

Future Imperfect: Behavioral Economics and Government Paternalism

Julian Le Grand

Marshall Institute, London School of Economics, UK; J.legrand@lse.ac.uk

ABSTRACT

Economists and others have used the results from behavioral economics to justify paternalistic government policies, aimed at changing an individual's behavior in the present so as to improve that individual's well-being in the future. Examples include the automatic enrollment in pension schemes and anti-smoking measures, such as banning smoking in public places or proposals for a smoking license. But these - and the economic analyses underlying them - have been challenged on the grounds that they arbitrarily privilege one set of preferences over another. The privileged preferences include those of an 'inner rational agent' and those of the future over the present. This paper addresses this criticism and puts forward two new conceptions of - and justifications for - these kinds of policy.

Economists, psychologists and others have used the growing number of empirical results from behavioral economics to justify paternalistic government policies (Camerer *et al.*, 2003; Thaler and Sunstein, 2008; Conly, 2013; Le Grand and New, 2015). The principal argument is that individuals make errors of judgment in the means they use to achieve their ends, and that governments can use policy interventions to correct those errors, and thus help those individuals better to achieve those ends. These interventions take a variety of forms and operate in a variety of areas. So they have been used to justify taxing, licensing or outright banning potentially unhealthy activities, such as

This paper owes a great deal to discussions with Robert Sugden and Dan Hausman. I also received helpful comments from Paul Niehaus and participants of the Behavioral Economics and the New Paternalism Conference in New York, April 2018.

smoking and drinking alcohol; government subsidies to cultural activities, such as operatic performances, museums or public broadcasting; legal restrictions on addictive activities, such as hard drug taking or gambling; and 'nudges' to encourage savings, such as switching the default for pension schemes from opt-in to opt-out.

However, these arguments – and the policy proposals associated with them – have been challenged as unjustifiably and arbitrarily privileging a particular set of preferences and the choices associated with those preferences. The defenders of these arguments and policies often describe those preferences as the individual's 'true' or the 'real' preferences, or, more precisely, as the ones that the individual would endorse if he or she 'had complete information, unlimited cognitive abilities and no lack of will-power' (Thaler and Sunstein, 2008). Others, more sceptical of their use as a justification for government paternalism, describe these preferences as those of an 'inner rational agent' (Sugden, 2018) or as those of the 'long-run self' (Pennington, forthcoming).¹

This paper addresses these challenges. I argue that there are at least two kinds of defense of government paternalistic interventions that can be mounted, one utilitarian in nature, the other on based on contractarianism, neither of which rely upon arbitrarily privileging one set of preferences, whether those of an inner rational agent or of the long-run or future self.

1 Types of Paternalism

It will help the discussion to differentiate between two types of paternalism. The basic definition of paternalism used here is that developed in Le Grand and New (2015). This is that a government intervention is paternalistic with respect to an individual if it is intended (a) to address a failure of judgment by that individual and (b) to further the individual's own good.²

This definition does not specify who decides on what constitutes the individual's own good: the individual, the government, or someone else entirely. Nor does the definition tell us which paternalistic interventions might be justified and which not. Given that most individuals make many judgments that both they themselves and others might judge as failures, we do need a further criterion to decide which failures the government should address (if any) and which it should not.

In order to find this criterion it is useful to distinguish between two kinds of paternalism: means-related and ends-related. The central idea here is that individuals have 'aims, ends, and values that at any point in time define or give content to [the individual's life plan] and which the person appeals to in

¹An older debate over unreliable preference revelations generated another term that might be applicable to these privileged preferences: 'laundered preferences' (Goodin, 1986).

²For further discussion of definitions of paternalism see Hausman (2018a).

choosing between alternative courses of action' and that he or she makes choices 'between those alternative courses of action themselves': that is, between the means of trying to achieve those ends (Brock, 1988, p. 561).

Ends-related paternalism is where a government agency considers that some individuals have failed appropriately to judge their own ends, and so replaces those individuals' ends with its own conception of what those ends should be its own conception of the good - and tries to alter the relevant individuals' behaviors so as to achieve them. Means-related paternalism is where the agency accepts the individuals' ends - that is, accepts their conception of the good life - but, observing that the individuals making mistakes in trying to achieve those ends, intervenes to correct those mistakes or their outcomes. Le Grand and New argue that means-related paternalism can be justified by reference to the empirical evidence produced by behavioral economists and psychologists that people do indeed make mistakes in trying to achieve their ends. But ends-related paternalism cannot be justified in this way, since ends are a matter of values and cannot be either refuted or supported by empirical evidence in this way.³

How do the challenges to behavioral economics listed above relate to this distinction? The basic argument concerns the nature of an individual's failure of judgement concerning the means he or she uses to achieve his or her ends especially if those ends concern the future, whereas the decision concerning means is in the present. But failure as compared to what? What choice of means would constitute a successful judgement?

