



Experimental validation bias limits the scope and ambition of applied behavioural science

David Gal¹✉ and Derek D. Rucker²

Applied behavioural science tends to overvalue interventions that can be readily tested using experiments. This experimental validation bias drives the popularity of light interventions and nudges and unnecessarily limits the scope and ambition of the field.

Recent years have witnessed significant excitement over real-world applications of psychology, traced to the idea that behavioural science can identify light-touch interventions, or ‘nudges’, to influence behaviour. Central to the nudge approach is its emphasis on simple, relatively superficial interventions over fundamental or structural change. For example, as opposed to increasing retirement security through expanding social security programs, a nudge might take the form of an opt-out program that directs some of individuals’ own income to a retirement savings account by default¹.

The appeal of nudges has been far reaching. ‘Nudge units’ have been formed in various governments, agencies and firms. For example, the UK government created a ‘Behavioural Insights Team’ and the US White House formed a ‘Social and Behavioral Science Team’ to design nudges for public policy aims.

We suggest that the dominance of the nudge approach in applied behavioural science is largely due to experimental validation bias, that is, the tendency to overvalue interventions that can be ‘validated’ by experiments. This bias results in interventions of limited ambition and scope, leading to an impoverished view of the relevance of behavioural science to the real world.

The limited impact of nudges

Despite their initial promise, some of the most heralded nudges from academic work have yielded limited impact when examined rigorously and at scale. For instance, an analysis of 126 randomized control trials of behavioural interventions carried out by the two largest US government ‘nudge units’ reported an average increase in adoption of the targeted behaviour of only 1.38% (REF.²). This is a small effect in an absolute sense, and it is particularly modest in comparison to a control adoption rate of 17.44%. Furthermore, it is one-sixth the size of the average effect from academic studies and less than one-quarter of the average effect estimate (5.8%) that researchers expected the nudges would yield. Notably, the most successful nudges were

also the most intuitive: they involved sending reminders and notifications.

Certainly, common-sense information design and clear communication can be highly beneficial to the public, and nudge interventions of these sorts might be economically worthwhile in some situations. However, it is discouraging that sending reminders and notifications currently passes as the main contribution of behavioural science to policy. Little behavioural insight is required to design these interventions, which are simply about making information more visible/salient.

A lightness-impact tradeoff

Most real-world behaviour reflects interactions among many, often shifting, variables that are difficult to anticipate or predict^{3–5}. Thus, nudges will tend to result in weak effects that get drowned out and/or unpredictably altered by the many background factors that exist in real-world environments.

Paradoxically, this limitation of the nudge approach — the emphasis on relatively light-touch interventions — is also the source of its appeal. As indicated, it aligns with what we term experimental validation bias — behavioural scientists’ tendency to overvalue interventions that can be readily tested using experiments (relative to more involved interventions with a potentially higher impact). Nudges are relatively easy to test experimentally and therefore match up with this bias.

Conversely, the sort of interventions likely to be impactful, namely those involving multifaceted and/or sustained structural change, are not readily tested using experiments. In many cases, experimental tests are not even possible owing to the scope and complexity of the intervention, as well as the sui generis and evolving nature of situations. For example, the impact of most of the policies implemented with the goal of averting the next mortgage crisis cannot be validated via experiments. Thus, an inherent and stark tradeoff often exists between the ease with which an intervention can be experimentally tested and its potential impact.

¹Department of Marketing, University of Illinois at Chicago, Chicago, IL, USA.

²Department of Marketing, Northwestern Kellogg School of Management, Evanston, IL, USA.

✉e-mail: davidgal@uic.edu

<https://doi.org/10.1038/s44159-021-00002-2>

Causes of experimental validation bias

Behavioural scientists might be susceptible to experimental validation bias because evidence from experiments is relatively precise and easy to evaluate, and individuals tend to overweight such evidence⁶. Moreover, experimental validation of real-world interventions appears ‘scientific’ because it means using the same method — experimentation — that behavioural scientists typically rely on in more basic research.

However, the notion that validating interventions through experiments is or should be scientific is largely illusory. It conflates the goals and methods of science with those of application. Whereas in science the goal is to generate general understanding, in application the goal is to solve a specific problem. In basic behavioural science, experiments are used to establish the mechanisms that drive behaviour. Conversely, application to solving a problem often demands multi-faceted interventions or structural changes that affect multiple disparate psychological (as well as non-psychological) processes.

