

# Some challenges to the new paternalism

JULIAN LE GRAND\*

*Professor of Social Policy, London School of Economics and Political Science, London, UK*

**Abstract:** Behavioural public policy analysts have examined cases of individuals' failures of reason or judgement to attain their ends and have used these to justify 'means' paternalism: a form of government intervention that tries to save individuals from the consequences of those reasoning failures and to enable them better to achieve those ends. This has been challenged on a number of grounds, including too great a focus on choice-preserving interventions such as nudges, the privileging of future preferences over current ones and the possibility of state failures as damaging to individual well-being as the original reasoning failure. This paper summarizes the principal arguments in favour of means paternalism and then addresses these challenges.

The only purpose for which power can be rightfully exercised over any member of a civilised community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. (John Stuart Mill, 1859/1972, p. 68)

In *Government Paternalism*, a book I published with Bill New, we argued that John Stuart Mill's assertion in the quotation above was wrong (Le Grand & New, 2015). Among other things, we used the growing number of empirical results from behavioural economics to argue that, contrary to Mill's classic statement on the illegitimacy of government paternalistic intervention to promote an individual's own good, there are situations where paternalistic interventions are justified. That is, there are circumstances where the state should intervene to save people from themselves – or, more precisely, from the adverse consequences of their own decisions, even if no one else is harmed by those decisions.

We were the first to demonstrate that these arguments could be applied to all types of governmental paternalistic interventions. However, earlier, Richard Thaler and Carl Sunstein used results from behavioural economics to justify so-called 'libertarian paternalism' (Thaler & Sunstein, 2003, 2008) and the

\* Correspondence to: Marshall Institute, London School of Economics, 5 Lincoln's Inn Fields, London WC2A 3BP, UK. E-mail: [j.legrand@lse.ac.uk](mailto:j.legrand@lse.ac.uk)

use of ‘nudges’ as instruments of paternalistic policy. George Lowenstein, Matthew Rabin and others came up with similar ideas to justify what they called ‘asymmetric paternalism’ (Camerer *et al.*, 2003), and Sarah Conly used behavioural economics to support what she called ‘coercive paternalism’ (2013).

All of these contributions to what we might call the ‘new paternalism’ have been extensively challenged. One recent challenge has come from Robert Sugden (2018); I discuss his arguments elsewhere (Le Grand, 2018). Another major critic, aiming primarily at the arguments in Le Grand and New (2015), but also at behavioural economics in general, is Mark Pennington (2016); addressing his arguments will be the central point of this paper.

This paper’s next section summarizes the principal arguments of *Government Paternalism*. It then addresses some of the critiques, concentrating on those by Pennington.

### **Government Paternalism: the principal arguments**

We begin with definitions. Most of the definitions of paternalism common in the literature are unsatisfactory, for they concentrate on the state, in the name of an individual’s own good, paternalistically restricting – or indeed even forbidding – the choices that he or she makes. But interpretations of paternalism defined in this way rule out many forms of state intervention that most would consider as paternalistic, including subsidies to the arts, which increase, not restrict, the choices available to individuals, and libertarian paternalistic policies that preserve choice, such as automatic enrolment in pension schemes or in organ donation programmes.

It seems better to define paternalism in terms of *intention*. So a paternalistic intervention is one where the state intends to replace an individual’s judgement as to what is the best decision for him or her. Hence, we define a state intervention as paternalistic with respect to an individual if it is intended to address a failure of judgement by that individual and if it is intended to further that individual’s own good.

This, of course, immediately raises the question as to what constitutes the individual’s own good. Here, it is useful to distinguish between the *ends* that individuals may have in pursuing their own good and the *means* by which they achieve those ends. Individuals’ ends are anything that they want to achieve by their actions: these are the elements that contribute to their overall well-being. These could be something long term, such as promoting happiness in old age, or they could be more immediate, such as stopping smoking. The means are the activities that people undertake to promote their ends: their savings, their repeated resolutions to give up cigarettes.

Now, while there is little evidence that people make mistakes in deciding upon their own ends, there is a growing volume of evidence from behavioural economics and psychology that even well-informed individuals do make mistakes, or at least undertake misjudgements, over the means for achieving those ends. Collectively, we term these misjudgements ‘reasoning failure’, and we argue that they arise from four basic sources: limited technical ability, limited imagination or experience, limited willpower and limited objectivity. In cases where such reasoning failure occurs and where the misjudgements have considerable consequences and can be readily identified, replacing the individual’s judgements concerning means by those of the state can be supported on the grounds of increasing well-being.

