Economics & Philosophy (2021), **37**, 139–161
doi: 10.1017/S0266267120000188

CAMBRIDGE UNIVERSITY PRESS

BOOK REVIEWS

Reciprocity and the Art of Behavioural Public Policy, Adam Oliver. Cambridge University Press, 2019, xvii + 194 pages.
Escaping Paternalism: Rationality, Behavioural Economics and Public Policy, Mario J. Rizzo and Glen Whitman. Cambridge University Press, 2020, xii + 496 pages.
doi:10.1017/S0266267120000188

From the late 1970s, the descriptive adequacy of rational choice models has increasingly been challenged by work in behavioural economics – the approach to economics that draws on the research methods and theoretical findings of psychology. From the early 2000s, there have been systematic attempts to apply 'behavioural insights' in the design of public policy. The main strand in this work has followed the approach of *libertarian* (or *soft* or *behavioural*) paternalism proposed by Cass Sunstein and Richard Thaler (2003) and Colin Camerer *et al.* (2003) and, under the catchier name of *nudging*, popularized by Thaler and Sunstein (2008). A string of later books have variously developed and opposed this approach. As the author of one of the books on the opposition side (Sugden 2018), I must declare an interest. The current review is of two significant new contributions to this literature. Adam Oliver is a founder of 'behavioural public policy' as a distinct research field; he presents his work as using behavioural insights in a way that is aligned with the 'liberal economic tradition' of John Stuart Mill and, in this respect, with my own work (148). Mario Rizzo and Glen Whitman are law-and-economics scholars of a classical liberal and Austrian persuasion, and are very definitely opposed to soft paternalism.

Oliver's two objectives are to 'outline the role of reciprocity from a multidisciplinary perspective' and to 'add to the efforts of making this concept more central in public policy design' (xv). The first two-thirds of his book give a clear and well-informed summary of the multi-disciplinary literature of reciprocity. The underlying thought is that a well-ordered society is structured by interpersonal relationships of reciprocity. For Oliver, this is a behavioural insight that has been unduly neglected.

How is this insight to inform public policy? Oliver gives some rather broad-brush answers to this question. His main concern is with traditional forms of public service provision, particularly health care, social care and education. He argues that, to ensure general acceptance of the principle that these services should be publicly provided, it is important that they are presented as schemes of social insurance, rather than as transfers from those who are not in need to those who are. Because (according to Oliver) reciprocity works most effectively in small and homogeneous communities, we should decentralize the political control of service provision. Oliver distances himself from soft paternalism in a very Millian way, declaring that paternalistic intervention in matters of private behaviour is not a proper role of public policy: 'The role of policy makers ought to be to provide opportunities for people to flourish – for them (i.e. the people) to pursue meaning and fulfilment to and in their lives … as they see fit' (157). Oliver also offers some smaller-scale behavioural insights, such as that messages to encourage people to register as organ donors should emphasize that everyone is a potential recipient of an organ transplant.

A running theme in the book is Oliver's disagreement with his LSE colleague Julian Le Grand about how to ensure quality in public service provision. Le Grand has famously argued that institutions should be designed to function effectively even if the individuals who populate them are 'knaves' (rational egoists) rather than 'knights' (altruists). Le Grand is an advocate of 'quasi-market' mechanisms of public service delivery. The idea is that services are low-priced or free at the point of use, but that service providers, instead of being directly funded as suppliers, compete to attract the demand of publicly funded customers or clients. Thus (it is claimed) demand-led competition will motivate knavish providers to achieve quality and efficiency. Oliver objects that demand-led competition encourages egoism by public sector professionals and crowds out their intrinsic motivation to serve the public, when better-designed institutions could harness motivations of reciprocity.

