Review of *The Origins of Behavioural Public Policy*, by Adam Oliver.

Adam Oliver has written an introduction to behavioural public policy, targeted at both students and laypersons. Oliver is a behavioural economist—much of his research has involved applying behavioural economics to the domain of health and healthcare—and he sees behavioural public policy through this lens. Economics starts with a model of a rational agent. However, there is plenty of evidence that people violate the economic assumptions of rationality in systematic and predictable ways. People use heuristics, or mental shortcuts, which evolved because they are fast and efficient ways of making decisions, giving our ancestors an advantage in the Pleistocene. However, when applied to the decisions we make in modern life, they sometimes lead to 'biases', or systematic deviations from the rational choice model. For a behavioural economist, behavioural public policy is about acknowledging these biases and therefore going beyond the traditional economic tools of regulation and pricing (i.e. taxes and subsidies).

Accordingly, Oliver starts by setting out the traditional rationality assumptions that are made by economists and the evidence that humans violate them. The basic set-up has an individual agent who is choosing between risky 'lotteries', i.e. it is not necessarily certain what outcome will result from her choice. Expected utility theory specifies a small set of intuitively plausible axioms that constrain the form of the agent's preferences over these lotteries. The seminal contribution made by von Neumann and Morgenstern (1947) was to show how an agent who satisfies these axioms can be mathematically modelled as maximizing a cardinal utility function: roughly speaking, rational agents choose the action that will give them the greatest possible subjective value from the outcomes, taking into account the probability that each outcome occurs. The expected utility theory framework allows economists to build mathematically tractable models from fairly parsimonious assumptions. From the beginning, it was ambiguous whether expected utility theory was supposed to be a descriptive model of behaviour or a normative model of the choices that a rational agent would take, and it is sometimes also called rational choice theory.

Oliver presents a number of 'paradoxes' of rationality, which are really patterns of choices or behaviours that are puzzling for expected utility theory because they imply that

people do not obey the axioms. He covers the classic examples, such as the 'Allais Paradox', where people seem to place a special value on certainty, disliking risk when certainty is an option, and 'preference reversals', where, when asked to choose between two lotteries x and y, people say that they would prefer to have x but, when asked to place monetary values on the lotteries, they place a higher value on y than x. In the following chapter, Oliver presents the main descriptive theories that aim to accommodate these findings. Then there three chapters on specific topics that challenge economic ideas about rationality. The first is on intertemporal decision-making. When making choices with outcomes that may be realised in the future, it may be rational to weight sooner outcomes more than later ones because, the further in the future the outcomes are realised, the less certain it is that one will actually get them. However, people display a 'present-bias'. They attach a particularly heavy weight to the present, which implies that if they have to make a choice now with an outcome that will be realised in the future, then they might reverse that choice if allowed to reconsider when the time comes. Second, Oliver covers the gestalt effects of experiences, where we value wholes differently from the sum of their parts. When evaluating an experience, we place an emphasis on the peak of the experience (high or low point) and the end moment. This implies that, if we take a painful experience, we can improve our evaluation of the experience if we add on an extra period of pain, providing that the extra pain is slightly less painful than the previous end moment. Paradoxically, more pain (as measured by summing across time periods) may be preferred to less. Finally, Oliver covers evidence that people's attitudes to incentives are different from those assumed by economics. Prices and fines can have perverse effects, for instance payment for blood leading to less being collected, or a fine for late pick-ups at nursery leading to parents arriving even later.

This part of the book is basically an exposition of relevant concepts from behavioural economics, whose relevance is demonstrated through policy examples, predominantly from healthcare. I found the material really clear—I especially liked that Oliver managed to explain and use the Marshak-Machina triangle—although obviously I'm not best placed to judge comprehensibility for the target audience. This material leads into discussion of non-standard policy interventions, or nudges, where Oliver draws on

his own research, applying a framework that he developed to classify various policy examples and using it to critique the theoretical foundations of nudge.

Oliver presents the following five requirements for an intervention to be a 'nudge', which he says are derived from Thaler and Sunstein (2003, 2008): the intervention must rely on automatic decision processes, be liberty preserving, not use large financial incentives, be informed by behavioural economics; and target internalities (in other words they are supposed to have beneficial effects for the decision-maker herself, rather than beneficial effects on others, as in the classic 'externalities' of economics). However, Thaler and Sunstein's definition of a nudge is much more expansive than this: "A nudge, as we will use the term, is any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler and Sunstein, 2008, p.6). I want to concentrate on two ways in which this definition differs from Oliver's criteria.

