Journal of Behavioral Economics for Policy, Vol. 1, Special Issue, 9-14, 2017
Behavioral Policy and its Stakeholders

Applying behavioural science to government policy: Finding the 'Goldilocks Zone'

Annalese Bolton[1,2], Ben R. Newell[1]*

Abstract
Using behavioural and social science to inform government policy has the potential to benefit society – provided good scientific practice is adopted. We argue that there are threats to scientific practice in the current Behavioural Insights environment that potentially undermine the validity and usefulness of such work. We discuss what constitutes good science and why this is important, and examine threats to scientific practice from the perspective of scientists conducting experiments on policy initiatives. We aim to assist researchers, governments and policy makers identify conditions where the fit between science and government policy is 'just right' (i.e., the 'Goldilocks Zone'). We discuss potential pathways for developing appropriate infrastructure and procedures to achieve this goal. In particular, we suggest that the early engagement of all parties is necessary to ensure projects incorporate sound science and deliver societal benefit.

Keywords
behavioural insights — government policy — good science

[1] *School of Psychology, UNSW, Sydney, Australia*
[2] *NeuRA, Sydney, Australia*
***Corresponding author**: ben.newell@unsw.edu.au

Introduction

Governments across the world and have been inspired by the potential benefits of using insights from behavioural and social science to inform policy (hereafter referred to as Behavioural Insights [BI]), and an increasing number have established Behavioural Insights or Nudge units to accomplish such goals (Organisation for Economic Co-operation and Development [OECD], (2017a, 2017b)). Applying science to policy is enticing for scientists. It creates potential pathways to apply and test the generalizability of ideas, access large data sets, and contribute to major societal issues. Governments and the public are the intended ultimate beneficiaries of these collaborations. Despite the mutual benefits, many scientists remain cautious about conducting experiments on policy initiatives, and this work remains rare (OECD 2017b). We argue that this caution is for good reasons. While there has been substantial public debate about the benefits (e.g. Thaler & Sunstein, (2008)) and ethics (e.g. Sunstein, (2016a)) of BI, there has been little discussion about *how best* to bring the two different worlds of government and science together, or what infrastructure, procedures and conditions best facilitate such collaboration. Currently, scientists have to creatively navigate their own way through systems that were not designed for conducting experiments. Moreover, scientists must rely on government personnel who may be unfamiliar with scientific processes and principles or what constitutes 'good science' and why this is important. This environment presents many practical challenges and threats to scientific practice, which decreases the ability to produce valid and useful outcomes.

This paper aims to create deeper appreciation of the need to think critically, and address infrastructure and procedural issues. We do not attempt to review the various ways behavioural science insights have been applied to policy, or the different infrastructure set up to facilitate this (although this would be worthwhile). Instead, we elucidate issues arising from this merger by raising challenges from the perspective of behavioural scientists conducting scientific experiments to test policy initiatives.

Our goal is to incite understanding as to why it is important to critically examine how best to apply behavioural science to government policy, and generate motivation for others to engage in the same critical thinking. Our intention is to assist researchers, governments and policy makers to identify conditions for the 'Goldilocks zone'[1] – that is where the fit between behavioural science and government policy is 'just right' for the production of quality scientific work with valid and useful conclusions.

What is good science and why is it important?

Governing bodies across the world have identified similar fundamental principles that researchers should adhere to without exception (e.g. see European Science Foundation [ESF],

[1] The reference to Goldilocks comes from the children's fairy story (Goldilocks and the Three Bears) in which a little girl (Goldilocks) chooses the items (chair, bowl and bed) which are 'just right' for her.

(2000); National Health and Medical Research Council & Australian Research Council, (2007); National Research Council Canada, (2013)). For example, researchers are to design and conduct investigations with the highest professional standards, carry out research and analyze data with a critical and open-minded approach, be frank and fair regarding others contributions, and demonstrate honesty in all stages of scientific enquiry (European Science Foundation 2000). The core of good practice requires facilitating "the vital, external processes of peer review, verification and repeatability" by the scientific community (European Science Foundation 2000, p.5). ESF and All European Academies (ALLEA) prescribe international Good Practice Rules addressing issues such as the availability and access of data, and proper and responsible research procedures (European Science Foundation and All European Academies 2011). ESF and ALLEA recognize that some rules may be subject to different cultural traditions, legislative regulations or institutional provisions. As such, nations, disciplines and institutions are encouraged to identify and incorporate legitimate differences. Instilling such values into individuals alone is not adequate to ensure good practice. Rather institutions employing scientific researchers are required to establish appropriate systems and procedures to facilitate scientific integrity.