Oliver seems to be thinking here about reciprocity between the government (as employer) and public service providers (as employees), rather than reciprocity between providers and users. He favours trusting professionals to judge public service quality rather than 'placing too much weight on feeding the wants of the demand-side' and so running 'the danger of allowing the (sightless) tail to lead the dog' (134). Opposing schemes of performance-related pay, Oliver favours paying workers 'a decent basic wage' and relying on their 'autonomous preferences' for delivering quality and efficiency (131, 136–137). Underlying all this is a fundamental presupposition that public services have *public* objectives that are properly determined in the political process. Oliver argues that public service professionals have special expertise in judging how best to achieve these objectives, and need not be constrained by what their clients actually want. Perhaps this is not paternalism: Oliver might reasonably say that the professionals are aiming to achieve public objectives rather than to benefit their clients as individuals. But it is hard to see why this kind of professional–client relationship is more reciprocal than the relationship between sellers and buyers in Le Grand's quasi-markets.

I think Oliver has a blind spot in not recognizing that markets are networks of voluntary interactions within which participants achieve mutual benefit. It is not self-evident (as Oliver sometimes seems to suggest) that the institutional structure of the market induces people to act on self-interested motivations. The demand-led competition that Oliver so much dislikes is a process in which suppliers compete to offer customers what they want to buy. If suppliers intend their transactions with their customers to be mutually beneficial, they are acting on intentions for reciprocity.

Moving outside the area of public service provision, and in opposition to the paternalism of nudges, Oliver commends the more robust approach of *budges*. A budge is justified when some activity by a firm 'relies on its effectiveness by being informed by behavioural insights, and where its effectiveness imposes potential harms on others' – or in other words, when *a firm* is using a nudge in a harmful way. The budge itself is a prohibition of that activity. One of his examples is the common practice by supermarkets of displaying confectionery in positions that tend to attract casual buyers. According to Oliver, if customers '[buy] more confectionery than they otherwise would [and] more than is good for them', then a regulation against the placement of confectionery products close to supermarket checkouts can be justified as correcting a negative externality (161–162). Perhaps surprisingly (but, it must be said, in agreement with Mill (1972 [1859]: 150–151)), Oliver does not treat this kind of regulation as paternalistic. I take it that this is because the regulation is supposed to prevent the firm from harming the customer, rather than preventing the customer from harming herself. Again, I think, Oliver is overlooking the mutuality of market transactions. His proposed regulation places a relatively small but deliberate obstacle in the way of a potential meeting between a buyer and a seller. Why does it matter whether the immediate effect of that obstacle is on the buyer or the seller?

A more fundamental problem, common to both budging and nudging, is contained in the claim that the supermarket regulation corrects a negative externality. On the usual definition, an economic transaction has a negative externality if it imposes a welfare loss on a third party, 'welfare' being measured by reference to the affected person's preferences. Leaving aside the question of whether the supermarket customer is a third party rather than a party to the transaction that generates the effect, *which* and *whose* preferences are relevant here? The answer seems to be that these are the true preferences of the customer herself, who truly prefers not to buy the confectionery. But what is the criterion of truth? Empirical psychology can explain why a person's tendency to buy something depends on the prominence with which it is displayed, but it cannot tell us whether he or she *truly* prefers to buy it.

This is just one of the long catalogue of objections to behavioural paternalism compiled by Rizzo and Whitman. Explaining the underlying structure of their book, they ask 'What's wrong with behavioural paternalism?' and continue: 'Our case will consist of a series of challenges – in effect, hurdles that behavioural paternalist proposals must clear in order to be justified as a matter of policy' (16). The book could well be called *The Case against Behavioural Paternalism*. It is long (439 pages, plus 36 small-print pages of references) and contains a huge amount of detailed discussion of behavioural research. Undoubtedly, as the authors say, it is the product of many years of thinking about the topic (xi). But I can't help feeling that they could have made a more forceful case in a shorter book.

The running metaphor of the book is of 'puppet masters' and 'puppets'. This metaphor stands both for the relationship between behavioural paternalists and the people whose behaviour they wish to improve, and for the relationship between economists and the abstract models they use to understand the world. The suggestion is that both relationships are problematic, and for similar reasons (2).

The book starts well, with three important lines of argument. The first is that, in its origins, the neoclassical analysis of rationality 'was not fundamentally a normative project' (41). It originated in early 20th century behaviourism and in the mathematical philosophy of Bourbakism. The aim was to produce logical and observational foundations for preference concepts and thereby for a pre-existing body of economic theory. But, having found evidence of contraventions of the neoclassical theory, behavioural economists have unquestioningly interpreted those contraventions as violations of rationality in a normative sense. It is by this false move that behavioural economics finds a role for paternalism in helping people to avoid decision-making errors.

