioural science also has much to say about pressing societal issues such as discrimination, pollution and economic mobility and the structures that produce them[32,33]. These ambitions have always been present in the behavioural insights movement[34], but the factors just outlined acted against their being realized more fully[35].

The first step towards achieving these ambitions is to replace the dominant metaphor of behavioural science as a tool. Instead, behavioural science should be understood as a lens that can be applied to any public or private issue. This change offers several advantages:

- A lens metaphor shows that behavioural science can enhance the use of standard policy options (for example, revealing new ways of structuring taxes) rather than just acting as an alternative to them.
- A lens metaphor conveys that the uses of behavioural science are not limited to creating new interventions. A behavioural science lens can, for example, help to reassess existing actions and understand how they may have unintended effects. It emphasizes the behavioural diagnosis of a situation or issue rather than pushing too soon to define a precise target outcome and intervention[31].
- Specifying that this lens can be applied to any action conveys the error of separating 'behavioural' and 'non-behavioural' issues: most of the goals of private and public action depend on certain behaviours happening (or not). Behavioural science should therefore be integrated into an organization's core activities rather than acting as an optional specialist tool[36].

It may seem odd to start with a change of metaphor, but the primary problem here is one of perception. Behavioural science itself shows us the power of framing: the metaphors we use shape the way we behave and therefore can be agents of change[37]. Metaphors are particularly important in this case because the task of broadening the use of behavioural science requires making a compelling case to decision makers[38]. The metaphor of behavioural science as a tool has established credibility and acceptance in a defined area; expanding beyond that area is the task for the next decade.

Build behavioural science into organizations

The second proposal is to broaden the scope of how behavioural science is used in organizations. Given that many dedicated behavioural science teams exist worldwide, it is understandable that much attention has been paid to the question of how they should be set up successfully. However, this focus has diverted attention from considering how to use behavioural science to shape organizations themselves[39]. We need to talk less about how to set up a dedicated behavioural science team and more about how behavioural science can be integrated into an organization's standard processes. For example, as well as trying to ensure that a departmental budget includes provisions for behavioural science, why not use behavioural science to improve the way this budget is created (for example, are managers anchored to outdated spending assumptions)[40]?

The overriding message here is for greater focus on the organizational changes that indirectly apply or support behavioural science

BOX 1

Glossary of main terms

Behavioural science. In its broadest sense, a discipline that uses scientific methods to generate and test theories that explain and predict the behaviour of individuals, groups and populations. This piece focuses particularly on the implications of dual-process theories of behaviour. Behavioural science is different from 'the behavioural sciences', which refers to a broader group of any scientific disciplines that study behaviour.

Behavioural insights. The application of findings from behavioural science to analyse and address practical issues in real-world settings, usually coupled with a rigorous evaluation of the effects of any interventions. In the current piece, this term is used interchangeably with 'applied behavioural science'.

Behavioural economics. The application of findings from behavioural science to the field of economics to create explanations for economic behaviour that often diverge from the principles of neoclassical economic theory.

Nudge. The design of choices so that non-conscious cognitive processes lead individuals to select the option that leaves them better off, as judged by themselves. Nudges do not involve coercion or any substantial change to economic incentives, leaving people with a meaningful ability to choose a different option from the one that the choice architect intends.

 https://doi.org/10.1038/s41562-023-01555-3

BOX 2

Criticisms of the behavioural insights approach

Limited impact. The approach has focused on more tractable and easy-to-measure changes at the expense of bigger impacts; it has just been tinkering around the edges of fundamental problems[29,50,172].

Failure to reach scale. The approach promotes a model of experimentation followed by scaling, but it has not paid enough attention to how successful scaling happens—and the fact that it often does not happen[18].

Mechanistic thinking. The approach has promoted a simple, linear and mechanistic approach to understanding behaviour that ignores second-order effects and spillovers (and employs evaluation methods that assume a move from A to B against a static background)[29,62,173].

Flawed evidence base. The replication crisis has challenged the evidence base underpinning the behavioural insights approach, adding to existing concerns such as the duration of its interventions' effects[79,174].

Lack of precision. The approach lacks the ability to construct precise interventions and establish what works for whom, and when. Instead, it relies either on overgeneral frameworks or on disconnected lists of biases[80,92,94].

Overconfidence. The approach can encourage overconfidence and overextrapolation from its evidence base, particularly when testing is not an option[175].

Control paradigm. The approach is elitist and pays insufficient attention to people's own goals and strategies; it uses concepts such as irrationality to justify attempts to control the behaviour of individuals, since they lack the means to do so themselves[176,177].

Neglect of the social context. The approach has a limited, overly cognitive and individualistic view of behaviour that neglects the reality that humans are embedded in established societies and practices[125,178,179].

Ethical concerns. The behavioural insights approach will face more ethics, transparency and privacy conundrums as it attempts more ambitious and innovative work[143,145,154].

Homogeneity of participants and perspectives. The range of participants in behavioural science research has been narrow and unrepresentative[164]; homogeneity in the locations and personal characteristics of behavioural scientists influences their viewpoints, practices and theories[124,166].

principles, rather than just thinking through how the direct and overt use of behavioural science can be promoted in an organization. One advantage to this approach is that it can help organizations to address problems with scaling interventions[36]. If some of the barriers to scaling concern cognitive biases in organizations, these changes could minimize the effect of such biases[41]. Rather than starting with a behavioural science project and then trying to scale it, we could start by looking at operations at scale and understanding how they can be influenced.

It is useful to understand how this approach maps onto existing debates about how to set up a behavioural function in organizations. Doing so reveals six main scenarios, as shown in Table 2. In the 'baseline' scenario, there is limited awareness of behavioural science in the organization, and its principles are not incorporated into processes. In the 'nudged organization', behavioural science awareness is still low, but its principles have been used to redesign processes to create better outcomes for staff or service users. In 'proactive consultancy', leaders may have set up a dedicated behavioural team without grafting it onto the organization's standard processes. This lack of institutional grounding puts the team in a less resilient position, meaning that it must always search for new work. In 'call for the experts', an organization has concentrated behavioural expertise, but there are also prompts and resources that allow this expertise to be integrated into business as usual. Expertise is not widespread, but access to it is. Processes stimulate demand for behavioural expertise that the central team can fulfil. In 'behavioural entrepreneurs', there is behavioural science capacity distributed throughout the organization, through either direct capacity building or recruitment. The problem is that organizational processes do not support these individual pockets of knowledge. Finally, a 'behaviourally enabled organization' is one where there is knowledge of behavioural science diffused throughout the organization, which also has processes that reflect this knowledge and support its deployment.

Most discussions make it seem like the meaningful choice is between the different columns in Table 2—how to organize dedicated behavioural science resources. Instead, the more important move is from the top row to the bottom row: moving from projects to processes, from commissions to culture. A useful way of thinking about this task is about building or upgrading the "choice infrastructure" of the organization[42]. In other words, we should place greater focus on the institutional conditions and connections that support the direct and indirect ways that behavioural science can infuse organizations.

Working out how best to build the choice infrastructure in organizations should be a major priority for applied behavioural science. Already we can see that some features will be crucial: reducing the costs of experimentation, creating a system that can learn from its actions, and developing new and better ways of using behavioural science principles to analyse the behavioural effects of organizational processes, rules, incentives, metrics and guidelines[36].

See the system

Many important policy challenges emerge from complex adaptive systems, where change often does not happen in a linear or easily predictable way, and where coherent behaviour can emerge from interactions without top-down direction[43]. There are many examples of such systems in human societies, including cities, markets and political movements[44]. These systems can create "wicked problems"—such as the COVID-19 pandemic—where ideas of success are contested, changes are nonlinear and difficult to model, and policies have unintended consequences[45].

This reality challenges the dominant behavioural science approach, which usually assumes stability over time, keeps a tight focus on predefined target behaviours and predicts linear effects on the basis of a predetermined theory of change[46]. The result, some argue, is a failure to understand how actors are acting and reacting in a complex system that leads policymakers to conclude they are being irrational—and then actually disrupt the system in misguided attempts to correct perceived biases or inefficiencies[47–49].

