doi:10.1017/S0266267115000243

*Behavioural Public Policy*, edited by Adam Olivier. Cambridge University Press, xv + 235 pages.

## 1. INTRODUCTION

The rapid expansion of behavioural economics into the field of policy interventions has received widespread attention, including outside of academia. 'Nudge' policy proposals, aiming to change people's behaviour without substantially affecting freedom or material incentives, have garnered news coverage and have been promoted by various newly created public and private policy institutions (e.g. the Behavioural Insights Team in the UK). The edited volume *Behavioural Public Policy* joins this development, by instructing its nine contributing authors 'to relate an aspect of behavioural economics to a policy concern of their choosing' (13) and by soliciting commentaries for each of the contributed papers. The result is an eclectic and interdisciplinary mix of papers, with contributors including economists, psychologists, political scientists and philosophers. The policies discussed include interventions in the health sector, in individuals' financial decisions, concerning the environment, and how to reward people for their work. The discussions take on different formats, ranging from two applications of extant models to particular policy cases, one report of a new empirical study, four reviews of previously performed empirical or theoretical work, and two papers that push the conceptual discussion beyond a mere review of previous work. Not all the work is original: two contributions are explicitly marked as recycles, while some of the other papers seem to provide only material that already has been published. Nevertheless, seeing the different approaches next to each other makes for interesting reading, and many of the commentaries provide valuable insights and correctives to the main articles. Anyone interested in behavioural public policy will find plenty of food for thought here.

For philosophers, a number of problems that still beset the field and that relate to issues of evidence quality, the rationality/irrationality divide, and the question of underlying mechanisms are of particular interest. These problems are sometimes addressed in the main papers and often in the commentaries accompanying them; however, some of them are not addressed at all. In those cases, the book serves rather as an illustration of the problems of behavioural public policy, than as a discussion of them. In order to highlight these problems, I will organize the rest of this review accordingly.

## 2. EVIDENTIAL SUPPORT

The first problem concerns evidential support. For example, Adam Oliver seeks to explain the UK government's response to the 2009 swine flu outbreak as driven by ambiguity aversion. Agents are ambiguity averse if they avoid strategies whose outcomes are uncertain, in such a way that they violate standard decision theory postulates (Ellsberg 1961). Olivier argues that the government's decisions to purchase sufficient vaccine for the worst-case scenario, to not include breaking clauses in these purchasing contracts, and to adopt a 'treat all' approach in the containment phase all point to the government's ambiguity aversion.

There are multiple problems with this argument. First, in contrast to Ellsberg's scenario (which compares *two* decisions under slightly altered conditions), Oliver can cite only *one* observed government choice. Without the second choice as a reference, one in principle cannot distinguish whether the choice was driven by ambiguity aversion or simply by risk-averse preferences.

Second, the evidence that Oliver cites seems to imply that the government didn't even realize the extent of the uncertainty involved. For example, 'the initial information … appeared to suggest that the virus was associated with rapid spread and high fatality' (21). Oliver also criticizes policymakers for 'over-rely[ing] on modellers because they are 'credible' … and give concrete, easily understood, seemingly robust answers' (21). These are instances of misinformation or misjudgement, not uncertainty: policymakers considered certain information more reliable than it was, and might have formed precise, albeit false, probabilistic beliefs. Without an agent realizing the uncertainty involved, however, ambiguity aversion does not apply.

Third, as Oliver acknowledges, there are other explanations for the government's behaviour beyond ambiguity aversion or risk aversion – in particular the desire to maintain public confidence. What remains is that the ambiguity aversion is a *how-possibly explanation* of the swine flu case, without much evidential support. Consequently, the policy recommendations following from this explanation are somewhat muddled. Battling ambiguity aversion presumably would require eliciting one's risk preferences in low-ambiguity contexts, and then deriving decision guidelines from them that could be employed in high-ambiguity contexts. Oliver, in contrast, suggests that 'the government … take a view on the most *likely* outcome of the pandemic' (29, my emphasis). Here the cure clearly denies the diagnosis: discerning more or less likely outcomes falls outside of the realm of uncertainty. What remains is the (possibly correct) criticism that the government has for various reasons disregarded or misjudged important evidence; but that issue has little to do with ambiguity aversion.