However, even if we can demonstrate that a particular means-related paternalistic intervention will increase individuals’ well-being, this may not be sufficient to justify that intervention. We argue that there is another important consideration that has to be taken into account: the impact on autonomy. Can any kind of paternalistic intervention be exempted from the accusation that it impacts on an individual’s autonomy? The answer is no; for, as long as autonomy has as one of its essential elements the individual’s *belief* that he or she is governing their own life, the object of the intervention will almost always *perceive* their autonomy to be offended. All forms of paternalism actually involve trading off improvements in well-being against compromises of perceived and actual autonomy, and, unless one believes that autonomy is a lexicographic right such that no such compromise is ever acceptable, no matter how small is the infringement of autonomy or how large is the gain in well-being, such trade-offs are inevitable in the realm of practical policy. The aim of policy, therefore, should be to minimize so far as possible the loss in autonomy, both actual and perceived, while maximizing the gain in well-being.

What are the practical implications of these arguments? Is it possible to identify the areas in practice where some form of paternalistic intervention is most likely to be justifiable? In fact, it is possible to identify categories of consumption or activity that are most likely to be subject to reasoning failures for a large proportion of people because they manifest reasoning failure in most or all of its forms. These are:

- Those that result in a harmful effect that only manifests itself a long time in the future (unhealthy eating, smoking, failing to provide for a pension);
- Those that have a very small chance of an immediate catastrophic outcome (driving without a seat belt or riding a motorcycle without a helmet).

These types of activity typically suffer from the type of reasoning failure that we have labelled as ‘limited imagination’: one where it is hard to visualize one’s future circumstances long into the future or when the event is very rare. But

they also are associated with another type of reasoning failure: limited will-power – temptation is harder to resist if the benefits from such resistance will not accrue for many years or if the consequences of a failure to resist seem vanishingly small. And they are also linked with a third kind: limited objectivity. Anything risky can be subject to an emotional belief that we are the kind of person that will beat the odds. Finally, they may also contain elements of the fourth kind: limited technical ability. A judgement on whether to consume or not, or engage in an activity or not, will involve an accurate assessment of probability, either of an event a long time in the future or an immediate event with a low probability – something human beings seem quite ill-equipped to do.

So, if it is accepted that the state is justified in certain circumstances to engage in a paternalist intervention to promote the well-being of its citizens, then how should it do so? There are four types of paternalistic intervention: legal restrictions, positive financial incentives, negative financial incentives and changes in the choice architecture (libertarian paternalistic or ‘nudge’ policies). Two criteria may be used for assessing them: the impact on well-being (the larger the better) and that on autonomy (the smaller the better). For instance, some forms of compulsion, such as banning all smoking, may score highly on well-being, but, since they involve direct coercion, would have a low score on any scale relating to autonomy. Each of the other three are less damaging to autonomy, with libertarian paternalism perhaps the least harmful.

It will not surprise the reader that these arguments have been subject to challenge. Notable among these are those emanating from a long review article of Le Grand and New by Pennington (2016). There, he makes a number of general points that are part of a broader critique that he is making of choice and libertarian paternalism. He then raises two specific issues of direct relevance to our arguments: ones that he describes as Coasean and Hayekian after the authors that he cites in their support. I consider each of these points in turn.

### Paternalism and choice

Pennington (2016) relates a general critique of libertarian paternalism to our book because, he argues, it falls “firmly within the libertarian paternalism canon. Le Grand and New oppose government measures that limit personal choice.” He goes on to note that, both in my writings and in my work as a government policy adviser, I have been a strong advocate of individuals having choices, especially in public services such as healthcare and education (Le Grand, 2007).

