

ARTICLE

The authors of our own lives: the limitations of the behavioural justification for paternalism

Adam Oliver

Department of Social Policy, London School of Economics and Political Science, London WC2A 2AE, UK
Email: a.j.oliver@lse.ac.uk

(Received 31 January 2023; accepted 31 January 2023; first published online 27 February 2023)

Abstract

In this article, the confidence that has been placed in hard and, in particular, soft paternalistic measures in the field of behavioural public policy is questioned. The four purported limitations of human reasoning – i.e. limited imagination, willpower, objectivity and technical ability – are considered, but ultimately it is concluded that these are insufficient justifications for paternalistic intervention, for two principal-related reasons. First, it is impossible for a policy maker to discern what people desire for their own lives, and second, so long as they are not harming others, people ought to be free to pursue their own desires. The vision for the future of behavioural public policy proposed here is thus consistent with classical liberal, and in particular, Millian thought: i.e. aim to educate people on the pros and cons of their actions and inactions so that they are better equipped to live the lives they wish to lead but do not interfere directly in guiding them towards any particular end.

Keywords: autonomy; behaviour; liberalism; liberty; paternalism

What is paternalism?

At a symposium at the London School of Economics in 2019, Lord Gus O'Donnell, the former head of the British Civil Service under Prime Ministers Blair, Brown and Cameron, stated that policy makers are all unabashed paternalists. In other words, they believe that they know better about what is best for people than do the people themselves.¹ This statement is unsurprising as it is perhaps this belief that attracts many policy makers to their vocation to begin with, but in this article I aim to challenge the confidence that one may place in paternalistic measures, specifically

¹Lord O'Donnell was not referring to policies directed at the very general ways in which policy makers must structure society to serve particular purposes, in relation to, for example, taxation or defence. He was referring to interventions that are intended to correct for supposed individual decision-making errors largely pertaining to personal lifestyle choices (e.g. financial and health-related decisions). Those sorts of decisions have been the main focus of behavioural public policy to date and are the also the focus of this article.

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

through the lens of the field of behavioural public policy (i.e. the application of behavioural science to policy design and implementation), where paternalism, at least in terms of the rhetoric, has thus far been dominant.

Before offering my critique of paternalism, a more precise definition of what the term denotes is necessary. Dworkin (1972: 65) defined paternalism as the ‘interference with a person’s liberty of action justified by reasons referring exclusively to the welfare, good, happiness, needs, interests or values of the person being coerced’. With the exception that ‘welfare, good, happiness, needs, interests or values’ be replaced with ‘utility, welfare or happiness’ (because most of those working in the field of behavioural public policy implicitly or explicitly embrace the maximisation of the latter as their normative goal), and that paternalism here is extended beyond acts of coercion to include also those of manipulation, Dworkin’s definition is taken as given in this article. While any actor can act paternalistically towards any other actor, this article is principally concerned with paternalistic intervention by the State.

As intimated, hard (or coercive) and, in particular, soft (or more manipulative) forms of paternalism have dominated behavioural public policy to date. Both approaches maintain that people are led astray from pursuing their own best interests by biases acting upon their automatic individual decision-making; they differ, however, in that hard paternalists would ban certain activities and behaviours that, in their view, cause people to harm themselves, while soft paternalists would instead reorganise people’s decision-making environments such that their automatic choices are more likely to align with their deliberative preferences.

The biggest objection to paternalism is that it undermines individual autonomy. To understand why autonomy is important we must also have an idea of what it means. According to Le Grand and New (2015: 106), ‘... autonomous people have the capacity to think, decide, and act for themselves. If we are autonomous, we are the authors of our own lives’. Incidentally, distinctions between autonomy, freedom and liberty are sometimes made in the literature. For example, as noted, Le Grand and New (2015) contend that autonomy is the ability to act as deliberating agents, which they maintain can be undermined even if freedom – i.e. the absence of physical constraints over what one might choose – remains untouched. However, in this article, autonomy, freedom and liberty – the violation of which leads to some infringement of people’s ability to think or act of their own volition – are taken as interchangeable terms.