While research organizations are well versed with such systems and procedures, not all government departments are. Furthermore, governments do not necessarily have the infrastructure to facilitate good science, particularly regarding experiments on policy initiatives. Discussions about evidence-based policy-making have unfortunately not come with exhortations about procedures required to establish a reliable evidence-base. Without appropriate guidance, BI is vulnerable to good-practice violations.

Science and government policy have intersected previously, and such circumstances highlight the importance of guarding scientific practice quality. For example, the 'sugar conspiracy' controversy offers many sobering lessons to both scientists and governments of a hasty, haphazard, and unscientific intersection of these two worlds (see Teicholz, (2014); Leslie, (il 7)). This example reveals how political connections and power can unduly influence which ideas are funded, quashed or enacted on in policy. It also demonstrates the difficulty shifting a paradigm once it is enshrined as 'scientifically proven', how academic careers can be unjustly ruined, and how devastating public consequences can occur.

Threats to scientific knowledge in BI

In the following section we point out how scientific experiments conducted within government can be time-consuming, subject to practical difficulties that can compromise methodological designs, are bound by delicate data release issues, and come with a risk of being derailed at any stage due to staff turnover or political shifts. These issues result in situations that are personally risky for scientists. Peer-reviewed publications (i.e. the production of scientific knowledge) represent the core key performance indicator for scientists, and therefore investing significant time and energy in projects with on-going instability and no guarantee of publication is risky. Consistent with this, the OECD's survey (OECD 2017a) suggests that often the results of BI studies are not published in any form (i.e. half the studies reported in their survey were not published) and it is rare that studies are published in peer-reviewed academic journals (i.e. less than 6% of surveyed studies). A number of BI journals have recently been established creating a greater platform for BI studies to be published (e.g. Journal of Behavioural Economics for Policy, Behavioural Science & Policy, Behavioural Public Policy). However, as the causes for this failure to publish are unidentified, it remains unknown whether discipline specific journals alone can sufficiently address this situation.

Threats from the cultural environment

Science's cultural environment influences practice quality – that is the extent to which scientist engage in falsification, fabrication, plagiarism, and other questionable research practices that distort the validity of scientific knowledge (e.g. see Munafo et al., (2017)). Issues such as pressure to publish, commercialization, and headline chasing due to job and funding competition (among others) within academic environments are likely contributing to increases in research misconduct (e.g. Nuffield Council on Bioethics, (2014); ESF & ALLEA, (2011)). The Nuffield Council on Bioethics (2014) raised concern regarding the loss of innovative ideas in science due to strategically directed funding. This includes funding particular 'solutions', rather than allowing science to progress through granting scientists freedom to investigate various approaches.

There are likely additional environment factors influencing scientific practice in the BI environment. This is because the drivers, pressures, key performance indicators, and decision structures (i.e. who has authority to make what decisions) within governments are different from academia (EMCR forum 2016).

The BI field comes with its own cultural microcosm, the origin of which can be traced to the original ideas that sparked governments' enthusiasm. Behavioural scientists put considerable effort into communicating the benefits to the public and politicians by claiming that governments could increase their effectiveness while saving millions of dollars using interventions – consisting primarily of minor wording tweaks to existing policies, such as changing default options – that would cost governments next to nothing to implement (e.g. see Sunstein, (2011); Thaler & Sunstein, (2008)). While this approach creates enthusiasm, opportunities and support, it can also produce expectations for behavioural scientists to produce large changes in public behaviour with small manipulations, few resources and little time (OECD 2017b). Over the last 10 years, behavioural scientists have acknowledged various limits to the original approach (e.g. Bhargava & Loewenstein (2015), Furman (2017)) and the scope of BI has broadened to

include evidence-based policy-making, including experiments on policy initiatives (e.g. see OECD, (2017a, 2017b); Shankar & Foster, (2016)). It remains unknown whether governments understand what this broader scope implies for 'quick fix solutions', or have adapted their expectations of what scientists can achieve.

Threats from assertions of haste

> "All research should be designed and carried out in a careful and well considered manner; negligence, haste, carelessness, and inattentions should be avoided, so as to prevent human errors". (European Science Foundation and All European Academies 2011, p.13)

In academic settings, scientists have few external deadlines. The pace of knowledge attainment is set by the gap between what is known and required, the availability of appropriate research techniques, and the level of interest (and funding) in particular topics. Scientists can conduct several studies to ensure sound methodology, replication, and generalizability, as deadlines rarely exist for producing definitive answers. Even if researchers move between research organizations, they can continue working on projects, as projects are tied to individuals.