I have a lot of sympathy for Oliver's idea that nudges involve automatic processes, but that is not the position that Thaler and Sunstein (2008) take. Not only is his criteria more restrictive than their definition, but Sunstein (2016) has recently been investigating the public acceptability of automatic ('System 1') nudges, such as default rules, vs deliberative ('System 2'), nudges, such as factual disclosures. One point in favour of Oliver's requirement of automaticity is that it fixes the boundaries of nudge in the right place. Consider information campaigns. As Oliver points out, they are not generally considered to be nudges. However, some nudges, for instance traffic light labelling systems, clearly do impart information. What makes these interventions nudges when information campaigns are not? One possible answer is that information campaigns are designed to change minds via deliberation, whereas a traffic light system is designed to influence the automatic system. Where Sunstein (2016) talks of 'System 1' vs 'System 2' nudges, we might wonder if it makes more sense to talk of non-conscious vs conscious apprehension of how the nudge works, or even just 'opaque' vs 'transparent' nudges (Osman, Lin & Gold, in prep).

Oliver's internality criterion is also more restrictive than Thaler and Sunstein's (2008) definition because they do not rule out the possibility that nudges target

externalities. For Oliver, targeting internalities rather than externalities is what makes an intervention a nudge rather than a 'budge'. So, for Oliver, changing the default so that people have to opt out of pension contributions is a nudge, while changing the default so that people have to opt out of organ donation is a budge. For Thaler and Sunstein, both would be nudges. (The third item in Oliver's helpful framework is a 'shove', a regulatory intervention that uses behavioural economics in order to resolve internalities. His example is a smoking ban in public places, though I have to say that I don't understand why this is a behavioural rather than a standard economic intervention.) Again, there is a rationale for the restriction. A lot of the debate around nudge focusses on 'Libertarian paternalism', or whether it is acceptable to nudge people for their own good. However, this discussion is not relevant to policies that are justified in terms of their externalities, i.e. Oliver's budges.

There is a broader point that can be made here. A lot of intellectual ink has been spilled on the ethics of nudge, indeed I am guilty of spilling some of it myself (Osman, Lin & Gold, in prep). However, in policy applications, Libertarian paternalism may not be so relevant after all. This is not to say that policy makers do not worry about the ethics of nudge—they do—but that, in my opinion, most behavioural public policy interventions have externality elements to them. Many behavioural interventions are simply straight budges. Oliver gives the example of defaults for organ donation, but there are many others. For instance, the Behavioural Insights Team at Public Health England (no relation to the 'nudge unit' discussed in the book) is using behavioural insights to prevent anti-microbial resistance, whereby over-using antibiotics now increases the speed at which they will stop working for other people in the future (Hallsworth et al., 2016). In the Treasury, there is a Behavioural Insights Unit, which uses behavioural interventions to get people to pay their taxes on time. In these cases, Libertarian paternalism does not apply. But even policy interventions that are not wholly budges often have an externality element to them. Of course, if we use behavioural insights to encourage people to attend their NHS health checks (Sallis et al., 2016; Sallis et al. in prep a; Sallis et al. in prep b), we want them to be healthier for its own sake. But when someone takes action in order to stay fit and well, there is also a positive externality (a benefit that does not accrue to the person taking the action) because they will need less healthcare in the long run therefore saving the NHS

money (Department of Health, 2008). It may be easiest to see the externality by considering a health behaviour with a negative externality. One justification for the taxation of cigarettes is that they lead to illness, which places a burden on healthcare services. In countries with socialized medicine, such as the UK, this is funded from general taxation, so it is a cost that does not accrue to the person smoking the cigarette. Justifications deriving from cost effectiveness and externalities do not need to invoke Libertarian paternalism.

That strikes me as a good thing. There is a long tradition of justifying the state and its authority by appealing to the reciprocal benefits of providing public goods or social insurance. Behavioural public policies that have beneficial externalities can be justified within that framework. However, there is no concensus on whether the apparatus of the state may be used paternalistically on citizens.

This book could be called 'Behavioural economics and public policy'. Oliver is open that his focus is on behavioural economics and the approach he takes to public policy is pretty standard for a behavioural economist. However, many practitioners of behavioural public policy think of it more broadly as 'behavioural science and public policy'. This has several implications.