Respecting autonomy thus necessitates that we allow people to live their lives as they see fit, so long as they are not harming others. Raz (1986) has suggested that viewing autonomy as an end in this way is in itself a form of paternalism, but respecting autonomy does not equate to manipulating or coercing people to behave in any particular way. Rather, it leaves people free to pursue any desire, including the desire to place a limited value on their own autonomy.² Moreover, educating people, absent indoctrination or deliberate attempts to misinform or control, is not paternalistic.

²Buchanan (1991) suggests that slavery is acceptable if the enslaved deliberately choose their enslavement (e.g., as an escape from even worse material circumstances). Those adopting this viewpoint presumably therefore find paternalism acceptable if it is deliberately desired by those who are its targets.

Education informed by a neutral assessment of the latest evidence, arguments and learning enhances autonomy by enabling people to improve their reasoned judgements, yet leaves them free to decide, behave and act in whichever ways they want. Indeed, it may be contended that refusing to educate people with the latest information and knowledge, or deliberately misinforming them, whether undertaken with or without paternalistic intentions, undermines the capacities of individuals to pursue their own objectives. In essence, it places unnecessary constraints upon their liberty, a view held by John Stuart Mill, arguably the most prominent of all anti-paternalists.

Mill's antipathy

Mill's antipathy towards paternalism is clear in his *Principles of Political Economy*, in which he wrote that freedom from authoritative intrusion on human existence '... ought to include all that part which concerns only the life, whether inward or outward, of the individual, and does not affect the interests of others ... it is allowable in all, and in the more thoughtful and cultivated often a duty, to assert and promulgate, with all the force they are capable of, their opinion of what is good or bad, admirable or contemptible, but not to compel others to conform to that opinion' (Mill, 1848/1970: 306).

Although the above, in one sense, underscores Mill's anti-paternalism, that he referred to 'force' and 'coercion' might, it may be claimed, indicate that Mill was opposed only to hard forms of paternalism. Berlin (1969) stated that Mill did not identify the types of behaviours that many soft paternalists believe are irrational today – i.e. the systematic behavioural patterns that depart from the assumptions of rational choice theory. Nonetheless, the following passage may offer a reason for soft paternalists to believe that Mill would not have been averse to their vision: 'When a government provides means of fulfilling a certain end, leaving individuals free to avail themselves of different means if in their opinion preferable, there is no infringement of liberty, no irksome or degrading restraint. One of the principal objections to government interference is then absent' (Mill, 1848/1970: 306–307).

Thus, soft paternalists might contend that Mill would not oppose interventions that help people to satisfy their deliberative preferences so long as the interventions do not force those who might not have those preferences to alter their behaviours. However, although we will never know for sure what Mill's view would have been on this issue, there is reason to doubt whether he would have supported the use of behavioural insights by the State to manipulate behavioural patterns in particular directions, even if it were the case that those directions aligned with the supposed deliberative preferences of many of those targeted for behaviour change. In his *Principles*, for instance, he went on to state that: 'I have reserved for the last place one of the strongest of the reasons against the extension of government agency ... Instruction is only one of the desiderata of mental improvement; another, almost as indispensable, is a vigorous exercise of the active energies; labour, contrivance, judgement, self-control: and the natural stimulus to these is the difficulties of life' (Mill, 1848/1970: 312). Mill therefore intimates that government-sponsored behaviour change interventions may erode the capacity for people to think for themselves. Mill would not, for example, have wanted people not to have to worry about their

own self-control; he would have wanted people to learn to exercise their own self-control explicitly and entirely on their own terms, as a means of improving their capacity to live the lives they wish to lead.

