Introduction

As noted in the Preface, this book forms the final part of my trilogy on behavioural public policy. The first book in the series, *The Origins of Behavioural Public Policy* (a title inspired by Charles Darwin's *On the Origin of Species*) was followed by *Reciprocity and the Art of Behavioural Public Policy* (a title inspired by Robert Pirsig's *Zen and the Art of Motorcycle Maintenance*). The title of the current volume, as also mentioned in the Preface, was inspired by Mill's *Principles of Political Economy*.[1] By way of introduction, before proceeding to the main body of the current text, it is therefore apt, I think, to detail briefly where, in my discussions of behavioural public policy, I have so far been, and where I intend to go.

WHERE I HAVE BEEN

My aim in the *Origins* book was to outline the foundations and development of the relatively new subfield of behavioural public policy.[2] Given that behavioural public policy is defined as the use of behavioural economics, and behavioural science more broadly, to inform public policy design, that book necessarily included a brief summary of the development of behavioural economics, and indeed of standard rational choice theory also.[3] In that vein, broadly speaking, the book considered two principal lines of inquiry: namely, the early challenges that questioned the descriptive validity of the axioms of expected utility theory (e.g. the Allais and Ellsberg paradoxes), and the suggestion that people employ decision-making heuristics – which challenges the assumption that they maximise expected utility – and which is associated heavily with Herbert Simon and, later, Daniel Kahneman and Amos Tversky.

The behavioural phenomena – present bias, loss aversion, anchoring, reciprocal motivations, etc. – that had been written about intuitively hundreds of years ago by writers such as Adam Smith (1759/2009), were validated empirically by modern behavioural economists and psychologists in those lines of inquiry, and it is these phenomena that form the building blocks of behavioural public policy. In the *Origins* book, I discussed many of the questions that these behavioural scientists had posed for standard economic and rational choice theory assumptions pertaining to time, to utility, and to money, while at the same recognising – indeed, advocating – that the foundations of behavioural public policy are built not just from contributions by economists and psychologists, but from all of the other social sciences, from the humanities, and from branches of the natural sciences (e.g. biology and zoology) also.

I also reviewed some of the principal conceptual behavioural public policy frameworks that had, at that point, been developed, and highlighted that hard and, particularly, soft forms of paternalism had dominated – and still dominate – behavioural public policy (at least in terms of the rhetoric if not actual policy applications).[4] This paternalistic focus is, I believe, a mistake, for reasons that I will consider in some depth in this third book of the trilogy.[5] At the end of the *Origins* book I chose to highlight how reciprocity, in contrast to the selfish egoism that is typically cited as a characteristic of rational choice, is perhaps the most powerful human motivational force, as recognised by its prominent place in the discourses within a range of disciplines. Yet it seemed to me that the consideration of how to use reciprocal motivations as an input to design better public policies had been strangely lacking.

The discourse on human motivation with respect to policy design has instead tended to focus on the dichotomy between selfish egoism and pure altruism. I argued in the *Origins* book that encouraging selfish egoism, even if done so with the intention of improving public sector performance, may well damage the ethos on which the provision of good public services is based, and relying on pure

altruism, which is a relatively rare form of human motivation, is naïve. Reciprocity assumes that we give and take, rather than unconditionally give or take, and serves to strengthen cooperative activities within groups. Therefore, it is important for any group-based dynamic that this tendency is not crowded out by policy interventions that assume that people are predominantly egoistical.

I will not be revealing any secrets by stating that the principal focus in the second book in my trilogy was reciprocity. The intention in that second book was to continue and greatly expand upon the themes that I introduced briefly at the end of the first book. The main objectives of the second book were to review reciprocity from a number of different disciplinary perspectives, and to explore in more depth how policy design might be informed by that motivational force. I began the book by contending that reciprocity lies deep within the human psyche (and in a basic form, is observed in the behaviours of many non-human species also), with there being evidence that very small children show tendencies towards both positive and negative reciprocity.[6]

I argued that the fundamental reason why people (and some other animals) act reciprocally is to bring forth benefits and protection to the group, and, by extension, to the individual members of the group. Thus, social norms of both positive and negative reciprocity evolved from a form of self-interest – albeit a more long-term form of self-interest than that associated with selfish egoism, the latter of which may serve some people very well in the short term, but which can erode group cohesion and ultimately be of detriment to one's longer-term objectives. Labelling positive reciprocity as reciprocal altruism is common, but is therefore something of a misnomer, because it is unlikely that a purely altruistic urge underpins this motivational force.

