A new manifesto for applying behavioral science

Michael Hallsworth, PhD

THE
BEHAVIORAL
INSIGHTS
TEAM

Where I’m coming from

2009 – 2010	MINDSPACE report for UK government (1,600 cites)	Dolan et al., 2012, *J.EconPsych*
2010 - 2011	Major study on policy-making	Hallsworth, 2011, *Political Insight,* 4.
2011 – present	Leader for The Behavioral Insights Team, running many field RCTs and giving policy advice to senior leaders	Hallsworth et al. 2015 *PLoS One* Hallsworth et al. 2016, *Lancet* Hallsworth et al. 2017, *J.Pub.Econ*
2014	EAST framework for policymakers	BIT 2014
2018	Behavioral Government framework	BIT 2018
2020	*Behavioral Insights* summary book	MIT Press
2018 -	Where are we going?	Sanders et al., 2018, *Behavioral Public Policy*

Criticisms of applied behavioral science

- Limited impact
- Mechanistic thinking
- Control paradigm
- Neglect of the social context
- Flawed evidence base
- Lack of precision
- Overconfidence
- Failure to reach scale
- Application over innovation
- Ethical concerns

After a decade of the current phase of applied behavioral science, we need **a new manifesto** to address these challenges and guide our practices

Use behavioral science as a lens

See the system

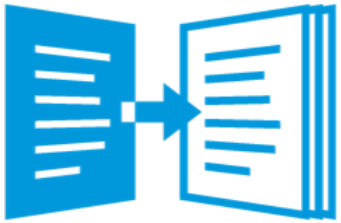
Put RCTs in their place

Be humble, explore and enable

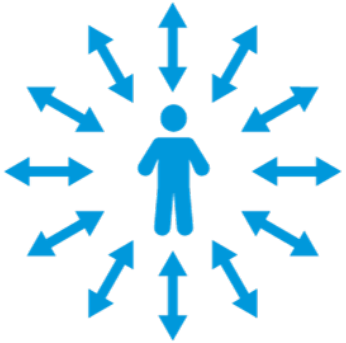
No “view from nowhere”

Predict and adjust

Build behavioral science into organizations



Replication, variation and adaptation



Data science for equity

Beyond lists of biases

Replication, variation and adaptation

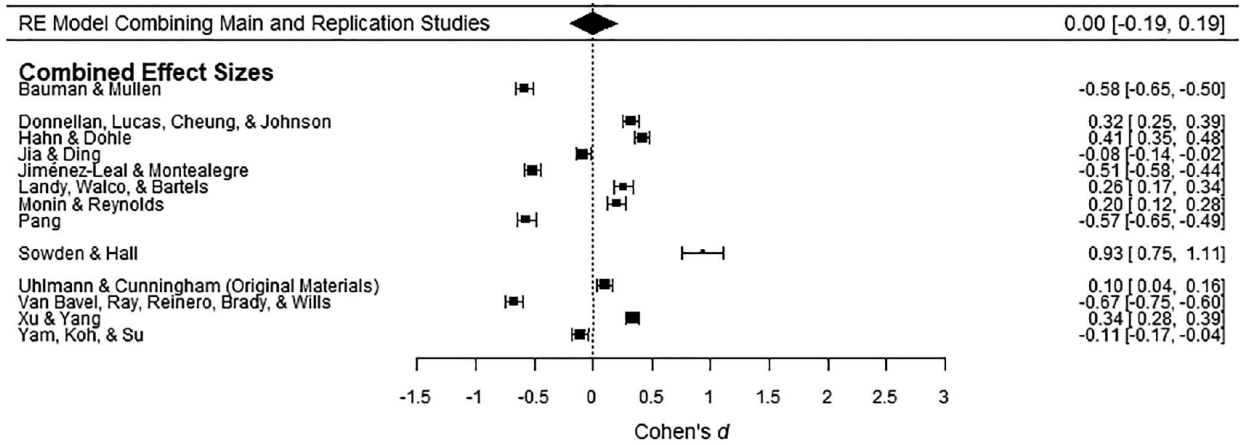
Contexts influence results

Experiments make myriad choices about things like the precise wording of messages, the time of day they are given, how participants are recruited, and so on. These choices vary greatly between studies and experimenters, in ways that often go unnoticed – even for replication studies.