The first point to make here is that, for reasons to be explained, in Le Grand and New (2015), we do indeed generally favour choice-preserving measures,

but we do not necessarily oppose measures that restrict choice. Rather, we propose a framework for systematically assessing the pros and cons of different paternalistic policies, including those that favour and those that limit choice. As should be apparent from the preceding section, we argue that a paternalistic policy can be justified when reasoning failure leads to individuals having lower levels of welfare or well-being than they would have if the policy were implemented and if these differences are of sufficient magnitude to outweigh any reduction in autonomy associated with the policy. If a hard paternalist policy that restricts choice has welfare benefits that outweigh the welfare costs and any associated infringement of autonomy, then we explicitly say that such a policy is justified.

We do believe that paternalistic policies involving choice-preserving interventions, such as taxes, subsidies, vouchers and the nudging and framing devices to which the libertarian paternalists have drawn attention, can often generate increases in well-being that are at least as great as otherwise equivalent choice-restriction policies, while infringing individual autonomy to a lesser extent. They would therefore be preferable from our perspective. Moreover, user choice in areas such as education and healthcare generally offer better incentives to providers to improve quality than command and control systems that incorporate more coercive paternalistic policies (Le Grand, 2007). However, we do not rule out the possibility that this may not be true in every case, and that, on occasion, choice-restrictive policies can be justified within our framework and, as such, may be preferable to the alternatives. Ultimately, which type of paternalistic policy is preferable is an empirical question that cannot be settled *a priori*, but has to be examined on a case-by-case basis.

### The Coasean argument

We now consider the two arguments that Pennington makes of particular relevance to our central arguments. The first of these Pennington derives from the seminal article on externalities by Ronald Coase (1960). Coase pointed out that, when externalities exist such that the actions of one individual or firm confer benefits or impose costs on others who are in no way involved in the activity concerned, there is no *a priori* reason to favour the interests of either the creator of the externality or of its recipient above the other. The size and distribution of the costs and benefits of this situation will depend on the allocation of the relevant property rights, and the answer to the question as to whether (and how much) government should intervene to 'correct' the externality should be neutral with respect to whether the individuals or firms concerned were recipients or creators.

How does this relate to our book? Pennington rightly notes that we are strongly opposed to intervening in people's ends, rather than in the means to those ends. However, he suggests that, during the course of our argument, we abandon our neutral stance over ends. He makes the analogy with a polluting factory and its neighbouring area that is affected by its pollution – a classic case of economic externalities. In that situation, he argues that an ends-neutral stance would be one that accepts a 'first possession' principle – whoever was there first having the right to pollute or to be free from pollution – rather than, for instance, systematically favouring 'amenity' interests over those of the producer.

With respect to an individual's decision that affects only him or herself but that has implications for both the present and the future, there are in a sense two people involved: the individual's current self and their future self. The current self makes the decision and in doing so benefits themselves; but in doing so they confer an 'external' benefit or cost on their future self. So, by picking up a cigarette today, they enjoy the narcotic effect of the tobacco in the present, but increase their future self's health risk. Pennington argues that in such a situation we would consider government intervention justified that prevented or in some way hindered them from picking up the cigarette. In doing so, "along with other behavioural economists," we would be automatically privileging the future self over the present, and therefore we are not being ends-neutral. True ends-neutrality would be something like the first possession principle. Applied to this case, this would presumably imply the giving of priority to the present (as the first in time) over the future and therefore not intervening to hinder or prevent the individual concerned from smoking.

There are a number of points here. First, Pennington's interpretation of ends-neutrality (or lack of ends-neutrality) is not as we would understand it. He argues that we are not ends-neutral because we apparently favour one individual's ends over another's: the future self's ends over the present self's ends. A clearer way of describing such favouritism, if we do display it (and, as we argue later, we do not), is as a lack of 'individual-neutrality', not ends-neutrality.

Pennington conflates ends- and individual-neutrality because he believes that the two individuals concerned – the 'present' and 'future' selves – have different ends. But we would argue that both the present self and the future self have the same end. This could be expressed formally as something like the maximization of lifetime utility (or, to take account of uncertainty, of the expected value of lifetime utility). Less formally, we take it that both 'selves', if it were possible to consult them at the same time, would like both a happy present and a happy future. The question is whether the means of achieving those ends is judged sufficiently well.