As noted in the introduction, many of the behavioral economists who are also government paternalists refer to 'true' or 'real' preferences. In perhaps the most substantial critique of this line of argument, Robert Sugden argues that the paternalists are fundamentally defining successful decisions as those that would have been made by an 'inner rational agent'; that is, an agent with the same preferences as the individual concerned but who, in making the decisions concerned, conforms to the neo-classical axioms of rationality. And, he goes on to argue, this presupposes 'a kind of rational human agency for which there is no psychological foundation' (2018, Ch. 4, p. 82).

However, it is possible to mount defenses of means-related paternalistic government interventions that do not rely upon a comparison with the judgements about means made by an inner rational agent or the equivalents proposed by other critics. One of these has utilitarian foundations and the other rests on contractarian ones.

³For the view that certain kinds of ends-paternalism can in fact also be justified by reference to misjudgements over ends, see Acland (2018).

2 Government Paternalism: A Utilitarian Defense

Suppose that the principal aim or end for most individuals is to maximize their utility over their lifetime.⁴ Now a government agency observes individuals undertaking a particular activity in the present that the agency judges will raise their utility - again in the present. The judgement is an empirical one; the agency has observed many people engage in this activity and therefore infers that such people derive benefit therefrom. However, in the case of this particular activity, the agency estimates that it is also likely to have consequences at some time in the future for the individuals concerned, and these will result in a substantial reduction in their utility. Again, this is an empirical judgment based on its observations of individuals experiencing those consequences.

Although obviously it would be well beyond the capacity of the agency to come up with exact calculations of the relevant increases and reductions of utility, in many cases it is likely to be able to assess whether one is larger than the other. In particular it may judge that the increase in utility from the activity in the present is substantially less than the likely fall in utility resulting from the consequences of the activity in the future. In such cases, it may conclude that the individuals concerned will experience a 'net' reduction in lifetime utility as compared to that which the individuals would experience if they had not undertaken the activity. So the government agency judges that the individuals have not chosen the right means to achieve their ends; and it

an outright ban on smoking for Jane (or anyone else) and towards in the direction of a sin tax, or a smoking permit or license (Le Grand and New, 2015, pp. 157—159).

But these kinds of concern about autonomy do not directly challenge the utilitarian justification for government intervention in Jane's behavior. What does challenge that justification is the issue of Jane's own decision-making. If the agency can make the lifetime utility assessment, then presumably Jane can too – and presumably she has made that calculation in deciding to purchase the cigarettes. Why then should the agency overrule her judgement? Of course, no-one is arguing that Jane – or anyone else - purchasing cigarettes actually makes formal calculations of the expected value of their lifetime utility. But, given that there cannot be anyone on the planet who is unaware that smoking has some degree of health risk, it has to be reasonable to assume that she knows there is a risk and has taken that knowledge into account in some way in the judgment that led to her decision. So on what basis can the government agency decide that her judgment over means is faulty and needs to be replaced by its own judgment?

There are three reasons why the agency may so decide. The first is that it believes that at '9gmj(wn)belrTJ/F19t8ide.132(expSu)-1(e)-mF15 9.9626 Tf 0.98 0 0 1 !

A second reason why the agency might that Jane might make the wrong decision is that, in the agency's judgment, she is using an incorrect discount rate in her calculations of expected value. Put less formally, when she assesses the relative contributions that present and future activities make to her lifetime utility, she 'weights' the latter as less than the former – incorrectly in the agency's view. Now, if the agency did intervene for this reason, it would undoubtedly be paternalistic, but would it be justifiably so? The principal reason why Jane might discount both the costs and benefits of future activities is likely to be because of her assessment of the uncertainties that are involved in any assessment of the future. Then, as with the risk discussion above, if the government agency has a different assessment of these uncertainties, then its role is the (non-paternalistic) one of providing Jane with its own assessment and then to allowing her to make her own decision. Otherwise, there seems to be little reason for the government to judge Jane's discount rate as 'wrong'; her uncertainty-free weighting of present and future presumably depend on her preferences, and the agency has no reason to judge those preferences as wrong.⁸

So, if the government agency were to hinder or prevent Jane's activities because of poor information on risks or because she uses 'incorrect' discount rates, then it would seem to be unjustifiably paternalistic. But there is a third reason as to why the agency might decide that Jane's judgement is likely to be faulty, even if she were in possession of the correct information concerning the relevant risks and uncertainties. This is because she is likely to suffer from a type of reasoning imperfection or 'failure'.

Now four types of reasoning failure are identified in Le Grand and New:

This kind of argument for paternalistic intervention does not seem to require *arbitrarily* privileging one set of the preferences over another. In the example, there is indeed some privileging going on; the government intervention is clearly giving the preferences of those of 65 year-old Jane priority over those of 18 year-old Jane. However, note that the privileging of Jane in the future in this way is not automatic. It is possible that the agency, in assessing the expected value of the impact of 18 year-old Jane's decision on the 65 year-old Jane's well-being, might conclude that it was less than the impact on the well-being of 18 year-old Jane of her being prevented from smoking; in which

in a contractarian world, FJ and PJ would discuss this matter and come to some arrangement of mutual benefit. But FJ is not around to argue with PJ. So someone else has to represent her interests in the contractual discussion - and the obvious 'person' to do that is the government.