However, in BI settings, investigations can be subject to deadlines extraneous to knowledge attainment needs. For example, in government, projects are tied to roles not individuals, but the individuals performing those roles often have discretion to choose which projects to support. Staff movement is commonplace and politicians come and go every few years with different ideas of what projects to support. Therefore, research projects within government have impending deadlines with every relevant staff or political changeover. Furthermore, government staff are under constant public scrutiny with high (realistic or perceived) pressure to produce positive results within their service term. In such environments, there is high pressure for scientists to quickly produce positive results and a risk for 'solutions' to be prematurely instantiated into policy.

Ironically, despite the increased pressure for haste, there are significant practical issues that slow progress. Conducting trials within government policy may require various government sub-departments' support and physical involvement due to compartmentalized roles. For example, in some jurisdictions different sub-departments are responsible for IT systems, data management, data extraction, data analysis, policy, service delivery, and so on. Navigating through this system can be time-consuming and there is a constant risk of miscommunication. However, the government systems' strength is that there are a number of checks and balances by different departments along the course of the project (this does not necessarily occur in academia) which can help to maintain the project's integrity.

Threats from inappropriate power assertions

> "Coercion of powerful persons or institutions, religious or political pressure, economic or financial interests can corrupt science". (European Science Foundation and All European Academies 2011, p.10)

ESF views this issue so seriously that their solution is for science to remain independent, and maintain its freedom to adhere to its own laws and criteria within the ethical and social context in which the science proceeds (European Science Foundation and All European Academies 2011).

However, with BI, science simply cannot maintain independence. Policy makers must consider science alongside other factors such as social, economic, political and legal issues. Government officials, rather than scientists, are responsible for ensuring the application of science to policy is optimal, relevant, complete, accurate, and ethical. Government staff therefore decide which projects and solutions to support.

Nevertheless, most government staff do not have science training. This places them at risk of being unable to make judgements about appropriate scientific evidence, evidence limits, or the adequacy of a scientist's knowledge or experience. Similar concerns have been raised in related fields. For example, the Society of Environmental Toxicology and Chemistry (SETAC) provide policy-makers with practical tips for understanding what constitutes 'sound science' and judging studies' and scientists' adequacies (Society of Environmental Toxicology and Chemistry (SETAC) 1999). It remains unknown how well government staff (or even scientists) can make such judgements. Unfortunately, such issues have not been publicly discussed within the BI field. The OECD (2017a) conducted an international survey of institutions applying BI. Of the institutions that directly employed BI staff, 45% employed 'expert' staff, defined as having degrees in 'relevant' disciplines (e.g. Psychology, Economics). Whether this covers undergraduate or postgraduate degrees was not reported and nor was the level of experience conducting experiments or trials. Other institutions involved in the survey reportedly engaged experts on a short-term or consultancy basis, but the experts' qualifications, experience, or level of involvement was not reported. This is worrying as experts early, in-depth and on-going involvement is required to conduct sound scientific investigations, particularly in such a complex arena as public policy.

The BI field has provided little public guidance as to the limits of appropriate assertions by non-scientific decision makers over scientific applications. We are unaware of any formal advice within the field about exactly who should be accountable for what decisions, what are the appropriate processes for decision-making, and what factors should and should not be allowed to influence decisions. Work by organizations such as the Union of Concerned Scientists (e.g. see Union of Concerned Scientists (2012)) may be an appropriate starting place to identify and address potential threats from inappropriate

power assertions along with thorough analysis and discussion of BI scenarios. Public discussions (e.g. in fora, media, literature) are crucial to raise the awareness, and to ensure that quality scientific practice is maintained in environments where science is not independent. As a discipline, we need to think carefully about how to identify and build safeguards into our systems and procedures to prevent and protect against inappropriate or misguided power assertions.