https://doi.org/10.1017/S0266267115000243 Published online by Cambridge University Press

A similar problem arises with the paper by Kate Disney, Julian Le Grand and Giles Atkinson (DLA), who investigate the introduction of a 5p fee for plastic bags in a major UK supermarket chain. Based on a survey study, they conclude that the policy 'crowded in' pro-environmental behaviour in the form of increased plastic bag re-use. 'Crowding in' is an interesting new concept: instead of the well-known effect of financial incentives 'crowding out' existing altruistic motivations, DLA suggest that financial disincentives like the plastic bag fee might *increase* altruistic motivations, which in this case supposedly leads to a reduction of plastic bag consumption independently of price considerations. In the study, DLA observed that shoppers increased their re-use of plastic bags after the introduction of the 5p fee. After the fee introduction, a third of these shoppers also claimed that they would re-use plastic bags at other, non-fee-charging stores as well. DLA conclude 'since those who were more likely to reuse at other stores ... were clearly motivated by factors other than cost, this would suggest a crowding in of motivation' (79). Besides the methodological problems of such a conclusion (the question is prone to create dissonance-reducing responses), the purported behaviour change can also be explained without reference to crowding-in. For example, as Richard Cookson suggests in his commentary, it might be explained as reduced costs: once people start reusing bags, it is easy to maintain and expand this habit. So the DLA model again is no more than one possible account of the behaviour. Perhaps these are isolated incidents, but it seems to me that if such cases appear in a noted anthology of behavioural public policy, this is a worrying sign for a movement claiming that 'our quiet revolution is putting evidence at the heart of government' (David Halpern, director of the UK Behavioural Insights Team, in Halpern 2014).

In defence of Oliver and DLA, I should say that they at least face up to the difficult external validity problems that explanations of and policy recommendations for such cases pose. Others in this volume (such as Paul Dolan or Bruno Frey) avoid arguing why a certain intervention is or is not effective in a particular environment altogether. Instead, they review literature that shows that certain interventions are effective in *some* environments. Frey for example describes how monetary rewards can *crowd out* intrinsic motivation if the recipient perceives the reward as 'controlling' (169), while it might *crowd in* intrinsic motivation if the recipient perceives it as 'acknowledging the good work performed' (171). The crucial feature on which these antipodal predictions depend are the employees' interpretations. Unless systematic relationships between manipulable features of the environments and such interpretations can be identified, the abstractly described intervention effects will be of little help for public policy-making.

https://doi.org/10.1017/S0266267115000243 Published online by Cambridge University Press

## 3. MECHANISMS

Beyond the external validity issue, questions arise also about the internal validity of the behavioural models. Often the interpretation of a behavioural phenomenon and its influencing factors are in question. A pertinent example is the exchange between Paul Slovic & Daniel Västfjäll (SV) and Jonathan Wolff. SV diagnose a systematic 'insensitivity to mass tragedy' (94) in people's behaviour: when faced with suffering of large groups of victims, for example from genocide or natural disasters, people feel comparatively less compassion and give less aid than when confronted with individual victims. They propose a psychophysical model of *psychic numbing* that describes an inverse relationship between an affective valuation of saving a life and the number of lives at risk. SV argue that this affective valuation is the basis for most intuitive moral judgements about how much effort or how many resources to devote to saving lives. They also argue that such intuitive judgements are morally unacceptable (because each life should be valued equally, no matter how many other lives are at stake). Consequently, they propose a corrective education of moral intuitions through framing, individualization and harnessing the power of narratives.

By proposing these policy interventions, SV show that they take the psychic numbing model not just as a description of the data, but as a causal model: they seek to individualise victims in order to break the causal force of perceiving victims as groups. Wolff, in his commentary, disagrees with this causal interpretation. He suggests instead that the numbing effect might arise from the fact that some disaster situations apparently lack a clear 'cut-off point', or that the outcome of any intervention in such a situation is highly uncertain. Which side is right will have important effects for the success of the proposed policies. I find it therefore disheartening that so little evidence for the affective or cognitive mechanisms underlying the behaviours and policies in question is provided in this book (and elsewhere in the behavioural public policy literature).