So we would recast Pennington's criticism as our not being individual-neutral rather than not being ends-neutral. But are we in fact not being individual-neutral? Pennington would argue that we are biased because we *always* assume that people's long-term welfare should take precedence over their shorter-term pleasures and consumption. But why should the 'longer-term self' be automatically preferred over the 'current self'? Indeed, Pennington suggests that in some cases the current self may actually need protecting from the future self, particularly when the individual exercises 'too much' self-control rather than too little. He cites examples of workaholics and anorexics, both cases where the present self arguably damages their current well-being in order to protect the interests of their future selves: in these cases, surely, the correct response should be to curtail the interests of the future self?

In fact, Pennington's invocation of the 'first possession principle' in the pollution example when applied to the present/future decision seems to imply, if anything, not future individual-bias or even individual-neutrality, but something like present-individual bias. Putting that aside, it should be noted that in general the analogy does not match our own case for paternalistic intervention. First, an individual who makes decisions that adversely affect a future self does not operate like the pollution example – their future self is *automatically* put at risk by the current self's decisions and will very likely be adversely affected by them. One's future self cannot, as it were, 'stay away' from the polluter. A more accurate analogy would be of a polluter whose pollution would only affect the surrounding community (who are already living near the farm) many years after the polluting activity. But in this case a local community could agitate and protest at what is coming their way. However, our future self cannot speak up for their interests; instead, the state has to act as their advocate.

More generally, we would argue that it is wrong to assume that 'too much' self-control is any less likely to be a reasoning failure and thus a candidate for paternalism than 'too little' self-control. Excessive self-control involves current decisions damaging the interests of future selves. Workaholics may cause their personal relationships and family life to suffer in the long term; anorexics are very likely to suffer future ill-health and possibly death. The issue is not the degree of control, but whether the individual can properly balance current and future welfare. The anorexic/workaholic example is peculiar only in that it involves decisions to deny oneself current pleasures in the possibly mistaken belief that this will increase some kind of future welfare, when in fact it does the reverse.

But there is a wider point here. Why, in the case of a smoker, for example, where the current pleasures are perhaps more clear cut than with an anorexic,

should we favour protecting the long-run smoker's 'self' from becoming ill, rather than the current 'self' deriving pleasure from their cigarette?

Pennington argues that, with the use of examples such as these, we seem to be *systematically* favouring what we deem to be long-run over short-run interests. However, this is not quite right. The cases we choose for justifiable paternalism are only those where the reasoning failures seem most likely to pertain and where the harm done (or benefits forgone) is likely to be most severe. This *tends* to occur where decisions affect outcomes a long time in the future, but not inevitably. Our concern is not to bias systematically in favour of long-term selves, but rather to bias in favour of decisions that favour overall lifetime welfare.

However, it is correct to argue that we do believe – and indeed cite empirical evidence in support of this belief – that many individuals' reasoning abilities fail them in some of their short-term decision-making. As a result of those mistakes, their lifetime well-being will be lower than if they had not made the mistakes or if those mistakes had been corrected by the appropriate government intervention. Take the reasoning failure of limited imagination. This occurs because it is difficult to conceive of, for example, having lung cancer, and thus we tend erroneously to downplay its seriousness. As it happens, cancer tends to strike many years after the initial behaviour, but the reasoning failure would also pertain if there were the possibility of contracting lung cancer tomorrow. We cite a famous study (Le Grand & New, 2015, p. 91) in which people assessed a coffee mug as more valuable when they actually possessed it rather than when they were only able to see it or have it described to them. The 'immediacy' of possessing the mug changed people's assessment of its merits. So, although the adverse consequences of smoking for the individual are indeed normally felt some time in the future, the reasoning failure is more than just a matter of being short-sighted in the assessment of future costs. It is specifically concerned with their limited ability to imagine themselves in a particular situation.<sup>1</sup>

Another type of reasoning failure – weakness of the will – is more closely associated with problems of current and future selves. That weakness of the will presents difficulty for lifetime well-being has a long pedigree, going back at least as far as Aristotle, and its very longevity suggests that people understand that their actions do not always accord with their will or with their future interests. I tell myself that it is in my interests to put £20 away

1 The 'limited objectivity' reasoning failure is similar in that it often but not exclusively concerns long-term decisions. Far more people think they are 'better than the average' than statistics allow, whether it is avoiding a car crash tomorrow or lung cancer 30 years from now.



towards my pension every week, but as I walk past the pub on a Friday night, I can't seem to help going in and spending the lot on beer. The temptation is too great and, as a consequence, my long-term well-being suffers. In this case, we do argue that the future self should be protected to some degree, but only because the potential harm done to the individual is grave. We do not support paternalistic interventions in many other types of consumption decision with long-term payoffs, such as undertaking a personal development course, non-pension saving and investing, doing regular maintenance work on one's car or even planting trees rather than marigolds. In such cases, the future benefit is either unlikely to be an end that is universally supported or is simply trivial.