- 15 research teams
- Each asked to create interventions to test the same five hypotheses (all unpublished at the time)
- Hypotheses concern moral judgment, negotiations, implicit cognition
- 15,000 participants in two separate samples then randomly assigned to receive one of the interventions for each hypothesis

Effect sizes varied dramatically. For four of the five research questions, different teams' interventions 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'.



Landy, J. F., & Crowdsourcing Hypothesis Tests Collaboration. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin*, *146*(5), 451.

Effects vary between groups in a population

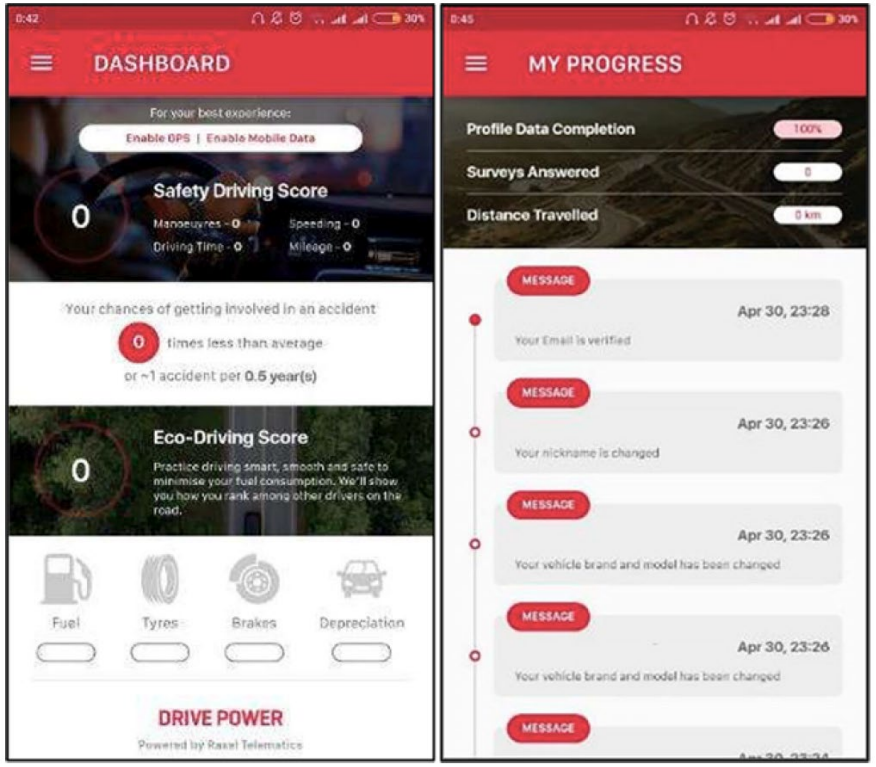
Increasingly revealed by machine learning.

Study in India used a smartphone app to give drivers feedback on their driving – three different kinds of messages.

“Personal best” nudge better for high-performance drivers who did not seek feedback often; “average driver” nudge worked best for low-performance drivers who sought feedback often.

Call for a “heterogeneity revolution: in behavioral sciences: most effects vary, effect doesn’t need to hold across all groups to be important or real.

This challenges approach that prioritizes achieving overall marginal shifts in the behavior of large populations.



Choudhary, V., Shunko, M., Netessine, S., & Koo, S. (2022). Nudging drivers to safety: Evidence from a field experiment. *Management Science*, *68*(6), 4196-4214.

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature human behaviour*, *5*(8), 980-989.