Table 1 | Summary of proposed actions to improve future applications of behavioural science

Category	Proposal	Recommended action(s)
Scope	Use behavioural science as a lens	Present behavioural science as a lens that improves the view of any public and private issue to break a self-sustaining pattern that has directed behavioural science away from the most important problems.
	Build behavioural science into organizations	Focus less on how to set up a dedicated behavioural science team and more on how the approach can be integrated into an organization's standard processes by upgrading its choice infrastructure.
	See the system	Use aspects of complexity thinking to improve behavioural science so that it can exploit leverage points, model the collective implications of heuristics, alter specific features of systems to create wider changes, and understand the longer-term impact on a system of a collection of policies with varying goals.
Methods	Put RCTs in their place	Strengthen RCTs to deal better with complexity by gaining a better understanding of the system interactions and anticipating how they may play out, setting up RCTs to measure diffusion and contagion in networks, and building feedback and adaptation into the design of RCTs and interventions.
	Replication, variation and adaptation	Identify the most reliable interventions, develop an accurate sense of the likely size of their effects and avoid the weaker options. Recognize that heterogeneity requires a much higher bar for claiming that an effect holds true across many unspecified settings. Create multi-site studies to systematically study heterogeneity in a wider range of contexts and participants. Codify and cultivate the practical skills that successfully adapt interventions to new contexts.
	Beyond lists of biases	Emphasize theories that are practical: they fill the gap between high-level frameworks and jumbled lists of biases; they are based on data and generate testable hypotheses, but they also specify the conditions under which a prediction applies; and they present actionable steps to solve real-world problems.
	Predict and adjust	Develop the practice of getting behavioural scientists to predict the results of experiments and then feeding back the results to them.
Values	Be humble, explore and enable	Avoid using the term 'irrationality', practice epistemic humility, and design processes and institutions to counteract overconfidence. Pay greater attention to people's own interpretations of their beliefs, feelings and behaviours. Reach a wider range of experiences, including marginalized voices and communities. Recognize how apparently universal cognitive processes are shaped by specific contexts. Use six criteria (detailed in the main text) to assess when to enable people to use behavioural science themselves.
	Data science for equity	Use data science to identify the ways in which an intervention or situation appears to increase inequalities and introduce features to reduce them. For example, groups that are particularly likely to miss a filing requirement could be offered pre-emptive help.
	No "view from nowhere"	Cultivate self-scrutiny, find new ways for the subjects of research to judge researchers, and take actions to increase diversity among behavioural scientists and their teams, such as building professional networks between the Global North and Global South.

These criticisms may overstate the case, but they point to a way forward. Behavioural science can be improved by using aspects of complexity thinking to offer new, credible and practical ways of addressing major policy issues. The first step is to reject crude distinctions of 'upstream' versus 'downstream' or the 'individual frame' versus the 'system frame'[50]. Instead, complex adaptive systems show that higher-level features of a system can actually emerge from the lower-level interactions of actors participating in the system[44]. When they become the governing features of the system, they then shape the lower-level behaviour until some other aspect emerges, and the fluctuations continue. An example might be the way that new coronavirus variants emerged in particular settings and then went on to change the course of the whole pandemic, requiring new overall strategic responses.

In other words, we are dealing with "cross-scale behaviours"[49]. For example, norms, rules, practices and culture itself can emerge from aggregated social interactions; these features then shape cognition and behavioural patterns in turn[51]. Recognizing cross-scale behaviours means that behavioural science could:

- Identify "leverage points" where a specific shift in behaviour will produce wider system effects[52]. One option is to identify when and where tipping points are likely to occur in a system and then either nudge them to occur or not, depending on the policy goal[53]. For example, if even a subset of consumers decides to switch to a healthier version of a food product, this can have broader effects on a population's health through the way the food system responds by restocking and product reformulation[54].
- Model the collective implications of individuals using simple heuristics to navigate a system. For example, new models show how small changes to simple heuristics that guide savings (in this case, how quickly households copy the savings behaviours of neighbours) can lead to the sudden emergence of inequalities in wealth[55].
- Find targeted changes to features of a system that create the conditions for wide-ranging shifts in behaviour to occur. For example, a core driver of social media behaviours is the ease with which information can be shared[46]. Even minor changes to this parameter can drive widespread changes—some have argued that such a change is what created the conditions leading to the Arab Spring, for example[56].

This approach also suggests that a broader change in perspective is needed. We need to realize the flaws in launching interventions in isolation and then moving on when a narrowly defined goal has been achieved. Instead, we need to see the longer-term impact on a system of a collection of different policies with varying goals[57]. The best approach may be "system stewardship", which focuses on creating the conditions for behaviours and indirectly steering adaptation towards overall goals[58].

Of course, not every problem will involve a complex adaptive system; for simple issues, standard approaches to applying behavioural science work well. Behavioural scientists should therefore develop the skills to recognize the type of system that they are facing (see the system) and then choose their approach accordingly. These skills can be developed through agent-based simulations[59], immersive technologies[60] or just basic checklists[61].

Methods

Put randomized controlled trials in their place

Randomized controlled trials (RCTs) have been a core part of applied behavioural science, and they work well in relatively simple and stable contexts. But they can fare worse in complex adaptive systems, whose

https://doi.org/10.1038/s41562-023-01555-3

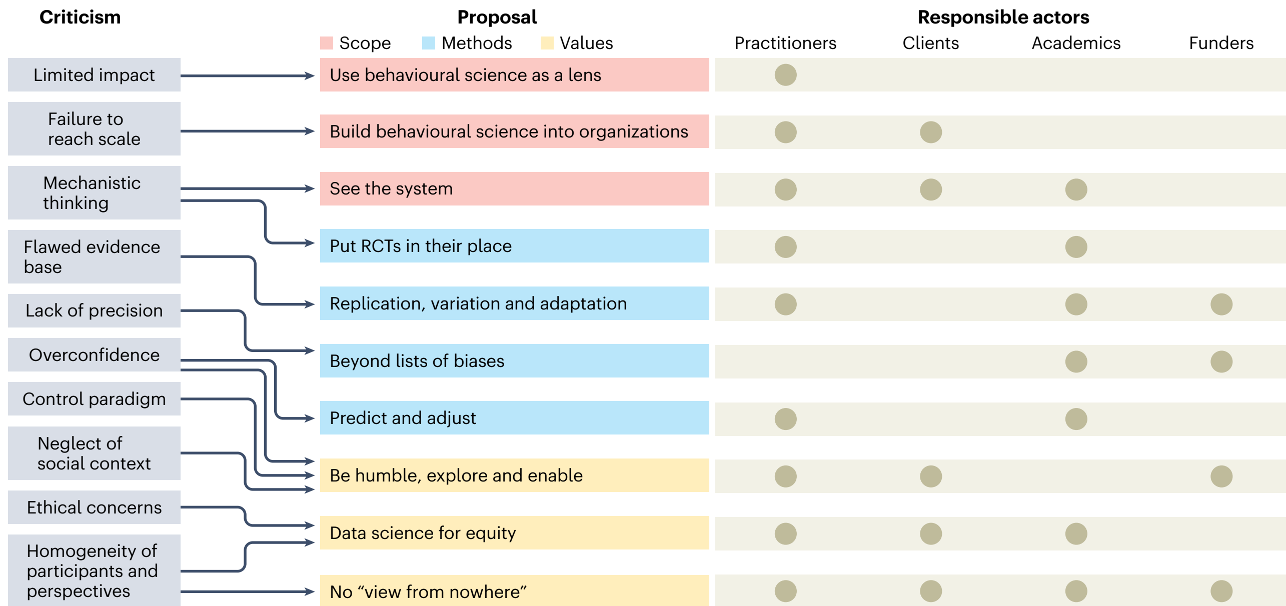

Fig. 1 | Manifesto overview. The left side shows common criticisms made of the behavioural insights approach. The middle column presents ten proposals to improve the way behavioural science is applied. These proposals are organized into three categories (scope, methods and values), which are represented by red, blue and yellow, respectively. The arrows from the criticisms to the proposals show which of the latter attempt to address the former. The matrix on the right shows the four main groups involved with implementing the proposals: practitioners, clients, academics and funders. The dots in each column indicate that the relevant group will need to make a substantive contribution to achieving the goals of the proposal in the corresponding row.

many shifting connections can make it difficult to keep a control group isolated and where a narrow focus on predetermined outcomes may neglect others that are important but difficult to predict[43,62].

We can strengthen RCTs to deal better with complexity. We can try to gain a better understanding of the system interactions and anticipate how they may play out, perhaps through "dark logic" exercises that try to trace potential harms rather than just benefits[63]. For example, we might anticipate that sending parents text messages encouraging them to talk to their children about the school science curriculum may achieve this outcome at the expense of other school-supporting behaviours—as turned out to be the case[64]. Engaging the people who will implement and participate in an intervention will be a key part of this effort.

