

# What works: When & why are nudges sticky, scalable and transferable?

Steven Johnson<sup>1\*</sup>

## Abstract

Evidence-based policy making assumes that the evidence in question has external validity i.e. that the conclusions drawn from research can be generalised to contexts outside of the research setting. In other words, that those conclusions are sticky, scalable and transferable.

The use of so-called 'nudge'-style behavioural interventions has increased considerably over the last decade in both public policy and commercial design contexts. This has put pressure on a fledgling evidence-base to answer questions of external validity with premature confidence. In turn, this can lead to replicability failures and the risk losing public confidence and implicit consent for the use of these approaches.

This paper argues that (at present) the only responsible answer to 'what works?' questions in relation nudges is that 'we don't know... yet'. It then goes on to argue that this is a perfectly natural and valid state for a young discipline to be in and points out the key areas that future research and practice needs to address to facilitate more confident and actionable claims of external validity.

<sup>1</sup> Chief Behavioural Officer, Voicescape

\*Corresponding author: [steven.johnson@voicescape.co.uk](mailto:steven.johnson@voicescape.co.uk)

"Now is not then, here not there...".

THOMAS MANN

## Introduction

The external validity of research findings lies at the foundation of evidence-based policy making. An understanding of the extent to which evidence can be applied beyond the research setting is essential if we are to apply that evidence to address real-world challenges.

The use of so-called 'nudge'-style behavioural interventions has increased considerably over the last decade. Classically, a nudge-style intervention consists in making small changes to the context in which a decision or behaviour occurs, with the intention of making a desired behaviour more likely. In the decision science literature, these contexts are referred to as 'choice architectures'.

Public acceptance of these techniques is essential if we are to unlock their full potential. Given that ethical concerns relating to 'nudging' have already made it into public consciousness, the potential reputational damage caused by overconfidence in the evidence and a resultant failure of replicability could be fatal for the discipline.

Considering this, if we are to protect the fledgling evidence-base associated with the application of nudges and behavioural insights more generally, we need to resist pressure to respond in inappropriately categorical terms and with premature confidence to 'what works?' questions and take proactive steps to

accelerate the integrity of the evidence base.

There are various areas to be addressed if we are to build confidence in the external validity of nudge-style research findings. These relate primarily to the operational definition of the variables at play. If these definitions fail to be clear, stable and consistent then we run the risk of merely collating a series of disparate examples rather than diligently strengthening a pillar of evidence.

Firstly, there is still confusion as to what we mean by a 'nudge'. The original Thaler and Sunstein definition is widely accepted to be difficult to operationalise and not conducive to development of a rigorous evidence base. However, despite the work that has been done subsequently, key confusions remain.

Many simply defer the definition to the notion 'choice architecture' without adequately defining the latter. Others try to narrow the parameters by incorporating specific reference to the physical environment or to the unconscious nature of the effect; others apply the concept only to specific domains such as health or sustainability for a similar purpose.

This is more than mere semantic quibbling. If the concept of a nudge eludes definition, we may be forced to admit that it isn't a robust unit of analysis in the first place, but merely a loose label reified by its coverage in popular literature and media. In turn, this makes the whole enterprise to establish an evidence base for the application of nudges redundant.

A second source of definitional confusion relates to the sorts of effect nudges are expected to have. We are working

with multiple definitions of effectiveness and multiple versions of what needs to obtain for an intervention to be deemed to have ‘worked’.

In particular, there is no reliable distinction being drawn between choices, decisions, one-off behaviours and sustained behaviour change. Nudge trials are being deployed to address all four, with no clear distinction being drawn between what are clearly separate (if overlapping) constructs.

One of the key confusions here results from the conflation of cognitive and volitional mechanisms: the former relating the application of ‘reason’ and the latter to the exertion of will power or ‘self-regulation’. In turn, this brings us to the third area of concern: theory.

One of the key criticisms of behavioural insights and behavioural economics is the alleged theory vacuum that sits at its heart. The appearance of a cohesive body of knowledge is merely an emergent property of disparate collection of psychological mechanisms we call heuristics and biases, with an absence of a cohesive theory to either explain or predict the phenomena triggered by its research studies.

Researchers have attempted to resist this criticism by grounding the evidence in Dual Systems Theories and their variants. This approach assumes that there are two cognitive systems at play in human thought and action: one slow and deliberative, the other quick and associative. Situations in which the former is required, but the latter dominates, make us prone to errors resulting in sub-optimal outcomes. Many nudge-style interventions, based as they are on Dual System Theories, aim to activate or deactivate these heuristics and biases in order to make positive behaviours more likely.

However, even if dual-system theories make sense of *cognitive* effects, they have limited use if sub-optimal behaviours result from *volitional* failures. If we consider the key real-world behaviour change domains, such as health and environment, the impacts we seek are reliant on sustained behaviour change. Sustained behaviour change is difficult because of volitional, rather than cognitive failure: limitations of self-regulation in the face of strong undercurrents of habit and identity that go untouched by surface nudges.

Whilst most of the issues raised by this paper are theoretical and/or conceptual, we now turn to a key methodological concern and the practical steps that could be taken to address it.

One of the key steps forward that nudging and the popularisation of behavioural economics has triggered is a wider recognition of the extent to which human behaviour is contextually-driven rather than (just) self-determined. However, when it comes to questions of external validity in real world contexts, tensions begin to emerge between theory and practice.

The first relates to the apparent contradiction contained in the notion of a ‘universal context-dependent principle’ – a contradiction that the wider literature on the psychology of human behaviour has spent considerable time and effort reframing as a resolvable paradox, rather than an out-and-out contradiction.