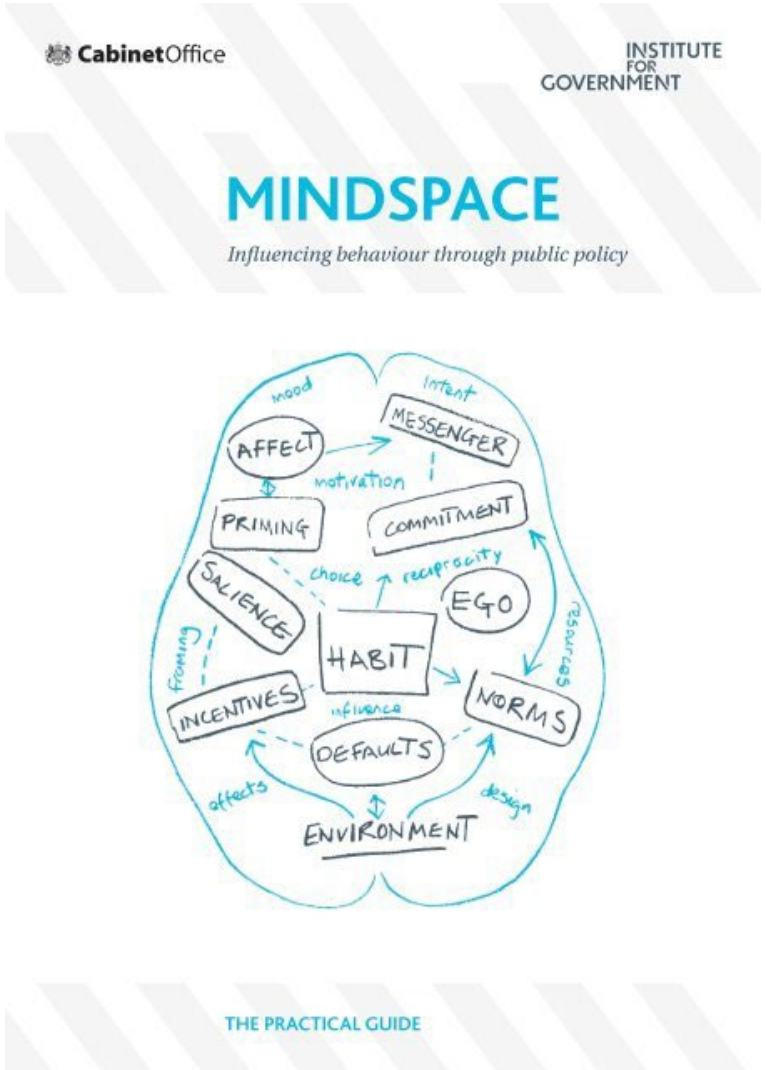
Make more conservative claims

Applied behavioral scientists need to set a higher bar for claiming that an effect holds true across many unspecified settings.

"we are strongly influenced by what others do"

Growing sense that interventions should be talked about as *hypotheses* that were true in one place, and which may need adapting for them to be true elsewhere as well.

Parallel to the way behavioral scientists have talked about incentives: impact "***clearly depends on factors such as the type, magnitude and timing of the incentive.***"



Expand contexts & participants, test systematically

Expand studies to include & examine a wider range of contexts and participants - and gather richer data about them.

Coordinated multi-site studies will be needed to collect enough data to explore heterogeneity **systematically** – e.g., BIT and hospital navigators.

“Crowdsourced” studies offer particular promise for testing context, methods, and expert judgment.

The truth is that collecting varied data better is going to be hard – realistically, academia will need to make the investments in research infrastructure (panels, coordination, data sharing).