Another option is to set up RCTs to measure diffusion and contagion in networks, either by creating separate online environments or by randomizing real-world clusters, such as separate villages[65,66]. Finally, we can build feedback and adaptation into the design of the RCT and the intervention, allowing adjustments to changing conditions[67,68]. Options include using two-stage trial protocols[69], evolutionary RCTs[70], sequential multiple assignment randomized trials[71] and 'bandit' algorithms that identify high-performing interventions and allocate more people to them[72].

Behavioural science can also be used to enhance alternative ways of measuring impacts—in particular, agent-based modelling, which tries to simulate the interactions between the different actors in a system[73]. The agents in these models are mostly assumed to be operating on rational choice principles[74,75]. There is therefore an opportunity to build in more evidence about the drivers of behaviour—for example, habits and social comparisons[49].

Replication, variation and adaptation

The 'replication crisis' of the past decade has seen intense debate and concern about the reliability of behavioural science findings. Poor research practices were a major cause of the replication crisis; the good news is that many have improved as a result[76,77]. Now there are sharper incentives to preregister analysis plans, greater expectations that data and code will be freely shared, and wider acceptance of post-publication review of findings[78].

Behavioural scientists need to secure and build on these advances to move towards a future where appropriately scoped meta-analyses of high-quality studies (including deliberate replications) are used to identify the most reliable interventions, develop an accurate sense of the likely size of their effects and avoid the weaker options. We have a responsibility to discard ideas if solid evidence now shows that they are shaky, and to offer a realistic view of what behavioural science can accomplish[18].

That responsibility also requires us to have a hard conversation about heterogeneity in results: the complexity of human behaviour creates so much statistical noise that it is often hard to detect consistent signals and patterns[79]. The main drivers of heterogeneity are that contexts influence results and that the effect of an intervention may vary greatly between groups within a population[80,81]. For example, choices of how to set up experiments vary greatly between studies and researchers, in ways that often go unnoticed[82]. A recent study ran an experiment to measure the impact of these contextual factors. Participants were randomly allocated to studies designed by different research teams to test the same hypothesis. For four of the five research questions, studies actually produced effects in opposing directions. These "radically dispersed" results indicate that "idiosyncratic choices in stimulus design have a very large effect on observed results"[83]. These factors complicate the idea of replication itself: a 'failed' replication may not show that a finding was false but rather show how it exists under some conditions and not others[84].

These challenges mean that applied behavioural scientists need to set a much higher bar for claiming that an effect holds true across many unspecified settings[85]. There is a growing sense that interventions should be talked about as hypotheses that were true in one place and that may need adapting to be true elsewhere[18,86].

Narrative changes need to be complemented by specific proposals. The first concerns data collection: behavioural scientists should expand studies to include (and thus examine) a wider range of contexts and participants and gather richer data about them. To date, only a small minority of behavioural studies have provided enough information to see how effects vary[87]. Moreover, the gaps in data coverage may result from and create systemic issues in society: certain groups may

Table 2 | Options for building behavioural science into organizations

		Behavioural science knowledge and capacity		
		Limited	**Concentrated**	**Diffused**
Behavioural science incorporated into organizational processes	No	Baseline	Proactive consultancy	Behavioural entrepreneurs
	Yes	Nudged organization	Call for the experts	Behaviourally enabled organization

The rows indicate whether behavioural science has been used to shape the organization's own structures or processes, using a deliberately crude yes/no distinction to make the table manageable. The columns deal with the extent and form of behavioural science knowledge and capacity in the organization.

be excluded or may have their data recorded differently from others[88]. Coordinated multi-site studies will be needed to collect enough data to explore heterogeneity systematically; crowdsourced studies offer particular promise for testing context and methods[83]. Realistically, this work is going to require a major investment in research infrastructure to set up standing panels of participants, coordinate between institutions, and reduce barriers to data collection and transfer[80]. These efforts cannot be limited to just a few countries.

Behavioural scientists also need to get better at judging how strongly an intervention's results were linked to its context and therefore how much adaptation it needs[81]. We should use and modify frameworks from implementation science to develop such judgement[89]. Finally, we need to codify and cultivate the practical skills that successfully adapt interventions to new contexts; expertise in behavioural science should not be seen as simply knowing about concepts and findings in the abstract. It is therefore particularly valuable to learn from practitioners how they adapted specific interventions to new contexts. These accounts are starting to emerge, but they are still rare[18], since researchers are incentivized to claim universality for their results rather than report and value contextual details[82].

Beyond lists of biases

The heterogeneity in behavioural science findings also means that our underlying theories need to improve: we are lacking good explanations for why findings vary so much[84]. This need for better theories can be seen as part of a wider 'theory crisis' in psychology, which has thrown up two big concerns for behavioural science[90,91].

The first stems from the fact that theories of behaviour often try to explain phenomena that are complex and wide-ranging[92]. If you are trying to show how emotion and cognition interact (for example), this involves many causes and interactions. Trying to cover this variability can produce descriptions of relationships and definitions of constructs that are abstract and imprecise[85]. The result is theories that are vague and weak, since they can be used to generate many different hypotheses—some of which may actually contradict each other[90]. That makes theories hard to disprove, and so weak theories stumble on, unimproved[93].

The other concern is that theories can make specific predictions, but they are disconnected from each other—and from a deeper, general framework that can provide broader explanations (such as evolutionary theory)[94]. The main way this issue affects behavioural science is through heuristics and biases. Examples of individual biases are accessible, popular and how many people first encounter behavioural science. These ideas are incredibly useful, but they have often been presented as lists of standalone curiosities in a way that is incoherent, reductive and deadening. Presenting lists of biases does not help us to distinguish or organize them[95–97]. Such lists can also create overconfident thinking that targeting a specific bias (in isolation) will achieve a certain outcome[98].

Perhaps most importantly, focusing on lists of biases distracts us from answering core underlying questions. When does one or another bias apply? Which are widely applicable, and which are highly specific? How does culture or life experience affect whether a bias influences behaviour or not[99,100]? These are highly practical questions when one is faced with tasks such as taking an intervention to new places.

The concern for behavioural science is that it uses both these high-level frameworks (such as dual-process theories) and jumbled collections of heuristics and biases, with little in the middle to draw both levels together[94]. Recent years have seen valuable advances in connecting and systematizing theories[101,102]. At the same time, there are various ongoing attempts to create strong theories: "coherent and useful conceptual frameworks into which existing knowledge can be integrated"[93] (see also refs. [91,103,104]). Naturally, such work should continue, but I think that applied behavioural science will benefit particularly from theories that are practical. By this I mean:

- They fill the gap between day-to-day working hypotheses and comprehensive and systematic attempts to find universal underlying explanations.
- They are based on data rather than being derived from pure theorizing[105].
- They can generate testable hypotheses, so they can be disproved[106].
- They specify the conditions under which a prediction applies or does not[85].
- They are geared towards realistic adaptation by practitioners and offer "actionable steps toward solving a problem that currently exists in a particular context in the real world"[107].

Resource rationality may be a good example of a practical theory. It starts from the basis that people make rational use of their limited cognitive resources[108]. Given that there is a cost to thinking, people will look for solutions that balance choice quality with effort. Resource rationality can offer a "unifying framework for a wide range of successful models of seemingly unrelated phenomena and cognitive biases" that can be used to build models for how people act[108].

A recent study has shown how these models not only can predict how people will respond to different kinds of nudges in certain contexts but also can be integrated with machine learning to create an automated method for constructing "optimal nudges"[109]. Such an approach could reveal new kinds of nudges and make creating them much more efficient. More reliable ways of developing personalized nudges are also possible. These are all highly practical benefits coming from applying a particular theory.

Predict and adjust

Hindsight bias is what happens when we feel 'I knew it all along', even if we did not[110]. When the results of an experiment come in, hindsight bias may mean that behavioural scientists are more likely to think that they had predicted them or quickly find ways of explaining why they occurred. Hindsight bias is a big problem because it breeds overconfidence, impedes learning, dissuades innovation and prevents us from understanding what is truly unexpected[111,112].

In response, behavioural scientists should establish a standard practice of predicting the results of experiments and then receiving feedback on how their predictions performed. Hindsight bias can flourish if we do not systematically capture expectations or priors about what the results of a study will be[113]. Making predictions provides regular, clear feedback of the kind that is more likely to trigger surprise and reassessment rather than hindsight bias[114]. Establishing the average expert prediction—which may be different from the null hypothesis in an experiment—clearly reveals when results challenge the consensus[115].