Of course, one reason for this might be that the domain of behavioural public policy is still immature. The oft-repeated 'more needs to be done' mantra seems to speak to this. Perhaps we still are in the phase of modelling merely *possible* mechanisms, and the task of evidencing them comes later. The papers by Matthew Rabin and Drazen Prelec are examples of that view. Rabin describes a habit-formation mechanism, where current consumption of a good increases the marginal utility of future consumption of that good, and argues that people systematically underestimate the impact of current lifestyles on future habit formation. While one can in principle see how unhealthy behaviour now might have future bad consequences through habit formation, Rabin stresses the

https://doi.org/10.1017/S0266267115000243 Published online by Cambridge University Press

'extreme lack of evidence on the degree to which eating and exercise are habit forming, and on the degree to which people may behave irrationally in the face of this habit formation' (116).

Prelec continues this discussion of intertemporal choice by presenting a theoretical model of a self-control strategy. People suffer from self-control problems because they steeply discount future outcomes in comparison to the present: satisfying the craving for just one cigarette *now* is more important than general health considerations, and what harm does a single cigarette do, anyhow? Yet the same person who prefers immediate gratification now would perhaps not have done so yesterday: yesterday, she might have said she would quit today. To explain how people might overcome such self-control problems, Prelec presents a model of self-signalling. According to the model, a potential smoker is able to abstain because she sees lighting up today not as an isolated event but as a signal about her character and about her inability to ever give up: 'what keeps the smoker from smoking is the immediate loss in expectations about long run health that would be triggered by a single cigarette' (219). This model is not new (it draws directly on George Ainslie's 1992 work on recursive self-prediction), but Prelec presents it in a formal way that makes it easily comparable to evidential decision theory and Newcomb's paradox (a fact that Luc Bovens points out in his commentary). This is interesting, but purely theoretical work.

However, there is at least one paper in this collection that seeks to provide evidence for cognitive mechanisms. Sunita Sah, Daylian Cain and George Loewenstein (SCL) argue that mandatory disclosure of conflicts of interest in medicine can have perverse effects both on advisees and advisors. Reviewing studies of their own and others, they show how careful study design allowed them to identify two psychological processes affecting advisees. *Insinuation anxiety* lets advisees fear that rejecting advice may signal to the advisor that they believe the advisor is corrupt; the *panhandler effect* lets advisees feel the pressure to help advisers obtain their personal interests once the adviser discloses this interest. Based on this mechanistic evidence, they propose policy interventions – including making disclosures to advisees at temporal and spatial distance from the advisers and providing disclosures secretly – that break the causal influence of these processes. Work like the papers that SCL review shows that evidential support for mechanism identification is possible, if hard to come by. More of this kind of work hopefully will appear in the future.

## 4. RATIONALITY

So far, I have discussed methodological and philosophy of science issues with behavioural public policy as described in this book. Yet even if interventions were effective in relevant environments, there remains

https://doi.org/10.1017/S0266267115000243 Published online by Cambridge University Press

the question of the justifiability of such intervention. Robert Sugden's commentary on SCL is a good example of this. He does not deny the causal effectiveness of the interventions that SCL propose. Rather, Sugden questions the rationale for disclosing conflict of interest in the financial and medical domains and our preoccupation with rectifying perverse effects of such disclosures. When purchasing produce in the supermarket, we don't expect impartial advice and hence don't need full disclosures. So why in the financial industry?

The dominant answer that behavioural economists give to such challenges is that people are systematically irrational and make predictable mistakes in particular environments, which policy interventions should rectify. Many authors in this anthology consequently seek to establish such irrationality, including Oliver's claim of ambiguity aversion, SV's claim that 'the rationality of these responses [lowering donation rates when victim numbers increase] can be questioned' (100), and Rabin's discussion of 'two basic errors that may lead people to engage in too many bad habits and too few good habits' (116).