That being said, of the various forms that paternalistic intervention might take, for many anti-paternalists soft forms of paternalism perhaps come the closest to being persuasive. I write ‘perhaps’ because some anti-paternalists are likely to view the covert nature of certain forms of soft paternalism as highly statist – indeed more statist, and thus more objectionable, than explicit forms of hard paternalism that can be more openly considered and challenged (see, e.g., Sugden, 2009). Nonetheless, if we assume that soft paternalism is generally perceived as less objectionable, and we remember that this form has been the most prominent in the field of behavioural public policy to date, it seems apt to consider in a little more depth the justifications offered by those who advocate for this approach.

The four limitations

Le Grand and New (2015) identify four limitations of human reasoning that they argue might distort behaviours, choices and actions, such that people often fail to end up where, on deliberation, they would ideally like to be. They define these as limited imagination, limited willpower, limited objectivity and limited technical ability, and each, they claim, gives the State a potential justification for paternalistic intervention.

Le Grand and New illustrate the limited imagination argument by referring to the tendency for many people to save insufficiently during their lives to ensure a financially comfortable retirement. Specifically, they write that: ‘The best advice on pensions is to begin contributing to a scheme as young as possible: say, before thirty. Yet it is very difficult in one’s twenties to imagine oneself in one’s sixties or seventies. It is as though one is contemplating a different person: one to whom one has a connection, certainly, but a fairly tenuous one’. Sunstein (2020) makes a similar point when he asks us to ‘Recall that choosers must solve a prediction problem; they must decide, at some point in advance of actual experience, about the effects of one or another option on experience. To solve that problem, knowing “how they feel” is not enough. At a minimum, they must know “how they *will* feel,” and they might not know enough to know that’.

One could make the philosophical argument that if we have only a fairly tenuous connection to our future self, then that future self might better be thought of as a different person. If so, then interventions intended to protect our future self from our current self are perhaps aimed at addressing an externality rather than being paternalistically driven. A different externality-related point can be made if we recognise that the most significant future harm generated by people who save little for their retirements now might not be imposed on their future selves, but on the future young, who may be required to cross-transfer far more resources than they otherwise would in order to support their elders.³ However, for this article, I will follow the

³This particular externality argument may not be so readily applied to other actions that might be argued to harm our future selves due to limited imagination. For example, it could be argued that a 20-year-old

convention that State-sponsored efforts to increase retirement savings are acts of paternalism.

Thus, those who believe that limited imagination causes mistakes in people's decision-making vis-à-vis their future selves contend that State intervention to address these errors is justifiable on paternalistic grounds. However, one may retort that it is impossible to truly discern whether those who save little for their retirements, for example, are genuinely making mistakes in relation to their own personal goals in life. For example, some of these people may have excellent foresight of their future selves and yet still prefer to live in the moment; others might deliberately prefer not to have any foresight of what the future has in store for them, and are willing to accept whatever life offers up. For those who do wish to save more for their futures, manipulating or coercing them into particular pension plans may disincentivise them to search for plans that better match their desires. There are a host of plausible reasons why paternalistic pension policy interventions do more harm than good.⁴

With respect to limited willpower, Weale (1978: 162) maintains that there may be occasions when people would like to be forced away from or towards actions and behaviours that cause harms or benefits to them. He states that: 'Citizens may prefer to be paternalistically coerced if they know they are prone to periodic bouts of the weakness of will which allow them to perform actions which they subsequently come to regret'. One could counter this statement by questioning whether the regret that people anticipate is aligned with the regret that they subsequently come to feel, and thus whether the latter really does outweigh the benefits that they would have accrued from the behaviours that they are coerced away from. This is not to argue that those who wish to be controlled should not be controlled on an individual basis, assuming, of course, that policies can be targeted at individuals rather than populations (unless, as seems unlikely, a whole population that is targeted for behaviour change wishes to be controlled, after each individual in that population has given serious thought to all of the possible implications of such a strategy). Rather, it is to propose that it is difficult to justify such coercion on the grounds of increasing welfare, the normative criteria that behavioural public policy paternalists embrace. Moreover, if an individual's desire to be paternalistically coerced (for whatever reason that is personal to him) is respected, he is not really being paternalistically coerced at all – his autonomy is maintained (as argued earlier).