Threats to the scientific community's access to data and replication

Advances in science usually come from engaging the scientific community – many minds addressing the same issue from different perspectives, with each study refining and building our knowledge. This collaborative and incremental approach acts to guard against fabrication and falsification, increases the likelihood that methodological flaws (which render studies invalid and unreliable) will be detected and ensures that studies can be appropriately replicated (European Science Foundation and All European Academies 2011). Replication is vital because it assists science to be self-correcting – it allows for determining whether an observed phenomenon is reliable or was a chance one-off (i.e. false positive). Failure to adhere to this model has, some argue, led to the current 'replication crisis' faced by disciplines underpinning the BI knowledge base (e.g. Psychology). The debate that has ensued highlights the need for multiple replication studies, large sample sizes, pre-registration of studies, and sharing of data to detect and rule out unsubstantiated claims and/or false positive results (e.g. Munafo et al., (2017)). However, adopting such practices in BI presents several challenges. Replication of large scale trials may simply not be feasible given the time, resources and populations involved.

Even if replication of interventions is possible, the release of government or politically sensitive data can be complicated. Governments and the public alike are not necessarily familiar with data release for research purposes, resulting in a risk of misuse. Furthermore, governments do not necessarily have appropriate procedures, systems or trained staff to consider such issues. Good practice involves being mindful of confidentiality when 'legitimately required' (European Science Foundation and All European Academies 2011). The balance between protecting such sensitivities and allowing the scientific community access to data is understandably delicate, and requires thoughtful decisions. Without appropriate processes or guidelines, what is considered 'legitimate' is open to subjective interpretation and is at risk of encapsulating unfounded fears. Concerns need to be identified and addressed. For example, governments may be unwilling to allow studies producing null or negative effects to be published for fear of public perception or how such 'failures' may be used against them politically. Scientists and policy-makers have a role in educating the public that finding out what does not work can be as valuable as finding out what does (e.g. see Sunstein, (2016b)).

Threats to designing and conducting investigations with the highest professional standards

Experimental designs within policy are constrained by what data is recorded or practically obtainable, and by data reliability. Government systems differ in their capacity to identify, randomize, and track participants' behaviour. However, tracking participants is crucial for data required to examine cause and effect. In the academic world, scientists first decide what data they require and then create systems to do and record exactly what they want. However, in the policy world, scientists have to work creatively within the confines of existing systems often designed for non-research purposes. In addition, the intersection of science and government policy is ripe for commercialization. In some cases, governments employ third party commercialized products in their service delivery, which can present significant barriers when designing research methodologies.

Creative 'workarounds' to overcome these practical issues can reduce an experiment's ability to rule out alternative hypotheses, establish cause and effect or measure outcomes. This impedes what can be concluded from a study. Scientists need to maintain understanding of which methodological principles are non-negotiable and which are more flexible in their study designs, and importantly they need to communicate these non-negotiables to their government partners.

How do we get it 'just right'?

As a field, we need to think creatively about solving these issues. A starting point may be to bring together scientists and policymakers who have conducted such experiments to develop a practice guide for the field by identifying issues and solutions. In addition, identifying and comparing the strengths and drawbacks of different partnership infrastructures (i.e. BI units within government, partnerships between universities and government, use of academic advisors, diffuse models, etc.; OECD, (2017a)) may be helpful to clarify appropriate ways forward.

It would be beneficial to think more strategically about what partnerships between government agencies and research institutions look like. One potential solution is to divorce policy issues from specific government departments as major policy issues likely share commonalities across jurisdictions and countries. For example, public misuse of ambulances represents a policy problem identified across multiple countries (e.g. see Chen, Bullard, & Liaw (1996), Snooks, et al.(1998)). We suggest that governments identify issues and desired outcomes, while researchers within institutions identify and address knowledge gaps and investigate a range of creative solutions prior to testing within the policy field. Platforms facilitating appropriate connections between scientists and policy makers (e.g. the BSPA's matchmaker portal) may be useful. Close relationships are required between policy makers and scientists to ensure that scientists appropriately consider relevant government context and systems factors. Such relationships would also assist all parties to be involved

at the beginning of a project rather than having academics brought in simply to lend a veneer of scientific credibility to a project after key design decisions have been made. On-going collaborations may also allow for greater buy-in from other funding sources (e.g. the Australian Research Council's Linkage grants require academic institutions to collaborate with other sectors such as government agencies), and create greater stability for completing projects. The OECD (2017a) similarly recommend that governments maintain specialization internally but develop partnerships with external bodies with relevant expertise.