A radical reply to this argument is to deny that policymakers are in the business of ensuring preference maximisation. Sugden proposes this view, arguing that the policy maker should provide opportunities instead. Consequently, even if one accepts the evidence that people make predictable mistakes, this might not constitute a reason for intervention.

A less radical response accepts that sufficient evidence for people making predictable mistakes would constitute a reason for intervention, but asks: *what mistakes*? Some of the papers of this anthology might invite such a criticism. SV, for example, argue that psychic numbing is irrational, because 'we should not be deterred from helping one person, or 4,500, just because there are many others we cannot save' (100). I disagree. If, for example, information about the extent of the disaster affects one's belief in the effectiveness of attempting to help an individual, then it might be rational to choose as if one were 'numbed'. Furthermore, SV themselves argue that psychic numbing can occasionally be considered rational (as when it 'enabled rescue workers to function during the horrific aftermath of the Hiroshima bombing', 102), but then fail to distinguish such 'beneficial' cases from those that are 'not beneficial', except in their consequences. Unless one is willing to fully reduce rationality to the assessment of consequences, such an approach does not answer the 'what mistakes?' criticism and thus leaves the justificatory question of behavioural public policies unanswered. Alex Voorhoeve, in his commentary on Rabin, makes a similar point: that the 'descriptions of these biases as 'irrational' is not always appropriate – sometimes, for example, they are merely a form of preference change' (142). In those cases, the justification from irrationality for intervention collapses.

Finally, I should mention that not all of behavioural economics relies on rationality/irrationality divide for the justification of interventions.

https://doi.org/10.1017/S0266267115000243 Published online by Cambridge University Press

Although the heuristics and biases tradition – with its focus on irrationality – has been dominant in behavioural economics, some of its research has instead focused on enriching mainstream economic models by a few relevant causal factors. Research on altruism comes to mind here. In this anthology, Frey's work on crowding out and Gwyn Bevan & Barbara Fasolo's work on reputation effects fall into this category. Both these approaches are untroubled by the policy justification problem, because they (i) have a clear policy objective (e.g. higher productivity, lower hospital waiting times) that does not depend on a contentious welfare metric, and because they (ii) compare effect sizes of different types of interventions in similar environments. A distinction between rational and mistaken behaviour is not necessary here – effect size is all that counts.

To conclude, *Behavioural Public Policy* offers a wealth of material for critical reflection on behavioural economics and its impact on policy design. Not all of this material was intended for this purpose – some of the papers unwittingly divulge the deficits and troubling issues that still mark this field. This is great news for the philosopher willing to engage with these problems. To those enthusiastic to begin devising and implementing new policy tools *now*, however, this book might give pause.

**Till Grüne-Yanoff**[*]

## ACKNOWLEDGEMENTS

Financial support for this research from the Swedish Research Council is gratefully acknowledged.

## REFERENCES

Ainslie, G. 1992. *Picoeconomics*. New York, NY: Cambridge University Press.

Ellsberg, D. 1961. Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics* 75: 643–669.

Halpern, D. 2014. Nudge unit: our quiet revolution is putting evidence at heart of government. *The Guardian*, 4.2.2014. Retrieved 26.2.2014, http://www.theguardian.com/public-leaders- network/small-business-blog/2014/feb/03/nudge-unit-quiet-revolution-evidence.

## BIOGRAPHICAL INFORMATION

**Till Grüne-Yanoff** is an Associate Professor of Philosophy at the Royal Institute of Technology, Stockholm. His research focuses on the philosophy of science and on decision theory. In particular, he investigates the practice of modelling in economics and other social sciences, develops formal models of preference change, and discusses the use of models in policy decision-making. Till is also a member of the TINT Finnish Centre of Excellence in the Philosophy of Social Science in Helsinki.

[*] Division of Philosophy, Royal Institute of Technology (KTH), Brinellvägen 34, 10044 Stockholm, Sweden. Email: gryne@kth.se. URL: http://people.kth.se/~gryne/.

https://doi.org/10.1017/S0266267115000243 Published online by Cambridge University Press