Sunstein (2019: 9) takes up the theme of limited willpower (and, implicitly, some of the other limitations too): 'People might have no idea how to get where they want to go. Like Adam and Eve, they can be tempted. Sometimes they lack self-control. Background conditions greatly matter. Sometimes people's choices are not, in the

smoker may not be able to imagine himself as a 70-year-old lung cancer patient. He has imposed harms on *himself* that cannot be directly alleviated by giving him more money. That said, the costs of his health care will still substantially be borne by those who are subject to income tax when he receives his cancer treatment, so the issue of externalities at least in part perhaps remains.

⁴If one accepts that people are limited in their imaginations with respect to their future selves, could not one also make the claim that people have limited imaginations (or remembrances) of their past selves. The person in his sixties or seventies, for example, in lamenting that his younger self did not save enough for his retirement may now find it difficult to imagine the enjoyments or purposes that he was able to meet by saving less earlier in life.

deepest sense, their own; they are deprived, deceived, or manipulated. Sometimes they lack crucial information. Sometimes their preferences are a product of injustice or deprivation. Sometimes they simply blunder. As a result, their lives go much worse'. It is difficult to dispute this statement, but one might contend that it offers few grounds for the paternalistic (counter-)manipulation or coercion of citizens. Rather, the concerns may best be addressed by regulating the activities of those who would otherwise manipulate or exploit others, to educate people so that they are less likely to lack crucial information, for people to learn to discipline themselves so that they might better exercise self-control whenever they wish to, and to legislate to address unacceptable injustices and deprivations. Moreover, our blunders, as intimated by Mill (noted earlier), may serve a useful purpose in incentivising us to think, to reflect and to learn; in short, to help us to develop as human beings.

The notion that people often have limited objectivity relates to the overconfidence heuristic – i.e. that people are sometimes perceived to be overconfident regarding their abilities, decision-making skills and susceptibilities to life's uncertainties – and to so-called confirmation bias, or in other words, the tendency for people to pay most attention to evidence and arguments that support their pre-existing opinions and conclusions. Limited technical ability refers to the difficulties people may face in performing the sometimes complicated calculations that are necessary for them to make optimal decisions (from the welfare or utility maximising point of view, which assumes, of course, that such optimisation is what people want and ought to do). Taking the four limitations together, Le Grand and New (2015) conclude that there are two kinds of actions that are most subject to reasoning failure: (1) those that may cause harms in the distant future and (2) those that pose a small chance of an immediate catastrophic outcome (e.g., due to overconfidence, confirmation and poor technical abilities, people may underweight the immediate threats of climate change, and, until recently in many parts of the world, pandemic threats).⁵

Le Grand and New are sympathetic to soft paternalistic measures to address certain actions, although the original proponents of nudging go further with their approach than the two kinds of action summarised above. Sunstein (2019), for instance, argues that some construct of the environments in which people live is inevitable with or without government intervention, and will affect people's decisions. He maintains that even the weather is a form of what Thaler and Sunstein (2008) label 'choice architecture': 'On snowy days, for example, people are especially likely to purchase cars with four-wheel drive, which they do not always love' (Sunstein, 2019: 11–12). If we define choice architecture in terms of our general surroundings, then Sunstein is of course right – it is always there. But if we consider choice architecture in the sense of it being used, often covertly, by one actor to guide specifically the behaviours of another actor that align with the first actor's intended directions, then it can be problematic.⁶ Private actors do this all the time, but if they overstep

⁵Note, however, that one could equally claim that that these reasoning failures can cause people to overweight these threats.