There are existing practices to build on here, such as the practice of preregistering hypotheses and trial protocols and the use of a Bayesian approach to make priors explicit. Indeed, more and more studies are explicitly integrating predictions[116,117]. However, barriers lie in the way of further progress. People may not welcome the ensuing challenge to their self-image, predicting may seem like one thing too many on the to-do list, and the benefits lie in the future. Some responses to these challenges are to make predicting easy by incorporating it into standard processes; minimize threats to predictors' self-image (for example, by making and feeding back predictions anonymously)[118]; give concrete prompts for learning and reflection, to disrupt the move from surprise to hindsight bias[119]; and build learning from prediction within and between institutions.

Values

Be humble, explore and enable

This proposal is made up of three connected ideas. First, behavioural scientists need to become more aware of the limits of their knowledge and to avoid fitting behaviours into pre-existing ideas around biases or irrationality. Second, they should broaden the exploratory work they conduct, in terms of gaining new types of qualitative data and recognizing how experiences vary by group and geography. Finally, they should develop new approaches to enable people to apply behavioural science themselves—and adopt new criteria for judging when these approaches are appropriate.

Humility is important because behavioural scientists (like other experts) may overconfidently rely on decontextualized principles that do not match the real-world setting for a behaviour[29]. Deeper inquiry can reveal reasonable explanations for what seem to be behavioural biases[120]. In response, those applying behavioural science should avoid using the term 'irrationality', which can limit attempts to understand actions in context; acknowledge that diagnoses of behaviour are provisional and incomplete (epistemic humility)[121]; and design processes and institutions to counteract overconfidence[122].

How do we conduct these deeper inquiries? Three areas demand particular focus in the future. First, pay greater attention to people's goals and strategies and their own interpretations of their beliefs, feelings and behaviours[123]. Second, reach a wider range of experiences, including marginalized voices and communities, understanding how structural inequalities can lead to expectations and experiences varying greatly by group and geography[124]. Third, recognize how apparently universal cognitive processes are shaped by specific contexts, thereby unlocking new ways for behavioural science to engage with values and culture[125,126]. For example, one influential view of culture is that it influences action "not by providing the ultimate values toward which action is oriented but by shaping a repertoire or 'toolkit' of habits, skills, and styles"[127]. There are similarities here to the heuristics-and-biases toolkit perspective on behaviour: behavioural scientists could start explaining how and when certain parts of the toolkit become more or less salient.

More can and should be done to broaden ownership of behavioural science approaches. Many (but far from all) behavioural science applications have been top-down, with a choice architect enabling certain outcomes[8,128]. One route is to enable people to become more involved in designing interventions that affect them—and "nudge plus"[129], "self-nudges"[130] and "boosts"[131] have been proposed as ways of doing this. Reliable criteria are needed to decide when enabling approaches may be appropriate, including whether the opportunity to use an enabling approach exists; ability and motivation; preferences; learning and setup costs; equity impacts; and effectiveness, recognizing that evidence on this point is still emerging[132,133].

But these new approaches should not be seen simplistically as enabling alternatives to disempowering nudges[134]. Instead, we need to consider how far the person performing the behaviour is involved in shaping the initiative itself, as well as the level and nature of any capacity created by the intervention. People may be heavily engaged in selecting and developing a nudge intervention that nonetheless does not trigger any reflection or build any skills[135]. Alternatively, a policymaker may have paternalistically assumed that people want to build up their capacity to perform an action, when in fact they do not. This is the real choice to be made.

A final piece missing from current thinking is that enabling people can lead to a major decentring of the use of behavioural science. If more people are enabled to use behavioural science, they may decide to introduce interventions that influence others[136]. Rather than just creating self-nudges through altering their immediate environments, they may decide that wider system changes are needed instead. A range of people could be enabled to create nudges that generate positive societal change (with no central actors involved). This points towards a future where policy or product designers act less like (choice) architects and more like facilitators, brokers and partnership builders[137].

Data science for equity

Recent years have seen growing interest in using new data science techniques to reliably analyse the heterogeneity of large datasets[138,139]. Machine learning is claimed to offer more sophisticated, reliable and data-driven ways of detecting meaningful patterns in datasets[140,141]. For example, a machine learning approach has been shown to be more effective than conventional segmentation approaches at analysing patterns of US household energy usage to reduce peak consumption[142].

A popular idea is to use such techniques to better understand what works best for certain groups and thereby tailor an offering to them[143]. Scaling an intervention stops being about a uniform roll-out and instead becomes about presenting recipients with the aspects that are most effective for them[144].

This vision is often presented as straightforward and obviously desirable, but it runs almost immediately into ethical quandaries and value judgements. People are unlikely to know what data have been used to target them and how; the specificity of the data involved may make manipulation more likely, since it may exploit sensitive personal vulnerabilities; and expectations of universality and non-discrimination in public services may be violated[143,145].

Closely related to manipulation concerns is the fear that data science will open up new opportunities to exploit, rather than to help, the vulnerable[146]. One aspect is algorithmic bias. Models using data that reflect historical patterns of discrimination can produce results that reinforce these outcomes[147]. Since disadvantaged groups are more likely to be subject to the decisions of algorithms, there is a particular risk that inequalities will be perpetuated—although some studies argue that algorithms are actually less likely to be biased than human judgement[148,149].

There is also emerging evidence that people often object to personalization. While they support some personalized services, they consistently oppose advertising that is customized on the basis of sensitive information—and they are generally against the collection of the information that personalization relies on[150]. To navigate this landscape, behavioural scientists need to examine four factors:

- Who does the personalization target, and using what criteria? Many places have laws or norms to ensure equal treatment based on personal characteristics. When does personalization violate those principles?
- How is the intervention constructed? To what extent do the recipients have awareness of the personalization, choice over whether it occurs, control over its level or nature, and the opportunity to give feedback on it[151]?
- When is it directed? Is it at a time when the participant is vulnerable? Would they probably regret it later, if they had time to reflect?

https://doi.org/10.1038/s41562-023-01555-3

- Why is personalization happening? Does it aim to exploit and harm or to support and protect, recognizing that those terms are often contested?

Taking these factors into account, I propose that the main opportunity is for data science to identify the ways in which an intervention or situation appears to increase inequalities, and reduce them[152]. For example, groups that are particularly likely to miss a filing requirement could be offered pre-emptive help. Algorithms can be used to better explain the causes of increased knee pain experienced in disadvantaged communities, thereby giving physicians better information to act on[153].

I call this idea data science for equity. It addresses the 'why' factor by using data science to support, not exploit. 'Data science for equity' may seem like a platitude, but it is a very real choice: the combination of behavioural and data science is powerful and has been used to create harm in the past. Moreover, it needs to be complemented by attempts to increase agency (the 'how' factors), as in a recent study that showed how boosts can be used to help people to detect micro-targeting of advertising[154], and studies that obtain more data on which uses of personalization people find acceptable.

No "view from nowhere"

The final proposal is one of the most wide-ranging, challenging and important. For the philosopher Thomas Nagel, the "view from nowhere" was an objective stance that allowed us to "transcend our particular viewpoint"[155]. Taking such a stance may not be possible for behavioural scientists. We bring certain assumptions and ways of seeing to what we do; we are always situated in, embedded in and entangled with ideas and situations[124]. We cannot assume that there is some set-aside position from which to observe the behaviour of others; no objective observation deck outside society exists[156].

Behavioural scientists are defined by having knowledge, skills and education; many of them can use these resources to shape public and private actions. They are therefore in a privileged position, but they may not see the extent to which they hold elite positions that stop them from understanding people who think differently (for example, those who are sceptical of education)[157]. The danger is that elites place their group values and preferences on others, while thinking that they are adopting a view from nowhere[158,159]. This does not mean that they can never act or opine, but rather that they need to carefully understand their own positionality and those of others before doing so.

There have been repeated concerns that the field is still highly homogeneous in other ways as well. Gender, race, physical abilities, sexuality and geography also influence the viewpoints, practices and theories of behavioural scientists[160,161]. Only a quarter of the behavioural insights teams catalogued in a 2020 survey were based in the Global South[162]. An over-reliance on using English in cognitive science has led to the impact of language on thought being underestimated[163]. The past decade has shown how behaviours can vary greatly from culture to culture, even as psychology has tended to generalize from relatively small and unrepresentative samples[164]. Behavioural science studies often present data from Western, educated, industrialized, rich and democratic samples as more generalizable to humans as a whole[165]. So, rather than claiming that science is value-free, we need to find realistic ways of acknowledging and improving this reality[166].