The second line of argument is similar to the 'inner rational agent' critique of behavioural welfare economics presented by Gerardo Infante *et al.* (2016). Behavioural paternalists construe decision-making error as failure to act on one's true preferences, and assume that each individual has true preferences that satisfy the consistency properties of neoclassical theory. True preferences are (it is supposed) revealed in the choices that individuals would make in the absence of psychologically induced 'errors' and 'biases'. These preferences are then interpreted as indicators of individuals' welfare, 'as judged by themselves' (a clause that is now in common use). But the concepts of true preference and error lack psychological foundations: the theoretical principles of behavioural economics give us no reason to expect that true preferences exist.

The third line of argument is the claim that there *is* a genuinely normative concept of rationality, *inclusive rationality*, defined as 'purposeful behaviour based on subjective preferences and beliefs, in the presence of both environmental and cognitive constraints' (26). Rizzo and Whitman are convincing in describing many diverse ways in which a person can be inclusively but not neoclassically rational. These include many forms of behaviour that behavioural paternalists have classified as error. In a passage that illustrates the two meanings of the puppet analogy, Rizzo and Whitman say: 'Behavioural researchers have made the mistake of conflating their models with reality – and, when reality fails to conform to the model, judging it deficient' (180). They acknowledge that, in these arguments, they are drawing heavily on Peter Todd *et al.*'s (2012) analysis of 'ecological rationality', but they claim that inclusive rationality is a broader concept (27). Rizzo and Whitman maintain that *some* types of behaviour are irrational, even according to their inclusive standard, but say that it is difficult 'to clearly and definitively identify which actions are, in fact, mistakes' (17). However, they give almost no guidance about how, even in principle, one might define and identify mistakes. I have a strong hunch that, were they to try to develop this idea, they would run into the same 'inner rational agent' problem as do the behavioural economists they criticize.

I think Rizzo and Whitman would have done well to stick with and expand on these three strong lines of argument. Instead, they seem to want to find as many objections as possible to behavioural paternalism. They end up weakening their case.

One way in which I think they go astray is by using what they call 'a series of "even if" arguments' (399). For example, they have argued that the true-preference welfare standard of behavioural economics assumes the existence of preferences that do not in fact exist, and that the non-existence of those preferences is a consequence of basic properties of human psychology. But they devote a chapter to arguing that, if true preferences *did* exist, policymakers would lack the knowledge necessary to discover them. Another example: They have argued that many of the effects that behavioural economists have attributed to errors and biases are in fact evidence of the workings of inclusive rationality. But they devote another chapter to arguing that, if these effects *were* the result of error and bias, policymakers would be subject to those errors and biases too. Rizzo and Whitman correctly describe their argumentative strategy as 'immanent critique' (311). But it could also be called overkill.

Rizzo and Whitman go further astray in a chapter that questions the scientific status of behavioural research. In so far as their criticisms are directed at the external validity of the findings of this research, and at the theories that have been proposed to explain them, Rizzo and Whitman are on firm ground. It is a fair criticism of behavioural economics that many of its experimental designs presuppose that subjects comply with particular properties of rational choice theory. (For example, investigations of behaviour in games often assume that subjects act on well-specified beliefs which can be elicited in incentivized experimental tasks.) It is also true that most of the best-known theories in behavioural economics use rational choice theory as a template and introduce psychological effects as add-on biases. The Gigerenzian critique of this approach has substance, even if behavioural economists can offer reasonable counter-arguments. But it is a different matter to suggest that there is an inadequate evidence base for the effects that behavioural economists have observed in experiments and surveys and have tried to explain.