In the *Reciprocity* book I also noted that a further way to discern how deep within our species this or any other motivational force lies – in addition to studying the behaviours of human children and non-human animals – is to attempt to glean some understanding of

how our ancestors lived, which can by proxy be done by observing extant tribal communities. It has widely been reported that both positive and negative reciprocity, in relation to, for example, the sharing of meat, childcare responsibilities and wisdom, and as a means of building obligations and discouraging power grabs, are crucial to the proper functioning of those communities.

In addition to touching upon evolutionary psychology, animal behaviouralism and anthropology, I referred to some of the work that behavioural economists, often in collaboration with psychologists, have undertaken on reciprocity, principally using economic games under controlled laboratory conditions. If one leans heavily on reciprocal motivations in building a conceptual framework for behavioural public policy, it is obviously important to consider the now quite extensive body of work that behavioural economists – principally, Ernst Fehr and his colleagues – have undertaken on the topic. These experiments demonstrate that the extent to which reciprocity is observed and sustained is likely to be dependent on context, with repetition of the game, the inclusion or not of punishment, anonymity between partners, whether the money on offer is windfall or earned, and many other factors having an influence. Therefore, although the tendency to act reciprocally appears to be almost hardwired within us, the context – the environment – that people face can crowd in or crowd out this motivational force. Thus, the two-word essence of behavioural public policy is evident – i.e. context matters – and thus an environment conducive to the reciprocal instincts needs to be in place if they are to produce their full benefit.

It ought to be admitted – and it was indeed emphasised in the second book – that there are also some devilish aspects of reciprocity, including, but not limited to, its potential to strengthen mutually reinforcing subgroups who discriminate against perceived outsiders, its capacity to generate power imbalances by imposing unwanted obligations and, with respect to negative reciprocity, the possibility of instituting seemingly endless spirals of retaliation. The increased polarisation in the political discourse in several countries over recent

years, with social media and a heightened readiness by certain politicians and their followers to stoke widespread paranoia playing an interrelated disharmonising role, demonstrates that reciprocity and cooperation within particular groupings poses an ever-present threat to the social fabric of society as a whole. But if harnessed carefully so that it helps to create and sustain opportunities for all people to flourish, reciprocity can serve substantively as a beneficial force. In the *Reciprocity* book, I discussed some ways in which the policy-making environment, and society in general, might be designed in order to realise the benefits of our reciprocal instincts. These included an emphasis on public policy decentralisation, because securing reciprocal motivations and abating egoistical ones is easier in relatively smaller groups, and the reversal of the income and wealth concentrations in very small percentages of populations that have become much more pronounced in many countries over the last four decades, because they erode social cohesion and may legitimise unconditional "taking".

At the end of the book, I briefly introduced a new political economy of behavioural public policy informed substantively by reciprocity, one that eschewed the paternalism that is central to the dominant forms of behavioural public policy to date. I argued that affording people a great deal of individual autonomy within some very broad environmental parameters (e.g. localised decision-making, reasonably equitable income and wealth distributions) is an effective way by which to secure social cooperation, and will help people to flourish in both achieving the objectives of their public sector services, and in them realising their multifarious privately held desires in life. That said, I recognised that affording people a great deal of individual autonomy will likely result in those who are egoistically inclined to attempt to exploit others by using, implicitly or explicitly, behavioural-informed tactics. Therefore, in addition to offering a great deal of autonomy as a means of nurturing reciprocity among those who impose no substantive harms on others, I suggested that an overarching conceptual framework of behavioural public policy must

also include a second arm, one that calls for regulations against behavioural-informed harms to others – i.e. the threat of a form of negative reciprocity. The framework that I suggested fits within the liberal tradition, particularly that postulated by Wilhelm von Humboldt (1791–1792/1993) and John Stuart Mill (1859/1969).[7] The principal objective of this, the final part of the trilogy, is to further develop my preferred overarching political economy of behavioural public policy.

WHERE I AM GOING

As already stated, the objective of the book you have before you is to build an overarching conceptual framework – a political economy – of behavioural public policy that gives guidance on where applications of this relatively new subfield of public policy are legitimate. This is the first all-enveloping policy framework in this policy domain and is a framework that is offered in the spirit of the liberal tradition.[8] First, however, I offer a disclaimer, with reference to the circumstances that we are living through at the time of writing. This is not a book about the Covid-19 pandemic or climate change or any other (relatively fleeting or more sustained) policy challenge. I do refer to these and other policy challenges throughout the book by way of illustrating the various arguments, but to have focused on, for example, the Covid-19 pandemic would have given the book a more limited shelf life, and ultimately a narrower influence, than I hope for.[9] Rather, the book provides a framework within which the legitimacy of any policy challenge that is, at least in part, addressable with behavioural public policy interventions can be considered; and that probably encompasses all conceivable policy challenges.