A starting point is for behavioural scientists to cultivate self-scrutiny by querying how their identities and experiences contribute to their stance on a topic. Hypothesis generation could particularly benefit from this exercise, since arguably it is closely informed by the researcher's personal priorities and preferences[167]. Behavioural scientists could be actively reflecting on interventions in progress, including what factors are contributing to power dynamics[168]. Self-scrutiny may not be enough. We should also find more ways for people to judge researchers and decide whether they want to participate in research—going beyond consent forms. If they do participate, there are many opportunities to combine behavioural science with co-design[128].

Finally, we should take actions to increase diversity (of several kinds) among behavioural scientists, teams, collaborations and institutions. Doing this requires addressing barriers such as the lack of professional networks connecting the Global North and Global South, and the time needed to build understanding of the tactics required to write successful grant applications from funders[169]. In many countries, much more could be done to increase the ethnic and racial diversity of the behavioural science field—for example, through support for starting and completing PhDs or through reducing the substantial racial gaps present in much public funding of research[170,171].

Conclusion

Applied behavioural science has seen rapid growth and meaningful achievements over the past decade. Although the popularity of nudging provided its initial impetus, an ambition soon formed to apply a broader range of techniques to a wider range of goals. However, a set of credible critiques have emerged as levels of activity have grown. As Fig. 1 indicates, there are proposals that can address these critiques (and progress is already being made on some of them). When considered together, these proposals present a coherent vision for the scope, methods and values of applied behavioural science.

This vision is not limited to technical enhancements for the field; it also covers questions of epistemology, identity, politics and praxis. A common theme throughout the ten proposals is the need for self-reflective practice that is aware of how its knowledge and approaches have originated and how they are situated. In other words, 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.

Realizing these proposals will require sustained work and experiencing the discomfort of disrupting what may have become familiar and comfortable practices. That is a particular problem because incentives for change are often weak or absent. Improving applied behavioural science has some characteristics of a social dilemma: the benefits are diffused across the field as a whole, while the costs fall on any individual party who chooses to act (or act first). Practitioners are often in competition. Academics often want to establish a distinctive research agenda. Commissioners are often rewarded for risk aversion. Impaired coordination is particularly problematic because coordination forms the basis for several necessary actions, such as the multi-site studies to measure heterogeneity.

Solving these problems will be hard. Funders need to find mechanisms that adequately reward coordination and collaboration by recognizing the true costs involved. Practitioners need to perceive the competitive advantages of adopting new practices and be able to communicate them to clients. Clients themselves need to have a realistic sense of what can be achieved but still be motivated to commit resources. Stepping back, the starting point for addressing these barriers needs to be a change in the narrative about what the field does and could do—a new set of ambitions to aim for. This manifesto aims to help to shape such a narrative.

References

1. Straßheim, H. The rise and spread of behavioral public policy: an opportunity for critical research and self-reflection. *Int. Rev. Public Policy* **2**, 115–128 (2020).
2. Gilovich, T., Griffin, D. & Kahneman, D. (eds) *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge Univ. Press, 2002).
3. Lieberman, M. D. Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.* **58**, 259–289 (2007).
4. Ajzen, I. The theory of planned behaviour. *Organ. Behav. Hum. Decis. Process.* **50**, 179–211 (1991).

5. Thaler, R. *Misbehaving: The Making of Behavioral Economics* (W.W. Norton, 2015).
6. Thaler, R. Mental accounting and consumer choice. *Mark. Sci.* **4**, 199–214 (1985).
7. Pykett, J., Jones, R. & Whitehead, M. (eds) *Psychological Governance and Public Policy: Governing the Mind, Brain and Behaviour* (Routledge, 2017).
8. Thaler, R. & Sunstein, S. *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Yale Univ. Press, 2008).
9. Hallsworth, M. & Kirkman, E. *Behavioral Insights* (MIT Press, 2020).
10. Oliver, A. *The Origins of Behavioural Public Policy* (Cambridge Univ. Press, 2017).
11. Straßheim, H. & Beck, S. in *Handbook of Behavioural Change and Public Policy* (eds Beck, S. & Straßheim, H.) 1–22 (Edward Elgar, 2019).
12. Ball, S. & Feitsma, J. The boundaries of behavioural insights: observations from two ethnographic studies. *Evid. Policy* **16**, 559–577 (2020).
13. Afif, Z., Islan, W., Calvo-Gonzalez, O. & Dalton, A. *Behavioural Science Around the World: Profiles of 10 Countries* (World Bank, 2018).
14. Science that can change the world. *Nat. Hum. Behav*. https://doi.org/10.1038/s41562-019-0642-2 (2019).
15. *Behavioural Insights and New Approaches to Policy Design: The Views from the Field* (OECD, 2015).
16. *Behavioural Insights and Public Policy: Lessons from Around the World* (OECD, 2015).
17. Feitsma, J. N. P. The behavioural state: critical observations on technocracy and psychocracy. *Policy Sci.* **51**, 387–410 (2018).
18. Mažar, N. & Soman, D. (eds) *Behavioral Science in the Wild* (Univ. of Toronto Press, 2022).
19. Oliver, A. Towards a new political economy of behavioral public policy. *Public Adm. Rev.* **79**, 917–924 (2019).
20. Grimmelikhuijsen, S., Jilke, S., Olsen, A. L. & Tummers, L. Behavioral public administration: combining insights from public administration and psychology. *Public Adm. Rev.* **77**, 45–56 (2017).
21. Cadario, R. & Chandon, P. Which healthy eating nudges work best? A meta-analysis of field experiments. *Mark. Sci.* **39**, 465–486 (2020).
22. Damgaard, M. T. & Nielsen, H. S. Nudging in education. *Econ. Educ. Rev.* **64**, 313–342 (2018).
23. Ferrari, L., Cavaliere, A., De Marchi, E. & Banterle, A. Can nudging improve the environmental impact of food supply chain? A systematic review. *Trends Food Sci. Technol.* **91**, 184–192 (2019).
24. Fishbane, A., Ouss, A. & Shah, A. K. Behavioral nudges reduce failure to appear for court. *Science* **370**, eabb6591 (2020).
25. Mertens, S., Herberz, M., Hahnel, U. J. & Brosch, T. The effectiveness of nudging: a meta-analysis of choice architecture interventions across behavioral domains. *Proc. Natl Acad. Sci. USA* **119**, e2107346118 (2022).
26. Maier, M. et al. No evidence for nudging after adjusting for publication bias. *Proc. Natl Acad. Sci. USA* **119**, e2200300119 (2022).
27. Levitt, S. D. & List, J. A. What do laboratory experiments measuring social preferences reveal about the real world? *J. Econ. Perspect.* **21**, 153–174 (2007).
28. Hansen, P. G. *Tools and Ethics for Applied Behavioural Insights: The BASIC Toolkit* (OECD, 2019).
29. Schmidt, R. & Stenger, K. Behavioral brittleness: the case for strategic behavioral public policy. *Behav. Public Policy*, https://doi.org/10.1017/bpp.2021.16 (2021).
30. *United Nations Behavioural Science Report* https://www.uninnovation.network/assets/BeSci/UN_Behavioural_Science_Report_2021.pdf (United Nations Innovation Network, 2021).
31. Hansen, P. G. What are we forgetting? *Behav. Public Policy* **2**, 190–197 (2018).
32. Van Rooij, B. & Fine, A. *The Behavioral Code: The Hidden Ways the Law Makes Us Better or Worse* (Beacon, 2021).
33. Andre, P., Haaland, I., Roth, C. & Wohlfart, J. *Narratives about the Macroeconomy* CEBI Working Paper Series (Center for Economic Behavior and Inequality, 2021).
34. Dolan, P., Hallsworth, M., Halpern, D., King, D. & Vlaev, I. *MINDSPACE: Influencing Behaviour Through Public Policy* (Institute for Government and Cabinet Office, 2010).
35. Meder, B., Fleischhut, N. & Osman, M. Beyond the confines of choice architecture: a critical analysis. *J. Econ. Psychol.* **68**, 36–44 (2018).
36. Soman, D. & Yeung, C. (eds) *The Behaviorally Informed Organization* (Univ. of Toronto Press, 2020).
37. Thibodeau, P. H. & Boroditsky, L. Metaphors we think with: the role of metaphor in reasoning. *PLoS ONE* **6**, e16782 (2011).
38. Feitsma, J. Brokering behaviour change: the work of behavioural insights experts in government. *Policy Polit.* **47**, 37–56 (2019).
39. Battaglio, R. P. Jr, Belardinelli, P., Bellé, N. & Cantarelli, P. Behavioral public administration *ad fontes*: a synthesis of research on bounded rationality, cognitive biases, and nudging in public organizations. *Public Adm. Rev.* **79**, 304–320 (2019).
40. Cantarelli, P., Bellé, N. & Belardinelli, P. Behavioral public HR: experimental evidence on cognitive biases and debiasing interventions. *Rev. Public Pers. Adm.* **40**, 56–81 (2020).
41. Mayer, S., Shah, R. & Kalil, A. in *The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do about It* (eds List, J. et al.) 41–57 (Routledge, 2021).
42. Schmidt, R. A model for choice infrastructure: looking beyond choice architecture in behavioral public policy. *Behav. Public Policy*, https://doi.org/10.1017/bpp.2021.44 (2022).
43. *Magenta Book 2020: Supplementary Guide: Handling Complexity in Policy Evaluation* (HM Treasury, 2020).
44. Boulton, J. G., Allen, P. M. & Bowman, C. *Embracing Complexity: Strategic Perspectives for an Age of Turbulence* (Oxford Univ. Press, 2015).
45. Angeli, F., Camporesi, S. & Dal Fabbro, G. The COVID-19 wicked problem in public health ethics: conflicting evidence, or incommensurable values? *Hum. Soc. Sci. Commun.* **8**, 1–8 (2021).
46. Bak-Coleman, J. B. et al. Stewardship of global collective behavior. *Proc. Natl Acad. Sci. USA* **118**, e2025764118 (2021).
47. Dunlop, C. A. & Radaelli, C. M. in *Nudge and the Law: A European Perspective* (eds Alemanno, A. & Simony, A. L.) 139–158 (Hart, 2015).
48. Scott, J. C. *Seeing Like a State* (Yale Univ. Press, 1998).
49. Schill, C. et al. A more dynamic understanding of human behaviour for the Anthropocene. *Nat. Sustain.* **2**, 1075–1082 (2019).
50. Chater, N. & Loewenstein, G. The i-Frame and the s-Frame: how focusing on individual-level solutions has led behavioral public policy astray. *Behav. Brain Sci.* https://doi.org/10.1017/S0140525X22002023 (2022).
51. DiMaggio, P. & Markus, H. R. Culture and social psychology: converging perspectives. *Soc. Psychol. Q.* **73**, 347–352 (2010).
52. Abson, D. J. et al. Leverage points for sustainability transformation. *Ambio* **46**, 30–39 (2017).
53. Andreoni, J., Nikiforakis, N. & Siegenthaler, S. Predicting social tipping and norm change in controlled experiments. *Proc. Natl Acad. Sci. USA* **118**, e2014893118 (2021).
54. Hallsworth, M. Rethinking public health using behavioural science. *Nat. Hum. Behav.* **1**, 612 (2017).