First, although behavioural economics has the limelight, other disciplines have been influential. As one health psychologist that I work with put it, 'Kahneman won the Nobel prize for economics and suddenly we had all these people coming on our turf and telling us it was behavioural economics.' Oliver gives a friendly critique of the Mindspace report, a tool which draws heavily on behavioural economics (Cabinet Office and Institute for Government, 2010). However, psychology had also recognised that there is an 'intention-behaviour gap' (Sheeran, 2002), so that traditional policy interventions that are aimed at changing minds have at best moderate effects (Webb & Sheeran, 2006) and the discipline has therefore developed its own tools. To take Oliver's preferred domain of health, there are tools such as the Behaviour Change Wheel and the Theoretical Domains Framework (TDF), which are used to diagnose the cognitive, affective, social, and environmental influences on behaviour change (Michie, Atkins, & West, 2014; Michie et al, 2005; Cane et al, 2012). The tools from psychology put more emphasis on analyzing the situation,

diagnosing the barriers to behaviour change, and matching interventions to barriers, compared to the Mindspace checklist of possible behavioural techniques. Sociologists are also in on the act, criticizing both psychology and economics for paying insufficient attention to institutional structures, leading to them to propose the Individual, Social and Material (ISM) tool, which has been used by the Scottish government (Darnton & Horne, 2013). That leads to an issue of disciplinary translation. The TDF synthesises psychological theories of behaviour change that are relevant to implementation of new behaviours, identifying the constructs that are used in these theories and grouping them into domains. Health psychologists sometimes ask of the heuristics and biases approach 'where is the theory?'. But, of course, in economics there is a theory of rational choice and then a catalogue of deviations from that theory, which arguably correspond to a suite of psychological mechanisms. As we go forwards, it will be important to develop ways of translating between the languages of different disciplines.

Second, although people working in behavioural insights predominantly focus on changing choice architecture, they do not exclusively do that. They may also use traditional price instruments, for instance, behavioural teams have worked on alcohol and sugar taxation (Harper, Ravenscroft & Service, 2018) and on incentivising general practitioners to perform NHS health checks (Public Health England, 2018). Policies that use financial incentives to change behaviour are not what gives behavioural public policy its originality. However, the two approaches may not be so easy to disentangle. When an item is on sale, its price is decreased with the aim of increasing purchases. But being on sale may also make the item salient, having a similar effect through a non-rational avenue. We also know that, when a relative price changes, responses to the policy may depend on how the change is presented, for instance a cash discount is considered more acceptable than a credit card surcharge (Kahneman, Knetsch, & Thaler, 1991). (Some of these ideas are covered in Oliver's chapter on motivation crowding.) So, although there is a rationale to Oliver's focus on deviations from the rational actor model, one shouldn't lose sight of the fact that the practice of behavioural public policy is broader than this.

Third, behavioural public policy takes methodology, as well as content, from behavioural science. One of the innovations, which Oliver does not emphasize, has been to

run randomized controlled trials (in the language of economics, 'field experiments') to learn what policy interventions will be most effective (Haynes et al, 2012). The idea of evaluating policies is not new. However, in policy circles, the approach of using a control group in order to establish that an intervention caused a change in behaviour is novel. If, as well as giving an interevntion to a group, there is a closely matched control group, who do not receive the intervention, then we can deduce that any difference in behaviour between the two groups is due to the intervention. Without a control group, we cannot tell whether a change in behaviour is due to the intervention or to other external factors that may also have changed during the course of the evaluation.

Oliver is very circumspect in his presentation of the material. He is judicious in his criticism, not revealing his own views until the final chapter. In both the cases where Oliver deviates from Thaler and Sunstein, I would have preferred it if he had presented his criteria as critique, rather than clarification, of their position. In general, it is not hard to imagine that this book's origins were a seminar series; it would be a very even-handed teaching aid. The flip side is that positions are often taken from the mouths of other authors, so the book can be a little 'he said, she said' at points. This is a shame because, when it comes through, Oliver's narrative voice has a wonderful wry wit.

Oliver ends by speculating whether behavioural public policy will maintain such a high profile in the future; he guesses that it will not, as the publicity around it dies down. On this, I disagree with him. At least in policy circles, the influence of behavioural insights, understood as a broad spectrum of interventions that are based on behavioural science, is only increasing. Oliver reports the worry of some theorists that nudges are less effective at changing behaviour than traditional regulation (Loewenstein et al. 2012; Marteau et al, 2011). While it is true that nudges generally do not bring about such large changes as regulations, they are cost effective and easy to implement. In an organization the size of the NHS, an eminently achievable 5% improvement in performance saves a lot of lives and a lot of money. For this sort of reason, there are many behavioural units across government and the private sector. Oliver focuses on the work of the Behavioural Insights Team or 'nudge unit', which originated in the cabinet office but was spun off as a social purpose company. However, most government departments will have a behavioural unit; as do

other public bodies, such as the Financial Conduct Authority; and private sector firms, such as consultancies and banks. So I don't think behavioural insights is going away soon. Also, Oliver's main focus is on the theoretical application of behavioural science to public policy. But the other innovation of behavioural public policy is a shift to more empirically informed policy-making, scientifically testing policy changes in a small area before rolling them out on a larger scale. Again, I expect that this approach is here to stay. All the more reason to read this book.