The outcome of any such contractarian 'negotiation' might be rather different from the standard government ways of correcting negative externalities, such as taxing or regulating the activities concerned. It is unlikely that, if it were possible to have a discussion between PJ and FJ, that PJ would simply agree to have her behavior taxed or regulated without receiving any kind of compensation from FJ. Rather, FJ (or the government agency representing FJ) would have to offer PJ some kind of incentive to encourage her to change her behavior. This could take the form of a generous loan to attend a health farm from addiction problems, a loan to be repaid by FJ at a later date. Or perhaps it could involve simply paying PJ not to smoke in the present, financing this by government borrowing, ultimately to be repaid by a tax on FJ.

Interestingly, although in such scenarios there is clearly government intervention aimed at changing PJ's behavior in a similar direction to those discussed earlier (and labeled paternalistic), this would not be a paternalistic act as we have defined it. For the government agency would not be intending simply to replace PJ's judgements with FJ's or even with its own; nor would it be trying to benefit PJ. Rather it would be trying to come to some arrangement that it thinks FJ would support, if she were present. In doing so the agency is simply undertaking a process of correcting market failure.

Is it possible to devise contractarian arguments that have a similar outcome but do not involve the multiple selves model? Robert Sugden would argue that it is. To see his argument, let us start with a non-health example, adapted slightly from one in his book (2018, p. 133). An individual (which could be Jane's father) observes that Jane has inherited something from a deceased uncle at a relatively young age that she does not value highly: say, a crate of vintage wine. Jane would have much preferred the uncle's computer - or indeed the cash equivalent. The observer knows this but thinks that, in later years, her preferences will change and that she would then value the wine rather more. If the observer were a utilitarian paternalist, and, if he were in a position to do so (as he might be if he were her father), he might view himself as justified in preventing Jane from selling the wine, on the grounds that this was for her own good and her lifetime utility would thereby be increased. But suppose he is not a paternalist but actually a contractarian. So, instead of engaging in some sort of preventive action, he offers to buy the wine from Jane at a relatively low price. She accepts the offer. He then stores the wine for a few years, then offers the wine back to Jane at a somewhat higher price. Jane, whose preferences have indeed changed in accordance with the observer's prediction, accepts the price and buys back the wine. Both exchanges are completely voluntary; so everyone is satisfied with them. No judgements have

been displaced: no paternalistic interventions have been invoked; no autonomy infringed.

Problems arise, however, when goods are not transferable from one individual to another, or from one period to another, in this way. Take health. If health were transferable, and Jane had 'too much' health in the present, she could sell some of it to an entrepreneur, who would then sell it back to her in the future when her 'stock' of health was lower. That is obviously not possible. However, it could be argued, in a fashion similar to that adopted by Sugden in his paper for this volume, that the government can 'mimic' this transaction in a way similar to that already discussed: offering Jane a subsidy to avoid engaging in unhealthy activities in the present and then taxing her in the future. It might be then argued that this is not paternalistic because it is simply imitating a market transaction that cannot take place for technical reasons.

I have to say that, unlike the multiple selves model, this still seems to me to be an example of government paternalism. It certainly fits our definition. In offering the subsidy, the government is substituting its judgment for Jane's in the present in order to further Jane's own good in the future. It seems that, if contractarians wish to endorse government actions such as the subsidy/tax ones in the example and avoid the label of paternalism, their only route to do so is via the multiple selves model; and, if they are uncomfortable with that, then they cannot support any government action of the kind suggested.

4 Conclusion

Several critics have argued that government paternalist interventions, designed to correct the errors of judgment that individuals make when choosing means in the present to achieve their ends in the future, arbitrarily privilege one set of preferences over another. This paper has accepted that such interventions can indeed privilege a set of preferences (although not always the same set), but that this is not arbitrary and can be justified by reference either to the desire of governments to assist the individual to maximise their lifetime well-being or to the government's desire to represent the preferences of 'future individuals' in any contractarian discussion of policy. The future will never be perfect; but, one way or another, suitable government intervention can make it better.

References

- Acland, D. 2018. "The case for ends-paternalism: extending Le Grand and New's framework for justifications of government paternalism". *Review of Behavioral Economics*. 5(1): 1–22.

- Brock, D. 1988. Paternalism and autonomy: a review of Feinberg and VanDe-Veer . *Ethics*. 98(3): 550 565.
- Camerer, C., S. Isacharo , G. Lowenstein, T. O'Donoghue, and M. Rabin. 2003. Regulation for conservatives: behavioral economics and the case for 'asymmetric paternalism' . *University of Pennsylvania Law Review* 151(3): 1211 1254.
- Conly, S. 2013. *Against Autonomy: Justifying Coercive Paternalism*. Cambridge: Cambridge University Press.
- Frankfurt, H. 1971. Freedom of the will and the concept of a person .*Journal of Philosophy*. 68: 5 20.
- Goodin, R. 1986. Laundering Preferences . In *Foundations of Social Choice Theory*. Ed. by J. Elster and A. Hylland. Cambridge: Cambridge University Press. 75 101.
- Hausman, D. M. 2018a. *Behavioural Economics and Paternalism* .