55. Asano, Y. M., Kolb, J. J., Heitzig, J. & Farmer, J. D. Emergent inequality and business cycles in a simple behavioral macroeconomic model. *Proc. Natl Acad. Sci. USA* **118**, e2025721118 (2021).
56. Jones-Rooy, A. & Page, S. E. The complexity of system effects. *Crit. Rev.* **24**, 313–342 (2012).
57. Hawe, P., Shiell, A. & Riley, T. Theorising interventions as events in systems. *Am. J. Community Psychol.* **43**, 267–276 (2009).
58. Hallsworth, M. *System Stewardship: The Future of Policymaking?* (Institute for Government, 2011).
59. Rates, C. A., Mulvey, B. K., Chiu, J. L. & Stenger, K. Examining ontological and self-monitoring scaffolding to improve complex systems thinking with a participatory simulation. *Instr. Sci.* **50**, 199–221 (2022).
60. Fernandes, L., Morgado, L., Paredes, H., Coelho, A. & Richter, J. Immersive learning experiences for understanding complex systems. In *iLRN 2019 London-Workshop, Long and Short Paper, Poster, Demos, and SSRiP Proceedings from the Fifth Immersive Learning Research Network Conference* 107–113 http://hdl.handle.net/10400.2/8368 (Verlag der Technischen Universität Graz, 2019).
61. *Annex 1 Checklist for Assessing the Level of Complexity of a Program* (International Initiative for Impact Evaluation), https://www.3ieimpact.org/sites/default/files/2021-07/ complexity-blg-Annex1-Checklist_assessing_level_complexity.pdf (2021).
62. Deaton, A. & Cartwright, N. Understanding and misunderstanding randomized controlled trials. *Soc. Sci. Med.* **210**, 2–21 (2018).
63. Bonell, C., Jamal, F., Melendez-Torres, G. J. & Cummins, S. 'Dark logic': theorising the harmful consequences of public health interventions. *J. Epidemiol. Community Health* **69**, 95–98 (2015).
64. Robinson, C. D., Chande, R., Burgess, S. & Rogers, T. Parent engagement interventions are not costless: opportunity cost and crowd out of parental investment. *Educ. Eval. Policy Anal.* **44**, 170–177 (2021).
65. Centola, D. *How Behaviour Spreads: The Science of Complex Contagions* (Princeton Univ. Press, 2018).
66. Kim, D. A. et al. Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. *Lancet* **386**, 145–153 (2015).
67. Berry, D. A. Bayesian clinical trials. *Nat. Rev. Drug Discov.* **5**, 27–36 (2006).
68. Marinelli, H. A., Berlinski, S. & Busso, M. Remedial education: evidence from a sequence of experiments in Colombia. *J. Hum. Resour.* 0320-10801R2 (2021).
69. Anders, J., Groot, B. & Heal, J. Running RCTs with complex interventions. *The Behavioural Insights Team* https://www.bi.team/blogs/running-rcts-with-complex-interventions/ (1 November 2017).
70. Volpp, K. G., Terwiesch, C., Troxel, A. B., Mehta, S. & Asch, D. A. Making the RCT more useful for innovation with evidence-based evolutionary testing. *Healthcare* **1**, 4–7 (2013).
71. Kidwell, K. M. & Hyde, L. W. Adaptive interventions and SMART designs: application to child behavior research in a community setting. *Am. J. Eval.* **37**, 344–363 (2016).
72. Caria, S., Kasy, M., Quinn, S., Shami, S. & Teytelboym, A. *An Adaptive Targeted Field Experiment: Job Search Assistance for Refugees in Jordan. Warwick Economics Research Papers No. 1335* (2021).
73. *The Complexity Evaluation Toolkit* v.1.0, https://www.cecan.ac.uk/wp-content/uploads/ 2020/08/EPPN-No-03-Agent-Based-Modelling-for-Evaluation.pdf (CECAN, 2021).
74. Schluter, M. et al. A framework for mapping and comparing behavioural theories in models of social–ecological systems. *Ecol. Econ.* **131**, 21–35 (2017).
75. Wijermans, N., Boonstra, W. J., Orach, K., Hentati-Sundberg, J. & Schlüter, M. Behavioural diversity in fishing—towards a next generation of fishery models. *Fish Fish.* **21**, 872–890 (2020).
76. Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366 (2011).
77. Nelson, L. D., Simmons, J. & Simonsohn, U. Psychology's Renaissance. *Annu. Rev. Psychol.* **69**, 511–534 (2018).
78. Frias-Navarro, D., Pascual-Llobell, J., Pascual-Soler, M., Perezgonzalez, J. & Berrios-Riquelme, J. Replication crisis or an opportunity to improve scientific production? *Eur. J. Educ.* **55**, 618–631 (2020).
79. Stanley, T. D., Carter, E. C. & Doucouliagos, H. What meta-analyses reveal about the replicability of psychological research. *Psychol. Bull.* **144**, 1325–1346 (2018).
80. Bryan, C. J., Tipton, E. & Yeager, D. S. Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nat. Hum. Behav.* **5**, 980–989 (2021).
81. Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J. & Reinero, D. A. Contextual sensitivity in scientific reproducibility. *Proc. Natl Acad. Sci. USA* **113**, 6454–6459 (2016).
82. Brenninkmeijer, J., Derksen, M., Rietzschel, E., Vazire, S. & Nuijten, M. Informal laboratory practices in psychology. *Collabra Psychol.* **5**, 45 (2019).
83. Landy, J. F. et al. Crowdsourcing hypothesis tests: making transparent how design choices shape research results. *Psychol. Bull.* **146**, 451 (2020).
84. McShane, B. B., Tackett, J. L., Böckenholt, U. & Gelman, A. Large-scale replication projects in contemporary psychological research. *Am. Stat.* **73**, 99–105 (2019).
85. Sanbonmatsu, D. M., Cooley, E. H. & Butner, J. E. The impact of complexity on methods and findings in psychological science. *Front. Psychol.* **11**, 580111 (2021).
86. Cartwright, N. & Hardie, J. *Evidence-Based Policy: A Practical Guide to Doing It Better* (Oxford Univ. Press, 2012).
87. Yeager, D. To change the world, behavioral intervention research will need to get serious about heterogeneity. *OSF* https://osf.io/zuh93/ (2020).
88. Snow, T. Mind the gap between the truth and data. *Nesta* https://www.nesta.org.uk/blog/mind-gap-between-truth-and-data/ (9 October 2019).
89. Damschroder, L. J. et al. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement. Sci.* **4**, 50 (2009).
90. Oberauer, K. & Lewandowsky, S. Addressing the theory crisis in psychology. *Psychon. Bull. Rev.* **26**, 1596–1618 (2019).
91. Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A. & Haig, B. D. Theory construction methodology: a practical framework for building theories in psychology. *Perspect. Psychol. Sci.* **16**, 756–766 (2021).
92. Sanbonmatsu, D. M. & Johnston, W. A. Redefining science: the impact of complexity on theory development in social and behavioral research. *Perspect. Psychol. Sci.* **14**, 672–690 (2019).
93. Fried, E. I. Theories and models: what they are, what they are for, and what they are about. *Psychol. Inq.* **31**, 336–344 (2020).
94. Muthukrishna, M. & Henrich, J. A problem in theory. *Nat. Hum. Behav.* **3**, 221–229 (2019).
95. Schimmelpfennig, R. & Muthukrishna, M. Cultural evolutionary behavioural science in public policy. *Behav. Public Policy* https://doi.org/10.1017/bpp.2022.40 (2023)
96. Kwan, V. S., John, O. P., Kenny, D. A., Bond, M. H. & Robins, R. W. Reconceptualizing individual differences in self-enhancement bias: an interpersonal approach. *Psychol. Rev.* **111**, 94 (2004).
97. Mezulis, A. H., Abramson, L. Y., Hyde, J. S. & Hankin, B. L. Is there a universal positivity bias in attributions? A meta-analytic review