If nudging consists in making small changes to contexts, we acknowledge the power of context as a determinant of human behaviour. However, in our desire for external validity – for nudges to be sticky, transferable and scaleable – we also demand that those nudges have consistent effects *across* contexts.

As with all systematic research, there is a desire to isolate and quantify the impact of the independent variable on the dependent by controlling for confounders – in effect removing the influence of context. However, in doing so we begin to betray our commitment to behaviour being a context-dependent phenomenon.

Determining which elements of context are noise to be removed and which are parts of the signal being detected is a perennial methodological and epistemological challenge for scientific research, but one that most nudge-style studies fail to recognise, let alone rise to.

One of the most common expressions of this failure is the design and analysis of RCTs powered to detect only overall effects rather than differential effects. That is, trials that ascertain which of a suite of interventions worked better *overall*, without trying to understand which intervention *worked better for whom*.

To use overall impact as an evidence-base to scale up and roll-out a single intervention is to overlook the fact that there will be identifiable groups of people for whom that intervention is not only ineffective, but actively results in negative outcomes. This move towards understanding differential rather than just overall impacts represents a simple step forward we could take to better understand the modulating role of context on the effectiveness of behavioural intervention.

It could be argued that this is already catered for within the existing conceptual framework through through notions such as ‘choice architectures’ and ‘micro-environments’ and the tendency to look at the effectiveness of nudges in specific domains such as health and environment. At one level, these are defined contexts in which we can ascertain the effectiveness of nudge interventions and expect those effects to endure in other similar contexts.

However, these constructs are nothing like robust enough to deal with real-world questions of external validity. When applied in the real world, or even as part of field-experiments, nudges occur within a spatio-temporal snapshot rather than a fixed and predictable environment: a *nano-context* created through an interplay between the agent, and the physical and social environment and the values to which those variables are set *at that moment in time*<sup>1</sup>.

<sup>1</sup>As philosopher of science Nancy Cartwright has argued, “It’s a very long road from it ‘worked somewhere’ to... it will work here’ and it is not an easy one to traverse”. Her work on moving towards an epistemology based on ‘fixed contributions’ from ‘stable tendencies’ within a context, rather than simple cause and effect provides useful direction”.

## Conclusion

As we have seen, if we are to make significant progress in our desire to strengthen the evidence-base in relation to the application of nudges there are number of fundamental areas which require attention.

If indeed choices and behaviours are driven by context at the ‘nano’ level we are faced with significant implications for the stickiness, scalability and transferability of nudges and even for ontologically validity of ‘nudge’ as a construct.

As is often the case, the key to strengthen the evidence base in an emerging area is time. Inevitably, many of the questions raised by this paper will be answered organically through the accumulation of more and different studies, issues around publication bias notwithstanding.

However, there are at least two more definite, proactive steps that can be taken by practitioners and academics alike to accelerate this natural evolution:

## Operational variables

- Use the existing literature to define what is meant by a nudge. If an individual study requires its own definition, then its focus is likely to be something other than nudge.
- Ensure that your intervention actually conforms to the definition invoked.
- Be clear on what category of effect the intervention is designed to elicit prior to the study or trial: a choice, a decision, a one-off behaviour, a sustained behaviour.

## Differential impact

- Where possible, power trials to detect differential as well as overall impact.
- Treat each arm of the trial as a sub-sample and test for variance in impact within that sub-sample.
- Seek to identify participants for whom there was no, little or even negative impact in those sub-samples and test for their significance.

## References

- Cartwright, N. (2011). Predicting ‘it will work for us’: (way) beyond statistics. In R. A. Klein, et al. G. J. Hollands, et al. *Causality in the Sciences*. Oxford University Press.
- Halpern, D. (2015). *Inside the nudge unit: How small changes can make a big difference*. Random House.
- Hollands, G. J., et al. (2013). Altering micro-environments to change population health behaviour: towards an evidence base for choice architecture interventions. *BMC public health*, 13(1), 1218.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Klein, R. A., et al. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443–490.
- Lin, Y., Osman, M., & Ashcroft, R. (2017). Nudge: Concept, effectiveness, and ethics. *Basic and Applied Social Psychology*, 39(6), 293–306.
- Marteau, T. M., Ogilvie, D., Roland, M., Suhrcke, M., & Kelly, M. P. (2011). Judging nudging: can nudging improve population health? *BMJ*, 342(d228). doi.org/10.1136/bmj.d228.
- Observer (2018). Why you need a chief behavioural officer. [observer.com/2018/01/why-you-need-chief-behavioral-officer-business-innovation-startups/](http://observer.com/2018/01/why-you-need-chief-behavioral-officer-business-innovation-startups/).
- OECD (2017). *Behavioural Insights and Public Policy: Lessons from Around the World*. OECD Publishing, Paris. doi.org/10.1787/9789264270480-en.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors’ introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528–530.
- Pykett, J., & Johnson, S. (2015). Silver Bullets Need a Careful Aim: Dilemmas in Applying Behavioural Insights. *ESRC Seminar on Behaviour Change and Psychological Governance*, London, May. Vol. 11.
- Schubert, C. (2017). Green nudges: Do they work? Are they ethical? *Ecological Economics*, 132, 329–342.
- Science and Technology Committee (2011). House of Lords- Science and Technology Select Committee 2nd Report of Session 2010-12–Behaviour Change. *House of Lords Paper*, 179.
- Sunstein, C. R. (2016a). Do People Like Nudges? *Administrative Law Review*. Forthcoming. doi.org/10.2139/ssrn.2604084.
- Sunstein, C. R. (2015). The ethics of nudging. *Yale Journal on Regulation*, 32, 413.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.
- Wansink, B. (2004). Environmental factors that increase the food intake and consumption volume of unknowing consumers. *Annual Review of Nutrition*, 24, 455–479.