Typically, the paternalistic vision for behavioural public policy is that people, due to the behavioural phenomena that can influence their choices, make mistakes. That is to say, people often choose, decide and behave in violation of that which maximises their lifetime welfare (or utility), that they are thus being irrational, and that there is therefore a justification for manipulating, or even coercing, them for

their own good. However, if the behavioural phenomena that are frequently labelled as "biases" are so ingrained in the human psyche, then presumably they must have evolved for good reasons, and therefore, strictly speaking, they may not be biasing influences at all. It may be the case that other parties take advantage of their implicit or explicit knowledge of these influences in order to manipulate people into doing things that they would rather not do, but in the absence of such manipulation, and given that people may have multifarious objectives in their own personal lives (with many of those objectives having little to do with welfare or utility maximisation), can a third party really conclude that their actions and behaviours are mistaken? A challenge to the assumption that the behavioural phenomena are necessarily biasing influences, and thus an implicit challenge to the legitimacy of paternalistic behavioural public policy, is laid out in Chapter 1.

In Chapter 2, however, I feel that it is only fair to summarise the parameters of some of the existing (albeit narrower) behavioural public policy frameworks, partly so that readers can reach their own conclusions with respect to the legitimacy of each framework and partly to serve as a reference against which my proposed political economy can be contrasted and compared. In Chapter 3, I note that several scholars have expressed support for soft forms of paternalism by arguing that humans are limited in their reasoning abilities: namely, that they (or rather, we) are limited in imagination, willpower, objectivity and technical skills. These scholars argue that, as a consequence, individuals sometimes fail in their pursuit of that for which they ought to be striving – specifically, more welfare, utility or happiness. Therefore, they need the guiding hand of the policy maker.

Countering these claims, in Chapter 4 I question whether utility (or welfare or happiness) maximisation is a legitimate general normative goal of public policy by arguing that, often, the pursuit of utility does not drive desires. Rather, desires typically precede any consideration of utility. Desires, I contend, are multifarious and vary

across people and, with a nod to Chapter 1, may be meaningfully facilitated by the behavioural influences. Thus, I contend that over the private realm of individual decision-making, the policy maker's role should not equate with being a utility-maximising social planner; rather, a policy maker should seek to secure the general conditions that facilitate people in their pursuit of their own conception of the desired life. Building on the arguments I initiated in the *Reciprocity* book, I contend in Chapter 5 that this can best be achieved by allowing people a great deal of autonomy. Liberals (such as myself) maintain that autonomous actions are the most effective way of forging the social cooperation from which we all benefit.

In Chapter 6, I suggest that reciprocity is linked heavily to perceptions of desert, which itself underpins notions of justice. This again therefore may underline how a motivational force that arose organically to benefit the individual by protecting the group to which he belonged evolved into a normative concept.[10] I argue that desert-based reciprocity is informed by considerations both of intentions and outcomes, which is useful when thinking about how public support for public sector services and broader welfare systems might be maintained. In Chapters 7 and 8 I continue the theme of how one might use reciprocity to inform policy design, first over the domain of private individual decision-making and then over the domain of public sector decision-making. For the former, due to the differing and multifarious desires that people may have (including the desire to start one's own business and the desire to escape from a poor-quality service), I conclude that disallowing the competitive market would be too much of an infringement on individual liberty, but for the latter, where public sector services exist to deliver a limited range of a priori collectively agreed-upon goals, the risks associated with market failures and the exploitation of the behavioural influences ought to cause us to be more circumspect of the potential net benefit of demand-led competition. In these circumstances, it is argued, and from the perspective of my political economy of behavioural public policy, reputational competition is more legitimate.

In Chapter 9, I elaborate on how policy might be used to regulate against those who, through self-serving actions, abuse their autonomy by imposing behaviourally informed harms on others. Typically, these actions will distort the mechanism of a fair exchange and will thus undermine the very notion of a reciprocal relationship, but sometimes they involve inaction rather than action, when action could deliver easily won benefits. This arm of my political economy of behavioural public policy includes regulations against both negative externalities and regulations to produce positive externalities; in short, it proposes that in order to protect freedom in general, some specific freedoms will need to be curtailed.

Thus, as aforementioned, this book explicates the political economy of behavioural public policy that I initially mooted in the *Reciprocity* book, but the short synopsis of that framework remains the same. That is, that policy makers ought to nurture the reciprocal instincts that evolved among humans for group and individual benefit by influencing the very general conditions of society that facilitate that objective, while at the same time regulating those who harm others by using the behavioural influences to exploit the opportunities that have been granted them. I hope I have not given too much away in this short introduction to dissuade the reader from reading on, but that I have given away enough to entice the reader to read on, and, indeed, to read "back", if they have not yet read the first two books in the trilogy.