of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychol. Bull.* **130**, 711 (2004).
98. Smets, K. There is more to behavioral economics than biases and fallacies. *Behavioral Scientist* http://behaviouralscientist.org/there- is-more-to-behavioural-science-than-biases-and-fallacies/ (24 July 2018).
99. Rand, D. G. Cooperation, fast and slow: meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychol. Sci.* **27**, 1192–1206 (2016).
100. Gelfand, M. J. *Rule Makers, Rule Breakers: How Tight and Loose Cultures Wire Our World* (Constable & Robinson, 2018).
101. West, R. et al. Development of a formal system for representing behaviour-change theories. *Nat. Hum. Behav.* **3**, 526–536 (2019).
102. Hale, J. et al. An ontology-based modelling system (OBMS) for representing behaviour change theories applied to 76 theories. *Wellcome Open Res* **5**, 177 (2020).
103. van Rooij, I. & Baggio, G. Theory before the test: how to build high-verisimilitude explanatory theories in psychological science. *Perspect. Psychol. Sci.* **16**, 682–697 (2021).
104. Smaldino, P. E. How to build a strong theoretical foundation. *Psychol. Inq.* **31**, 297–301 (2020).
105. Abner, G. B., Kim, S. Y. & Perry, J. L. Building evidence for public human resource management: using middle range theory to link theory and data. *Rev. Public Pers. Adm.* **37**, 139–159 (2017).
106. Moore, L. F., Johns, G. & Pinder, C. C. in *Middle Range Theory and the Study of Organizations* (eds Pinder, C. C. & Moore, L. F.) 1–16 (Martinus Nijhoff, 1980).
107. Berkman, E. T. & Wilson, S. M. So useful as a good theory? The practicality crisis in (social) psychological theory. *Perspect. Psychol. Sci.* **16**, 864–874 (2021).
108. Lieder, F. & Griffiths, T. L. Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behav. Brain Sci.* **43**, e1 (2020).
109. Callaway, F., Hardy, M. & Griffiths, T. Optimal nudging for cognitively bounded agents: a framework for modeling, predicting, and controlling the effects of choice architectures. Preprint at https://doi.org/10.31234/osf.io/7ahdc (2022).
110. Roese, N. J. & Vohs, K. D. Hindsight bias. *Perspect. Psychol. Sci.* **7**, 411–426 (2012).
111. Henriksen, K. & Kaplan, H. Hindsight bias, outcome knowledge and adaptive learning. *Qual. Saf. Health Care* **12**, ii46–ii50 (2003).
112. Bukszar, E. & Connolly, T. Hindsight bias and strategic choice: some problems in learning from experience. *Acad. Manage. J.* **31**, 628–641 (1988).
113. DellaVigna, S., Pope, D. & Vivalt, E. Predict science to improve science. *Science* **366**, 428–429 (2019).
114. Munnich, E. & Ranney, M. A. Learning from surprise: harnessing a metacognitive surprise signal to build and adapt belief networks. *Top. Cogn. Sci.* **11**, 164–177 (2019).
115. Deshpande, M. & Dizon-Ross, R. *The (Lack of) Anticipatory Effects of the Social Safety Net on Human Capital Investment* Working Paper, https://faculty.chicagobooth.edu/-/media/faculty/rebecca-dizon-ross/research/ssi_rct.pdf (Chicago Booth, 2022).
116. DellaVigna, S. & Linos, E. RCTs to scale: comprehensive evidence from two nudge units. *Econometrica* **90**, 81–116 (2022).
117. Dimant, E., Clemente, E. G., Pieper, D., Dreber, A. & Gelfand, M. Politicizing mask-wearing: predicting the success of behavioral interventions among Republicans and Democrats in the U.S. *Sci. Rep.* **12**, 7575 (2022).
118. Ackerman, R., Bernstein, D. M. & Kumar, R. Metacognitive hindsight bias. *Mem. Cogn.* **48**, 731–744 (2020).
119. Pezzo, M. Surprise, defence, or making sense: what removes hindsight bias? *Memory* **11**, 421–441 (2003).
120. Dorison, C. A. & Heller, B. H. Observers penalize decision makers whose risk preferences are unaffected by loss–gain framing. *J. Exp. Psychol.* **151**, 2043–2059 (2022).
121. Porter, T. et al. Predictors and consequences of intellectual humility. *Nat. Rev. Psychol.* **1**, 524–536 (2022).
122. Egan, M., Hallsworth, M., McCrae, J. & Rutter, J. *Behavioural Government: Using Behavioural Science to Improve How Governments Make Decisions* (Behavioural Insights Team, 2018).
123. Walton, G. M. & Wilson, T. D. Wise interventions: psychological remedies for social and personal problems. *Psychol. Rev.* **125**, 617 (2018).
124. Lewis, N. A. Jr What counts as good science? How the battle for methodological legitimacy affects public psychology. *Am. Psychol.* **76**, 1323 (2021).
125. Lamont, M., Adler, L., Park, B. Y. & Xiang, X. Bridging cultural sociology and cognitive psychology in three contemporary research programmes. *Nat. Hum. Behav.* **1**, 866–872 (2017).
126. Vaisey, S. Motivation and justification: a dual-process model of culture in action. *Am. J. Sociol.* **114**, 1675–1715 (2009).
127. Swidler, A. Culture in action: symbols and strategies. *Am. Sociol. Rev.* **51**, 273–286 (1986).
128. Richardson, L. & John, P. Co-designing behavioural public policy: lessons from the field about how to 'nudge plus'. *Evid. Policy* **17**, 405–422 (2021).
129. Banerjee, S. & John, P. Nudge plus: incorporating reflection into behavioral public policy. *Behav. Public Policy*, https://doi.org/10.1017/bpp.2021.6 (2021).
130. Reijula, S. & Hertwig, R. Self-nudging and the citizen choice architect. *Behav. Public Policy* **6**, 119–149 (2022).
131. Hertwig, R. & Grüne-Yanoff, T. Nudging and boosting: steering or empowering good decisions. *Perspect. Psychol. Sci.* **12**, 973–986 (2017).
132. Hertwig, R. When to consider boosting: some rules for policy-makers. *Behav. Public Policy* **1**, 143–161 (2017).
133. Grüne-Yanoff, T., Marchionni, C. & Feufel, M. A. Toward a framework for selecting behavioural policies: how to choose between boosts and nudges. *Econ. Phil.* **34**, 243–266 (2018).
134. Sims, A. & Müller, T. M. Nudge versus boost: a distinction without a normative difference. *Econ. Phil.* **35**, 195–222 (2019).
135. Sunstein, C. R. *Choosing Not to Choose: Understanding the Value of Choice* (Oxford Univ. Press, 2015).
136. Miller, G. A. Psychology as a means of promoting human welfare. *Am. Psychol.* **24**, 1063 (1969).
137. Bason, C. *Leading Public Sector Innovation: Co-creating for a Better Society* (Policy Press, 2018).
138. Big-data studies of human behaviour need a common language. *Nature* **595**, 149–150 (2021).
139. Yarkoni, T. & Westfall, J. Choosing prediction over explanation in psychology: lessons from machine learning. *Perspect. Psychol. Sci.* **12**, 1100–1122 (2017).
140. Künzel, S. R., Sekhon, J. S., Bickel, P. J. & Yu, B. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proc. Natl Acad. Sci. USA* **116**, 4156–4165 (2019).
141. Wager, S. & Athey, S. Estimation and inference of heterogeneous treatment effects using random forests. *J. Am. Stat. Assoc.* **113**, 1228–1242 (2018).
142. Todd-Blick, A. et al. Winners are not keepers: characterizing household engagement, gains, and energy patterns in demand response using machine learning in the United States. *Energy Res. Soc. Sci.* **70**, 101595 (2020).