Where to draw the line between those 'important' long-term benefits and costs and others must ultimately depend on empirical evidence and public support. One study, for example, found that potential smokers were happier in areas where there were higher duties on cigarettes; public surveys regularly show that people wished they smoked less or saved more; and policies that restrict smoking or encourage saving for retirement are generally very popular *once implemented* (Le Grand & New, 2015, pp. 96–97). Of course, one cannot 'prove' that the judgement of the voter or survey respondent is more 'real' than that of the person making a weak-willed decision to enjoy pleasures now rather than in the future. But we would argue that in the context of something as important as an impoverished old age or painful and premature death, a dispassionate political decision should be favoured over those taken at a point in time when temptation and other distractions are likely to dominate. Perhaps it would be fairer to say that we have a bias in favour of the 'disassociated' individual removed from the point of their decision-making.

To summarize, Pennington's interpretation of ends-neutrality (or lack of it) is not as we would understand it. We would argue that the current and future selves have ultimately the same end: what we might term the promotion of lifetime well-being. So, *ceteris paribus*, a paternalistic intervention with respect to individual behaviour is justified if the result of that intervention is a net gain in lifetime well-being for the individuals (subject to the caveats that the gain is not trivial and there is a minimal loss of autonomy). And this does not necessarily privilege the future self over the present. If it can be readily demonstrated that the well-being reduction from, say, not smoking in regular visits to the pub outweighs the loss in well-being due to the reduction in life expectancy that results from such smoking, then banning smoking in pubs would not be justified according to our framework.

## The Hayekian argument

Pennington's second line of criticism parallels the well-established arguments of Friedrich Hayek against the state as an allocator of resources. People are generally better off trying to manage their own suboptimal decision-making rather than getting the state involved. This is partly because the state will inevitably do worse than disaggregated individual decision-making because of the complexity and context specificity of those decisions. Moreover, individuals are usually aware of their own reasoning failures and can take steps to correct them themselves, such as checking into a health farm to help overcome overeating problems.

Furthermore, even if the state could do better than the individual, it almost certainly will not do so in practice, for most state actors are driven by their own self-interests rather than the interests of those they govern, and even when they are not, they will be subject to reasoning failures of their own that will adversely affect their decision-making.

Again, there are a number of different points here. First, the reasoning failures to which the behavioural economics literature has drawn attention are generally so profound and so widespread that they are relevant to a large number of people – in other words, they are not, or are very rarely, context specific. Does any non-trivial group of people regret saving too much for retirement? Or regret having finally given up smoking?

Pennington argues that paternalistic policies may have unintended consequences, even if they are successful in their primary objective, because they ignore *other* biases that affect people's decision-making, thus potentially making it more difficult to achieve their ends as a whole. It is difficult to assess such a claim, as Pennington does not offer specific examples. He also argues that people could become less able to develop their own effective behavioural strategies – presumably because they now have this done for them by the state. In both cases, these are empirical questions. Certainly, careful monitoring of paternalistic policies would be necessary to ensure that they are not having unintended consequences, but it seems unlikely that providing incentives or nudges to achieve what individuals in other contexts profess they want to achieve will significantly divert them from such ends.

Second, only the government has the authority to enforce the required behaviour at the crucial point (what is to stop the individual from checking out of the health farm or ignoring the advice of a trusted friend, just when it matters most?). It is precisely because their decision-making is so poor at the point when it matters that the authority of the state is required. We should note that there is no reason to dismiss the effectiveness of 'self-help' strategies in many instances. Not everyone suffers from reasoning failures to the same

degree, and many people, whether through habit, insight or good fortune, are able to avoid the consequences of limited imagination or weak willpower. However, the rest of us will benefit from the restructuring of the choice context in which we live: it is precisely in order to help those who are more ill-disciplined that state intervention is necessary.