Behavioural scientists might also be susceptible to experimental validation bias because real-world field experiments correspond to the controlled clinical trials and standards of evidence used to evaluate medical interventions⁷. However, the analogy between behavioural science field experiments and controlled clinical trials in medicine is generally inapt. Although it is undeniable that the effect of medical interventions depends on interactions with individual-level characteristics, biology can be considered a more exact science than psychology⁸, and therefore the effects of medical interventions may be less variable than behavioural interventions. For example, remedying a deficiency in vitamin C will alleviate scurvy now just as it would have 500 years ago. Conversely, whether a deficit in savings is aided by a nudge will depend heavily on how the nudge is construed by the individual (for example, if it is seen as manipulative) at a particular time and in a particular situation. Furthermore, medical interventions are at the level of the individual, whereas impactful behavioural interventions are often at the level of the social environment, the institution, or society at large. We can therefore expect medical interventions to be more predictable in their effects than behavioural interventions.

Extrapolate insights, not effects

To truly fulfill the promise of ameliorating real-world problems with behavioural science, behavioural scientists must recognize two important realities.

First, the primary real-world relevance of our work as behavioural scientists lies not in identifying effects of interventions, but in offering insights that can be used by decision-makers — along with insights from other sources — to inform interventions or other courses of action.

For example, consider the classic experiment demonstrating the endowment effect, in which students given a mug at random demanded more to part with it

(about US\$7) than other students were willing to pay to acquire a mug (about \$3)⁹. To apply this effect, we might seek to influence behaviour by designing an incentive program where workers are endowed with money that is taken away if they fail to perform as opposed to a more traditional program where workers are paid to reward performance. However, even if effective, such an intervention would be impractical to apply in most real-world contexts.

By contrast, the insight from the endowment effect that peoples’ preferences are often ‘constructed’ at the time they are making a decision, on the basis of many decision-specific factors¹⁰, has broad practical relevance. For example, the insight of constructed preferences has been used to argue that marketing recommendation systems will usually offer only crude matches to peoples’ preferences. Consequently, consumers and policymakers should be less concerned that marketers will know exactly what consumers want or what buttons to push to manipulate them, and more concerned with the integrity of information consumers rely on to construct their preferences and make their choices³.

Second, we must accept that most potentially impactful interventions influenced by behavioural insights cannot be tested experimentally. This should not stop the adoption of policy informed by these insights. In fact, even before the use of experiments psychological insights informed policy, such as the separation of governmental powers promulgated by the US Constitution on the basis of the insight that even good people tend to be corrupted by power. However, this approach to applying psychological insights has lost significant ground to the predominant nudge approach in recent years. By recognizing and adopting (and in some sense returning to) these principles, applied behavioural science can meet its full potential in both scope and ambition.

1. Thaler, R. H. & Sunstein, C. R. *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Penguin, 2008).
2. DellaVigna, S. & Linos, E. RCTs to scale: Comprehensive evidence from two nudge units. *National Bureau of Economic Research Working Paper* <https://doi.org/10.3386/w27594> (2020).
3. Gal, D. & Simonson, I. Predicting consumers’ choices in the age of the internet, AI, and almost perfect tracking: Some things change, the key challenges do not. *Consum. Psychol. Rev.* **4**, 135–152 (2021).
4. Cronbach, L. J. Beyond the two disciplines of scientific psychology. *Am. Psychol.* **30**, 116–127 (1975).
5. Goswami, I. & Urminsky, O. When should the ask be a nudge? The effect of default amounts on charitable donations. *J. Mark. Res.* **53**, 829–846 (2016).
6. Hsee, C. K. The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organ. Behav. Hum. Decis. Process.* **67**, 247–257 (1996).
7. Reiley, D. H. & List, J. A. in *The New Palgrave Dictionary of Economics* 2nd edn (eds Durlauf, S. N. & Blume, L. E.) pp. 1–5 (Palgrave Macmillan 2008).
8. Fanelli, D. & Glänzel, W. Bibliometric evidence for a hierarchy of the sciences. *PLoS ONE* **8**, e66938 (2013).
9. Kahneman, D., Knetsch, J. L. & Thaler, R. H. Experimental tests of the endowment effect and the Coase theorem. *J. Polit. Econ.* **98**, 1325–1348 (1990).
10. Payne, J. W., Bettman, J. R. & Johnson, E. J. Behavioral decision research: A constructive processing perspective. *Ann. Rev. Psychol.* **43**, 87–131 (1992).

Competing interests

The authors declare no competing interests.