143. Mills, S. Personalized nudging. *Behav. Public Policy* **6**, 150–159 (2022).
144. Soman, D. & Hossain, T. Successfully scaled solutions need not be homogenous. *Behav. Public Policy* **5**, 80–89 (2021).
145. Möhlmann, M. Algorithmic nudges don't have to be unethical. *Harvard Business Review,* 22 April (2021).
146. Susser, D., Roessler, B., & Nissenbaum, H. Online manipulation: hidden influences in a digital world. *4 Georget. Law Technol. Rev.* 1 (2019).
147. Abbasi, M., Fridler, A., Schneidegger, C. & Venkatasubramanian, S. Fairness in representation: quantifying stereotyping as a representational harm. In *SIAM International Conference on Data Mining*, *SDM 2019,* (eds. Berger-Wolf, T. & Chawla, N.) 801–809 (Society for Industrial and Applied Mathematics, 2019).
148. Eubanks, V. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's, 2018).
149. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **366**, 447–453 (2019).
150. Kozyreva, A., Lorenz-Spreen, P., Hertwig, R., Lewandowsky, S. & Herzog, S. M. Public attitudes towards algorithmic personalization and use of personal data online: evidence from Germany, Great Britain, and the United States. *Hum. Soc. Sci. Commun.* **8**, 117 (2021).
151. Kotamarthi, P. This is personal: the do's and don'ts of personalization in tech. *Decision Lab* https://thedecisionlab.com/insights/technology/this-is-personal-the-dos-and-donts-of-personalization-in-tech (2022).
152. Matz, S. C., Kosinski, M., Nave, G. & Stillwell, D. J. Psychological targeting as an effective approach to digital mass persuasion. *Proc. Natl Acad. Sci. USA* **114**, 12714–12719 (2017).
153. Pierson, E., Cutler, D. M., Leskovec, J., Mullainathan, S. & Obermeyer, Z. An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nat. Med.* **27**, 136–140 (2021).
154. Lorenz-Spreen, P. et al. Boosting people's ability to detect microtargeted advertising. *Sci. Rep.* **11**, 15541 (2021).
155. Nagel, T. *The View from Nowhere* (Oxford Univ. Press, 1986).
156. Sugden, R. The behavioural economist and the social planner: to whom should behavioural welfare economics be addressed? *Inquiry* **56**, 519–538 (2013).
157. Liscow, Z. D. & Markovits, D. Democratizing behavioural economics. *Yale J. Regul.* **39**, 1217–1290 (2022).
158. Bergman, P., Lasky-Fink, J. & Rogers, T. Simplification and defaults affect adoption and impact of technology, but decision makers do not realize it. *Organ. Behav. Hum. Decis. Process.* **158**, 66–79 (2020).
159. Pereira, M. M. Understanding and reducing biases in elite beliefs about the electorate. *Am. Polit. Sci. Rev.* **115**, 1308–1324 (2021).
160. Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D. & Mortenson, E. Racial inequality in psychological research: trends of the past and recommendations for the future. *Perspect. Psychol. Sci.* **15**, 1295–1309 (2020).
161. Lepenies, R. & Małecka, M. in *Handbook of Behavioural Change and Public Policy* (eds Beck, S. & Straßheim, H.) 344–360 (Edward Elgar, 2019).
162. Common Thread. From Idea to Immunization:A Blueprint to Building a BI Unit in the Global South https://gocommonthread.com/work/global-gavi/bi (2022).
163. Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D. & Majid, A. Over-reliance on English hinders cognitive science. *Trends Cogn. Sci.* **26**, 1153–1170 (2022).
164. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–83 (2010).
165. Cheon, B. K., Melani, I. & Hong, Y. Y. How USA-centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. *Soc. Psychol. Pers. Sci.* **11**, 928–937 (2020).
166. Dupree, C. H. & Kraus, M. W. Psychological science is not race neutral. *Perspect. Psychol. Sci.* **17**, 270–275 (2022).
167. Mullainathan, S. Keynote address to the Society of Judgment and Decision Making Annual Conference (2022).
168. *A Guidebook for Community Organizations, Researchers, and Funders to Help Us Get from Insufficient Understanding to More Authentic Truth* https://chicagobeyond.org/researchequity/ (Chicago Beyond, 2018).
169. Asman, S., Casarotto, C., Duflo, A. & Rajkotia, R. Locally-grounded research: strengthening partnerships to advance the science and impact of development research. *Innovations for Poverty Action* https://www.poverty-action.org/blog/locally-grounded-research-strengthening-partnerships-advance-science-and-impact-development (28 September 2021).
170. *The PhD Project,* https://phdproject.org/ (PhD Project, accessed 9 December 2022).
171. Erosheva, E. A. et al. NIH peer review: criterion scores completely account for racial disparities in overall impact scores. *Sci. Adv.* **6**, eaaz4868 (2020).
172. Marteau, T. M. et al. Judging nudging: can nudging improve population health? *BMJ* **342**, d228 (2011).
173. Lambe, F. et al. Embracing complexity: a transdisciplinary conceptual framework for understanding behavior change in the context of development-focused interventions. *World Dev.* **126**, 104703 (2020).
174. Shrout, P. E. & Rodgers, J. L. Psychology, science, and knowledge construction: broadening perspectives from the replication crisis. *Annu. Rev. Psychol.* **69**, 487–510 (2018).
175. IJzerman, H. et al. Use caution when applying behavioural science to policy. *Nat. Hum. Behav.* **4**, 1092–1094 (2020).
176. Grüne-Yanoff, T. Old wine in new casks: libertarian paternalism still violates liberal principles. *Soc. Choice Welf.* **38**, 635–645 (2012).
177. Rizzo, M. J. & Whitman, G. *Escaping Paternalism: Rationality, Behavioral Economics, and Public Policy* (Cambridge Univ. Press, 2020).
178. Ewert, B. Moving beyond the obsession with nudging individual behaviour: towards a broader understanding of behavioural public policy. *Public Policy Adm.* **35**, 337–360 (2020).
179. Leggett, W. The politics of behaviour change: nudge, neoliberalism and the state. *Policy Polit.* **42**, 3–19 (2014).

Acknowledgements

I thank L. Tublin for her editorial support. I also thank S. Banerjee, E. Berkman, A. Buttenheim, F. Callaway, J. Collins, J. Doctor, A. Gyani, D. Halpern, P. John, T. Marteau, M. Muthukrishna, D. Perera, D. Perrott, K. Ruggeri, R. Schmidt, D. Soman, H. Strassheim, C. Sunstein and members of the Behavioural Insights Team for their feedback on previous drafts.

Competing interests

The author is the managing director, Americas, at the Behavioural Insights Team, which provides consultancy services in behavioural science.

Additional information

Correspondence should be addressed to Michael Hallsworth.

Peer review information *Nature Human Behaviour* thanks Peter John and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.