In his preface, Oliver says that he has not written a popular science book. Nevertheless, he has written a book that—as well as being informative—is immensely readable, and I highly recommend it.

Natalie Gold

References

Cabinet Office and Institute for Government (2010) *MINDSPACE: Influencing behaviour through public policy.* London: Cabinet Office and Institute for Government.

Cane, J., O'Connor, D., & Michie, S. (2012). Validation of the theoretical domains framework for use in behaviour change and implementation research. *Implementation science*, *7*(1), 37.

Department of Health. (2008a). *Putting prevention first vascular checks: risk assessment and management, impact assessment*. London: Department of Health.

Darnton, A. & Horne., J (2013) *Influencing Behaviours.* Edinburgh: Scottish Government.

Felsen, G., Castelo, N., & Reiner, P. B. (2013). Decisional enhancement and autonomy: public attitudes towards overt and covert nudges. Judgment and Decision Making, 8(3), 202. 10.1177/2332858416674007

Hallsworth, M., Chadborn, T., Sallis, A., Sanders, M., Berry, D., Greaves, F., Clements, L. & Davies, S. C. (2016). Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial. *The Lancet*, *387*(10029), 1743-1752.

Haynes, L., Service, O., Goldacre, B., & Torgerson, D. (2012). Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials. London: Cabinet Office.

Hugo Harper, Luke Ravenscroft and Owain Service (2018) Sugar tax: how will it affect behaviour? Available at http://www.behaviouralinsights.co.uk/health/behaviour-change-and-the-new-sugar-tax/, accessed 12th March 2018.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, *5*(1), 193-206.

Loewenstein, G., Asch, D. A., Friedman, J. Y., Melichar, L. A., & Volpp, K. G. (2012). Can behavioural economics make us healthier?. *BMJ: British Medical Journal (Online)*, *344:e3482*.

Marteau, T. M., Ogilvie, D., Roland, M., Suhrcke, M., & Kelly, M. P. (2011). Judging nudging: can nudging improve population health?. *BMJ: British Medical Journal (Online)*, *342*.

Michie, S., Johnston, M., Abraham, C., Lawton, R., Parker, D., & Walker, A. (2005). Making psychological theory useful for implementing evidence based practice: a consensus approach. *BMJ Quality & Safety*, *14*(1), 26-33.

Michie, S., Atkins, L., & West, R. (2014). The behaviour change wheel: a guide to designing interventions. *Needed: physician leaders*, *26*.

Neumann, J. von, and Morgenstern, O. (1947). *Theory of Games and Economic Behavior*, Second edition. Princeton, NJ, Princeton University Press.

Osman, Lin & Gold (in prep) 'What factors persuade the public on social policy interventions?

Public Health England (2018) NHS Health Check Commissioning: *current and potential use of weighted financial remuneration*. Available from www.healthcheck.nhs.uk

Sallis, A., Bunten, A., Bonus, A., James, A., Chadborn, T., & Berry, D. (2016). The effectiveness of an enhanced invitation letter on uptake of National Health Service Health Checks in primary care: a pragmatic quasi-randomised controlled trial. *BMC family practice*, *17*(1), 35.

Sallis, A., Agbebiyi, A., James, R., Berry, D., Gold, N. & Bonus, A. (in prep a) Increasing uptake of National Health Service Health Checks in Primary Care: A pragmatic randomised controlled trial of enhanced invitation letters in Northamptonshire, England.

Sallis, A., Sherlock, J., Bonus, A., Saei, A., Gold, N. &. Chadborn, T (in prep b) Increasing uptake of National Health Service Health Checks: A factorial randomised controlled trial using theoretically informed letters and SMS pre-notification and reminder text messages.

Sheeran, P. (2002). Intention—behavior relations: a conceptual and empirical review. *European review of social psychology*, *12*(1), 1-36.

Sunstein, C. R. (2016). People prefer system 2 nudges (kind of). *Duke Law Journal*, 66, 121-168.

Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *The American Economic Review*, *93*(2), 175-179.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological bulletin*, *132*(2), 249.

BIOGRAPHICAL INFORMATION

Natalie Gold is a Senior Research Fellow in the Philosophy Department at Oxford University and a Principal Behavioural Insights Advisor at Public Health England. Her research area is behavioural decision making, and she does both theoretical and empirical work, including laboratory experiments and randomized controlled trials. She has published on topics including framing, moral judgements and decision, cooperation and coordination, self-control, and trust in finance.