Table 2
Effect Sizes and Q, I^2, and τ^2 Statistics From Meta-Analyses of Main Studies a

Hypothesis	Description	k	Effect size [95% CI]
			Main studies
1	Awareness of automatic prejudice	13	d = .07 [−0.22, 0.37]
2	Extreme offers reduce trust	13	d = 1.04 [0.61, 1.47]
3	Moral praise for needless work	13	d = .33 [0.17, 0.50]
4	Proximal authorities drive legitimacy of performance enhancers	12	d = .07 [−0.05, 0.20]
5	Deontological judgments predict happiness	13	r = .06 [0.01, 0.11]
			Replication studies
1	Awareness of automatic prejudice	13	d = −0.07 [−0.33, 0.19]
2	Extreme offers reduce trust	13	d = .61 [0.32, 0.88]
3	Moral praise for needless work	13	d = .24 [0.11, 0.38]
4	Proximal authorities drive legitimacy of performance enhancers	12	d = .03 [−0.06, 0.12]
5	Deontological judgments predict happiness	13	r = .03 [−0.04, 0.09]

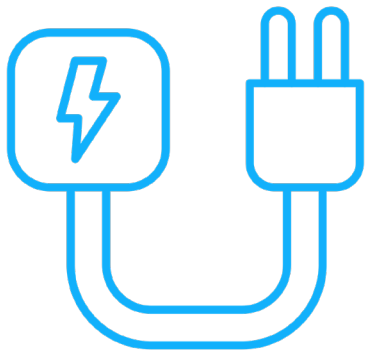
*** $p < .001$.

Landy, J. F., & Crowdsourcing Hypothesis Tests Collaboration. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin*, *146*(5), 451.

Be humble, explore and enable

Drivers of overconfidence in behavioral science

Behavioral science may provide a technical justification for seeing decisions as flawed, and thus needing corrective action.



- ‘Internalities’ / ‘present bias’ leading policy makers to prioritize the future self. But empirical evidence showing how economic environments increase discount rate.*
- Loss-gain framing effects may reflect awareness of punishment by others.§
- Anchoring effects = people trading off the cost of being wrong against the cost of taking time to get the answer more right.‡

The ready technical explanation offered by behavioral science could provide confidence that obscures the need to search more deeply for less obvious explanations.

These explanations may be based on overly cognitive and individual approach, with less attention to culture and society.

* Ruggeri, K., Panin, A., Vdovic, M. et al. The globalizability of temporal discounting. Nat Hum Behav (2022).
§ Dorison, C. A., & Heller, B. H. Observers penalize decision makers whose risk preferences are unaffected by loss-gain framing. Journal of Experimental Psychology: General
‡ Lieder, F., et al. (2018). Empirical evidence for resource-rational anchoring and adjustment. Psychonomic Bulletin & Review, 25(2), 775-784.

Recommendations

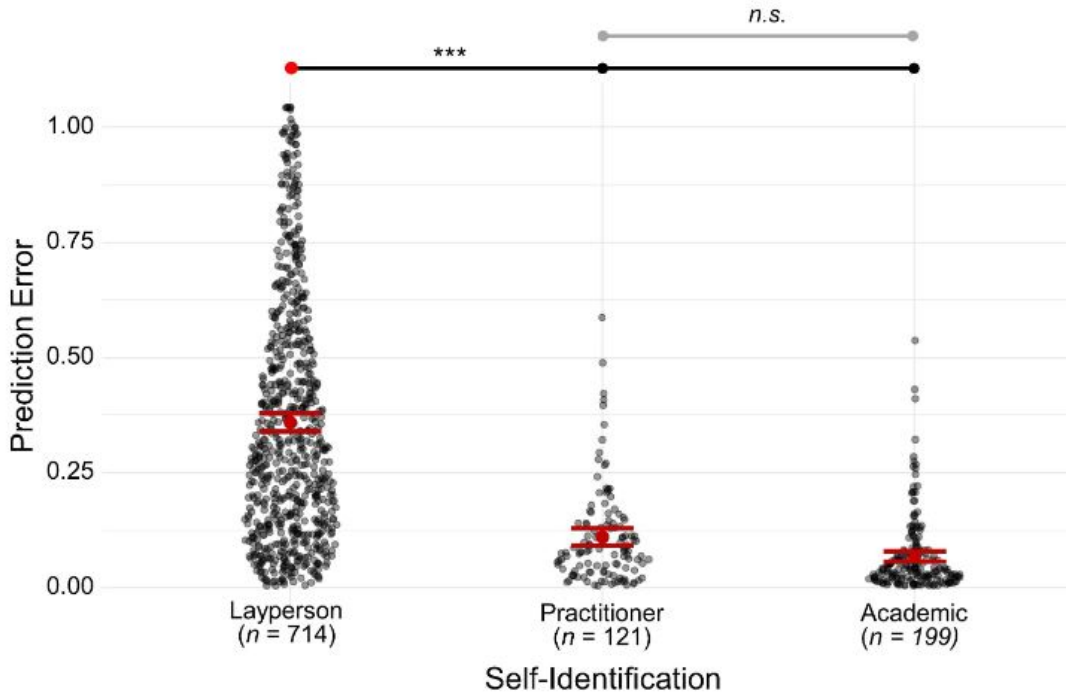
Avoid using the term **“irrationality”**, which can limit attempts to understand actions in context.