In trying to show that 'in many cases . . . the evidence for a bias is flawed or incomplete' (91), Rizzo and Whitman make some unjustified claims about experimental work in behavioural economics, and rely on a highly selective use of references. For example, the claim that 'Most, though not all, behavioural research demonstrating biases of decision-making rests on hypothetical choice experiments' (196) is surely wrong. The claim that there is a 'scarcity of genuine attempts at replication' (199) seems to rest on a misunderstanding of the process by which discoveries are made. It is true that few published papers consist entirely of exact replications of previous experiments. But for all the well-known 'effects' that feature in behavioural economics, replication is going on all the time. Take the case of the disparity between willingness-to-pay and willingness-to-accept valuations, which Rizzo and Whitman treat as a case of flawed or incomplete evidence (110–111, 196–197). There has been a huge amount of experimental work investigating the conditions under which this disparity is greater or less, the extent to which it decays with different kinds of experience, and the relative significance of possible explanatory factors. Every such experiment is a (usually) successful or (occasionally) unsuccessful replication of the basic experimental paradigm.

Rizzo and Whitman are uncritical in accepting the fashionable judgement that there is a 'replication crisis' in experimental economics. This judgement seems to depend on the idea that scientific journals are repositories of certified statements about general empirical regularities. But journals may be better thought of as repositories of certified reports of specific experiments, combined with conjectures about possible regularities that the experimental results might exemplify and that other scientists might be interested in investigating further. In a scientific community in which knowledge expands through an exchange of observations and ideas, we should not expect every reported experimental result to be replicable.

Having reached the start of the final chapter on page 398, the reader might reasonably expect to learn what, according to the authors and if their account of rationality and error is correct, public policymakers should do. But after a recapitulation of the criticisms of behavioural paternalism made in previous chapters, there are 28 pages which explain why various apparent 'escape routes' from these criticisms fail. That leaves less than seven pages for 'Recommendations'. We are told that we should 'replace puppet rationality with inclusive rationality' and that we should 'give individuals the benefit of the doubt when making judgments about their interests'. What behavioural economists need is 'a strong dose of

https://doi.org/10.1017/S0266267120000188 Published online by Cambridge University Press

humility'; they should 'approach [humanity] as fellow human beings doing the best they can, trying to improve their own choices, and offering friendly advice on how others might do the same' (433–438).

This is all sensible enough, but it seems a poor return on such a long Case Against. Some of the most important questions about Rizzo and Whitman's approach remain unanswered. Given the standard of inclusive rationality, how are mistakes defined and identified? Recommending caution to policymakers is all very well, but what are they to be cautious in aiming for? On a natural reading, Rizzo and Whitman's stance is one of *laissez faire*. But *laissez faire* arguments are typically based on theories about how the mechanisms that are to be left alone tend to produce good results. Welfare economics has familiar theorems about the workings of competitive markets in a world of neoclassically rational individuals. It is natural to ask what kinds of economic institution work well, and in what sense, when individuals are inclusively rational.

Robert Sugden
University of East Anglia
Emaill: r.sugden@uea.ac.uk

References

Camerer C.F., S. Issacharoff, G. Loewenstein, T. O'Donoghue and M. Rabin 2003. Regulation for conservatives: behavioral economics and the case for 'asymmetric paternalism'. *University of Pennsylvania Law Review* **151**, 1211–1254.

Infante G., G. Lecouteux and R. Sugden 2016. Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics. *Journal of Economic Methodology* **23**, 1–25.

Mill J.S. 1972 [1859]. *On Liberty*. London: Dent.

Sugden R. 2018. *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford: Oxford University Press.

Sunstein C.R. and R.H. Thaler 2003. Libertarian paternalism is not an oxymoron. *University of Chicago Law Review* **70**, 1159–1202.

Thaler R.H. and C.R. Sunstein 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Todd P.M., G. Gigerenzer and **ABC Research Group** (eds) 2012. *Ecological Rationality: Intelligence in the World*. Oxford: Oxford University Press.

Robert Sugden is Professor of Economics at the University of East Anglia, Norwich. His research uses a combination of theoretical, experimental and philosophical methods to investigate issues in behavioural economics, normative economics, choice under uncertainty, the foundations of decision and game theory, the methodology of economics, and the evolution of social conventions.

Measuring Utility: From the Marginal Revolution to Behavioral Economics, Ivan Moscati. Oxford University Press, 2019, vii + 326 pages.
doi:10.1017/S026626712000019X

Some philosophical problems seem to be in science to stay. The problem of measuring utility is one such problem for economists. At its heart lies the challenge that because it cannot be directly observed, utility cannot be measured