But Pennington's invocation of Hayek in fact has a broader focus than simply complexity or context specificity: it is that of government failure. State actors may be irrational themselves, or they may be self-interested; and, in order to serve their irrational and/or self-serving ends, they may inappropriately manipulate citizens by using their reasoning failures as a justification. Yet more fundamentally, democratic accountability itself is ineffective as a buttress against inappropriate paternalism because people cannot use democracy properly: reasoning failure itself implies limits to the effectiveness of majoritarian decision-making.

We would agree with the last point about the dangers of simple majoritarian decision-making. Indeed, we devote a whole chapter in *Le Grand and New (2015)* to the politics of paternalism that draws attention to the need to have some check on democratic government engaging in paternalism, constitutional or otherwise. We suggest various measures for meeting that need, including retrospective reviews of paternalistic policies by means of sunset clauses and post-hoc debates in legislative assemblies or referenda. In addition to their 'checking' function, such retrospective instruments have the advantage that they enable both those who benefit from the policy and those who do not to experience the policy in question before deciding upon its continuance.

In fact, Pennington comes close to rejecting the role of the democratic state in any context: democracy is simply insufficient to counteract the forces of inertia, maladministration and corruption that, in Pennington's view at least, are inherent in state activity. But it is not clear why these problems should apply with particular force to paternalistic policies. In fact, there is a whole history of apparently successful democratic state interventions, including paternalistic ones. To take just one example, consider the banning of smoking in public places in the UK. As mentioned in the book, one of us (Le Grand) was working in government at the time that this was being considered, and he can vouch for the fact that the forces of irrationality and self-interest worked overtime to try to prevent the measure ever becoming law. Powerful lobby groups in both the hospitality and tobacco industries worked with their friends in government to try to defeat the measure. And yet, despite the fact that the groups in favour of the measure (primarily those from the public health profession) had apparently far fewer resources and far fewer friends, the measure was eventually adopted – and has been retained despite several subsequent changes of minister and government.

Of course, this is just one example. But it is important to recognize that, on occasion, democratic government can work towards the public good and that the kind of blanket criticism made by Pennington and other critics of the state has its holes, just as does the blanket advocacy of the pro-democrats. Overall, democracy with checks and balances has its obvious flaws as a means of exercising control over the state, but most analysts would agree that it is better than the alternatives. Or do we throw the policy baby out with the imperfect accountability bathwater?

## Acknowledgements

The basis of this article was the 2019 Behavioural Public Policy annual lecture at the London School of Economics. I am grateful to participants at that lecture, my co-author, Bill New, and the editor of this journal, Adam Oliver, for many helpful comments.

## References

- Camerer, C., S. Issacharoff, G. Loewenstein, T. O'Donoghue and M. Rabin (2003), 'Regulation for Conservatives: Behavioral Economics and the Case of 'Asymmetric Paternalism'', *University of Pennsylvania Law Review*, 1151: 1211–1254.
- Coase, R. (1960), 'The Problem of Social Cost', *Journal of Law and Economics*, 3(1): 1–44.
- Conly, S. (2013), *Against Autonomy: Justifying Coercive Paternalism*. Cambridge: Cambridge University Press.
- Le Grand, J. (2007), *The Other Invisible Hand: Delivering Public Services Through Choice and Competition*. Princeton and Oxford: Princeton University Press.
- Le Grand, J. (2018), 'Future Imperfect: behavioural economics and government paternalism', *Review of Behavioral Economics*, 5(3–4): 281–299.
- Le Grand, J. and B. New (2015), *Government Paternalism: Nanny State or Helpful Friend?* Princeton and Oxford: Princeton University Press.
- Mill, J. S. (1859/1972), *On Liberty*. London: Dent.
- Pennington, M. (2016), 'Paternalism, Behavioural economics, Irrationality and State Failure', *European Journal of Political Theory*, 18(4): 565–577.
- Sugden, R. (2018), *The Community of Advantage: A Behavioral Economist's Defence of the Market*. Oxford: Oxford University Press.
- Thaler, R. and C. Sunstein (2003), 'Libertarian Paternalism', *American Economic Review*, 93: 175–179.
- Thaler, R. and C. Sunstein (2008), *Nudge: Improving Decisions about Health, Wealth and Happiness*. New Haven: Yale University Press.