⁶As touched up by, for example, John and Bicchieri in this issue, the liberty-preserving aspect of the nudge approach was intended to serve as a safeguard against misguided interventions, but if people are being (covertly) manipulated to alter their behaviours in particular directions, the notion that liberty is being preserved can be legitimately challenged.

the mark, their actions can be regulated. Many feel uncomfortable with public officials using similar tactics, particularly in relation to paternalistic as opposed to externality-focused interventions, not least because they undermine citizens' capacity to hold governments to account.

Sunstein also contends that since navigability – i.e. the ability to choose the best paths for ourselves in life – can pose serious problems for people, it is justifiable for the State to try to ensure that the choice architecture works in people's best interests. Some of those who oppose paternalistic measures do not deny that individuals might often need assistance or protection with respect to their own behaviours or the behaviours of others. However, as noted earlier, many anti-paternalists believe that attempts to educate people openly about the potential pitfalls they face, or to fill gaps in their knowledge, is a more appropriate strategy than reshaping people's environments often without their explicit knowledge, and if others are creating contexts that induce negative externalities then for many critics of paternalism, it is legitimate for the State to regulate explicitly against those activities. But to contend as such is not to assume that people are harming themselves just because their actions appear to conflict with the tenets of rational choice theory. People may, by and large, be pursuing the lives that they wish to lead, irrespective of how strange their behaviours and decisions might appear from the outside, and if they are imposing no substantive harms on others, the case can be made that they should not be coerced or manipulated into doing otherwise. That is the essence of a liberal society.

A return to liberalism

Mill, in his support for educating people so that they are better equipped to lead the lives that they wish to lead, and with his antipathy towards State interference in individual behaviours that impose no harms on others, offers, I contend, the quintessential form of classical liberalism. Over recent decades, however, some have equated liberalism to a school of thought that advocates for minimal State interference and for people to be guided by egoistic motivations (i.e. neoliberalism). Neoliberal thought (and interpretations of it) was perhaps (often inadvertently) nurtured by the members of the Austrian School of Economics, who offered a perspective that tended to be less interventionist than that proposed by Mill, although the belief system that many members of the Austrian School shared was somewhat more nuanced than that suggested by the neoliberal caricature.

Almost eighty years ago, Hayek expressed qualms with the way in which the concept of liberalism was being misused. He wrote: '... the essential features of individualism ... was first fully developed during the Renaissance and has since grown and spread into what we know as Western European civilisation – the respect for the individual man *qua* man, that is ... the belief that it is desirable that men should develop their own individual gifts and bents' (Hayek, 1944/2001: 14). He went on: 'Because of the growing impatience with the slow advance of liberal policy, the just irritation with those who used liberal phraseology in defence of anti-social privileges, and the boundless ambition seemingly justified by the material improvements already achieved, it came to pass that toward the turn of the century the belief in the basic tenets of liberalism was more and more relinquished' (Hayek, 1944/2001: 19).

Hayek thus appeared to be frustrated that the term 'liberalism' was being associated with egoistic self-interest. Liberalism ought to be rescued from this mischaracterisation, by reminding (or teaching) people that its approach embodies respect, tolerance and empathy for all people, and protects against those who transgress these ideals.

As alluded to above, however, the members of the Austrian School tended not to be as interventionist as Mill with respect to mitigating harms, which may be a reason why some misinterpret their attitude as being supportive of egoism. Unlike Mill, von Mises (1927/2005), for instance, did not believe in the regulation of externality-inducing personal behaviours, such as alcohol and drug addictions, because he was not convinced that such policy action would be effective, and, even if it was, he thought that legitimising State intervention could, via a slippery slope, lead to even greater harms. However, the Austrians were generally explicit in their view that there are some circumstances in which regulation against certain activities to afford protection is warranted. For instance, in offering a fairly representative view from the Austrian School, Friedman and Friedman (1980/1990: 7) wished to explore '... how we can limit government while enabling it to perform its essential functions of defending the nation from foreign enemies, protecting each of us from coercion by our fellow citizens, adjudicating our disputes and enabling us to agree on the rules that we shall follow'.