1 Setting the Scene

In his introduction to the Pelican Classics edition of John Stuart Mill's *Principles of Political Economy*, the economist Donald Winch wrote that:

> Like all liberal theorists, [Mill] took the individual as the basic unit of discourse. His contact with traditions antagonistic to the one in which he was brought up merely served to strengthen his attachment to individualism by enlarging his conception of what individuality should comprise. Institutional arrangements in society should be judged basically in terms of whether they enhanced this individuality by widening the sphere of independence and choice. In so far as social, political and economic conditions inhibited or prevented individuals, or groups of individuals such as the working classes, from partaking fully in the benefits of the social union, these should be removed by direct intervention or negative prohibition.
>
> *(Winch, 1970, p.48)*

In essence, Mill, according to Winch, believed that in order to protect liberty in general, some specific freedoms ought to be constrained, a view that had also been held by John Locke. From a behavioural public policy perspective, this will also be my conclusion in this book.

However, before reaching my conclusion it may prove instructive to detail how I arrived at it. As noted in the Introduction, in a previous book I contended that the motivational force of reciprocity – of responding in kind to good, and bad, intentions and/or actions – can and should be nurtured by policy makers to aid individuals in the pursuit of their own private predilections and public sector groupings in the pursuit of their collective objectives (Oliver, 2019). To borrow

from the economist Amartya Sen (1999), to the extent that public sector services, such as health and education, provide people with the capabilities to pursue their privately held goals in life, an environment that crowds in reciprocal motivations in those sectors is perfectly consistent with sustaining and extending liberty. As I previously contended, the urge to act reciprocally – and a concern with that which facilitates indirect reciprocity (namely, a good reputation) – lies deep within the human psyche, and probably evolved because this motivational force brings forth benefits and protection to the group.[1] Moreover, and importantly, since the individuals that comprise a group are more likely to fare well if their group is flourishing, a reciprocal cooperative spirit is compatible with – indeed, is probably principally driven by – the pursuit of individual long-term self-interest. Hence, out of this evolutionary process arose instinctive tendencies and social norms that favour conditional cooperation and that justify punishment of those who transgress.

Admittedly, that subgroups often act cooperatively and reciprocally to the detriment of the wider group is an unfortunate possibility that has long been recognised. The Scottish Enlightenment philosopher David Hume, for instance, wrote that 'Factions subvert government, render laws impotent, and beget the fiercest animosities among men of the same nation, who ought to give mutual assistance and protection to each other' (Hume, 1777/2018, p.155). Hume elsewhere noted that 'Robbers and pirates . . . could not maintain their pernicious confederacy, did they not establish a new distributive justice among themselves, and recall those laws of equity, which they have violated with the rest of mankind' (Hume, 1751/2018, pp.33–34). There are also risks associated with negative reciprocity, including undue or excessive retribution and spiralling retaliation, but if harnessed in the right way reciprocity can serve substantively as a force for good, as also emphasised by Hume, as we shall later see.

Reciprocity as a motivational force can, and sometimes is, embraced, if peripherally, by several different behavioural public policy frameworks, but is central to that favoured in this book;

namely, to nurture reciprocity in the positive sense so that people may be able to better pursue their own conception of a flourishing life, and also in the negative sense, to constrain those who might otherwise exploit the liberty that has been granted to them and in doing so impose unacceptable harms on others.[2] Given the centrality of reciprocity to this framework, it seems apt at this point to reflect a little further on the concept.[3]

THE ORIGIN OF RECIPROCITY

Those who write on the evolutionary origin of reciprocity present varied, if related, arguments. The evolutionary biologist Joseph Henrich, for instance, sees reciprocity as underpinning the mutual protection that became ever more necessary after our ancestors descended the trees and became ground apes (Henrich, 2016), whereas the ethologist and primatologist Christopher Boehm notes that reciprocal tendencies strengthen as a necessary feature of insurance when individual success in a hunt is uncertain (Boehm, 2012).[4] Boehm reports that when chimpanzees hunt, those that gain initial control of the carcass will share just enough to enable them to retain control, and there may be reciprocation between givers and receivers when their relative success (or lack thereof) is reversed in the future.[5] According to Boehm, archaic homo sapiens killed larger game than do chimpanzees, and thus there was more sharing, and in hunter-gatherer societies dominance over meat was often negated entirely by having it shared out by a neutral person (see also Sapolsky, 2017, p.323).

In *Utilitarianism*, Mill (1863/1969) wrote that we control ourselves in the face of internal and external sanctions, with the internal mediated by our conscience, but that our feelings for others, where they exist at all, are much weaker than our feelings for ourselves. Mill further contended that where feelings for others exist they do so due to a concern for their utility – a concern for *them*. One might acknowledge Mill's view, but it seems that a concern for others evolved because that is the best means to serve oneself, at least in the long