Conclusion

BI is only in its infancy and still largely finding its place. Scientists are obliged to encourage our work to be used in a constructive manner. There is little doubt that the use of behavioural and social science to inform government policy can benefit society - as long as it is based solidly on good science. As such, maintaining vigilance over scientific practice is of utmost importance. Our brief review identifies four key recommendations to improve the nexus between science and government institutions who want to employ BI. Governments and scientists need to work together to 1) develop appropriate infrastructure and procedures to facilitate good scientific practice, 2) identify potential threats to research integrity and develop appropriate safeguards, 3) identify and acknowledge the practical difficulties that arise within the intersection of government and science, and 4) provide strategic guidance to overcome such issues. Perhaps then, we may be closer to finding our 'Goldilocks zone'.

References

Bhargava, S. and G. Loewenstein (2015). Behavioural economics and public policy 102: Beyond nudging. *American Economic Review: Papers & Proceedings 105*(5), 396–401.

Chen, J.-C., M. J. Bullard, and S.-J. Liaw (1996). Ambulance use, misuse, and unmet needs in a developing emergency medical service system. *European Journal of Emergency Medicine 3*, 73–78.

EMCR forum (2016). *Discussion paper: Starting the conversation between academia and industry (Consultation draft)*. Australian Academy of Science. Retrieved from https://www.science.org.au/files/userfiles/support/emcr-activities/academia-industry-discussion-paper-consulation-draft.pdf.

European Science Foundation (2000). *Good scientific practice in research and scholarship*. Retrieved from http://archives.esf.org/fileadmin/Public_documents/Publications/ESPB10.pdf.

European Science Foundation and All European Academies (2011). *The European code of conduct for research integrity*. Revised Edition. Berlin: ALLEA. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf.

Furman, J. (2017). Applying behavioural sciences in the service of four major economic problems. *Behavioral Science & Policy 2*(2), 1–7.

Leslie, I. (2016, April 7). The sugar conspiracy. *The Guardian*. Retrieved from https://www.theguardian.com/society/2016/apr/07/the-sugar-conspiracy-robert-lustig-john-yudkin?CMP=fb_gu.

Munafo, M. R. et al. (2017). A manifesto for reproducible science. *Policy Briefing 1*, 1–9.

National Health and Medical Research Council & Australian Research Council (2007). *Australian Code for the Responsible Conduct of Research*. Retrieved from https://www.nhmrc.gov.au/_files_nhmrc/file/research/research-integrity/r39_australian_code_responsible_conduct_research_150811.pdf.

National Research Council Canada (2013). *Code of Conduct*. Retrieved from http://www.nrc-cnrc.gc.ca/obj/doc/about-apropos/ethics_integrity-ethiques_integrite/code_of_conduct_may_2013.pdf.

Nuffield Council on Bioethics (2014). *Culture of scientific research in the UK*. Retrieved from https://nuffieldbioethics.org/wp-content/uploads/Nuffield_research_culture_full_report_web.pdf.

OECD (2017a). *Behavioural insights and public policy: Lessons from around the world*. Paris: OECD Publishing. http://dx.doi.org /10.1787/978926427080-en.

OECD (2017b). Use of behavioural insights in consumer policy. *OECD Science, Technology and Industry Policy Papers* (36). Paris: OECD Publishing. http://dx.doi.org/10.1787/c2203c35-en.

Shankar, M. and L. Foster (2016). *Behavioural insights at the United Nations - Achieving agenda 2030*. United Nations Development Programme. Retrieved from http://www.undp.org/content/undp/en/home/librarypage/development-impact/behavioural-insights-at-the-united-nations–achieving-agenda-203.html.

Snooks, H. et al. (1998). Appropriateness of use of emergency ambulances. *Journal of Accident & Emergency Medicine 15*, 212–215.

Society of Environmental Toxicology and Chemistry (SETAC) (1999). Sound Science Technical Issues Paper. Pensacola, FL, USA.

Sunstein, C. R. (2011). Empirically informed regulation. *University of Chicago Law Review 78*, 1349–1430.

Sunstein, C. R. (2016a). *The ethics of influence: Government in the age of behavioural science*. New York, NY: Cambridge University Press.

Sunstein, C. R. (2016b). *Nudges that fail. July 18, 2016*. Retrieved from https://papers.ssrn.com/sol3/ papers.cfm?abstract_id=2809658.

Teicholz, N. (2014). *The big fat surprise: Why butter, meat and cheese belong in a healthy diet*. New York, NY: Simon & Schuster.

Thaler, R. and C. Sunstein (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Penguin Books.

Union of Concerned Scientists (2012). Heads they wins, tails we lose: How corporations corrupt science at the public's expense. Retrieved from http://www.ucsusa.org/sites/default/files/legacy/assets/documents/scientific_integrity/how-corporations-corrupt-science.pdf.