Although the classical liberals, including Mill and the Austrians, were united in their anti-paternalism, there are paternalistic interventions that some commentators suggest it is almost churlish to oppose – i.e. that fall outside the range of any legitimate anti-paternalistic concern. This is the case, some suggest, when the interference with autonomy is minimal, the potential benefits are great, and the interventions are not costly to enforce – for example, regulations on the wearing of car seatbelts and motorcycle crash helmets (see Le Grand & New, 2015). Similar arguments can be made in relation to some classic nudges, such as the strategic placement of food items on canteen and supermarket shelves to promote healthier choices. Yet there remain grounds on which to oppose such interventions, irrespective of how little impact on autonomy their advocates claim them to have. For example, can a third party, such as a policy maker, really know how intrusive are the enforced wearing of seatbelts/helmets and the strategic placement of food items? For many, it could be very intrusive indeed; for millions, it may be only slightly intrusive, but this will add up to a lot of intrusion in the aggregate. Is it unreasonable to suggest that a small infringement on the autonomy of a great many people is worse than extreme curtailments being placed on the liberty of a few? Moreover, with respect to manipulation rather than coercion, it might leave some with the uneasy feeling that, although relatively innocuous at face value, such interventions represent the unpalatable thin end of the wedge that von Mises, among others, was so concerned about (see Rizzo & Whitman, 2020: 349–397, for a discussion). One may conclude with the following question: in the realm of internality considerations, given that no third-party policy maker really knows the reasons why people do the things that they do, why not just aim to educate people on the pros and cons of their possible actions and inactions, and then leave their choices and behaviours to them?

References

- Berlin, I. (1969), 'Introduction', in J. S. Mill (ed.), *1863/1969. On Liberty and Utilitarianism*, Oxford: Oxford University Press, vii–xxxix.
- Buchanan, J. M. (1991), *The Economics and the Ethics of Constitutional Order*. Ann Arbor: University of Michigan Press.
- Dworkin, G. (1972), 'Paternalism', *Monist*, **56**: 64–84.
- Friedman, M. and R. Friedman (1980/1990), *Free to Choose: A Personal Statement*. Orlando: Harvest.
- Hayek, F. A. (1944/2001), *The Road to Serfdom*. Abingdon: Routledge Classics.
- Le Grand, J. and B. New (2015), *Government Paternalism: Nanny State or Helpful Friend*. Princeton: Princeton University Press.
- Mill, J. S. (1848/1970), *Principles of Political Economy*. Pelican Classics: Middlesex.
- Raz, J. (1986), *The Morality of Freedom*. Oxford: Oxford University Press.
- Rizzo, M. J. and G. Whitman (2020), *Escaping Paternalism: Rationality, Behavioural Economics, and Public Policy*. Cambridge: Cambridge University Press.
- Sugden, R. (2009), 'On nudging. A review of nudge: improving decisions about health, wealth and happiness by Richard H. Thaler and Cass R. Sunstein', *International Journal of the Economics of Business*, **16**: 365–373.
- Sunstein, C. R. (2019), *On Freedom*. Princeton: Princeton University Press.
- Sunstein, C. R. (2020), 'Behavioral welfare economics', *Journal of Benefit-Cost Analysis*, **11**: 196–220.
- Thaler, R. H. and C. R. Sunstein (2008), *Nudge: Improving Decisions About Health, Wealth and Happiness*. New Haven: Yale University Press.
- von Mises, L. (1927/2005), *Liberalism. The Classical Tradition*. Indianapolis: The Liberty Fund.
- Weale, A. (1978), 'Paternalism and social policy', *Journal of Social Policy*, **7**: 157–172.