Acknowledge that our diagnoses of behavior are provisional and incomplete (**“epistemic humility”**).

Design processes and **institutions** to counteract overconfidence.

- Building predictions and feedback loops into standard processes, to counter hindsight bias.
- Pre-mortems and ‘dark logic’ exercises
- Expanded reference class forecasting from infrastructure to social policy
- Build in break points
- Require two estimates

Hypothesis 6: Accuracy of predictions of nudge effects differs among laypeople, academics, and practitioners



Dimant, E., Clemente, E. G., Pieper, D., Dreber, A., & Gelfand, M. (2022). Politicizing mask-wearing: predicting the success of behavioral interventions among republicans and democrats in the US. *Scientific Reports*, *12*(1), 1-12.

A **new manifesto** for applied behavioral science

Use behavioral science as a lens

See the system

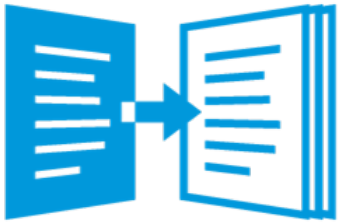
Put RCTs in their place

Be humble, explore and enable

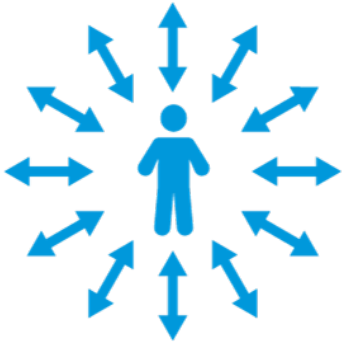
No “view from nowhere”

Predict and adjust

Build behavioral science into organizations

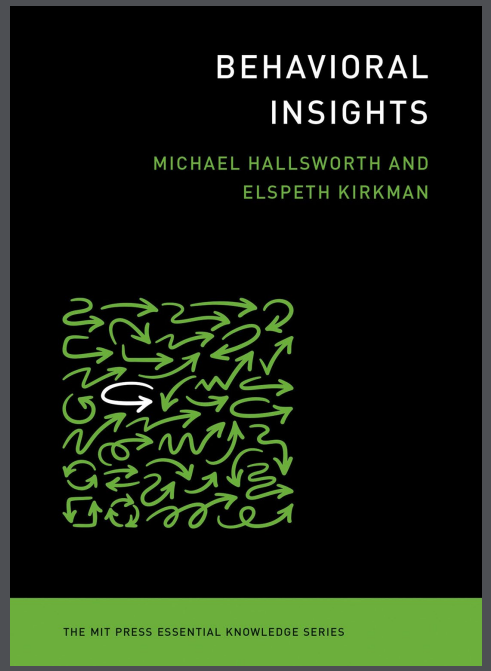


Replication, variation and adaptation



Data science for equity

Beyond lists of biases



michael.hallsworth@bi.team

@mhallsworth

www.michaelhallsworth.com