

The Problem With Nudges: Paternalism,
Autonomy, and Transparency

O Problema com os Nudges: Paternalismo,
Autonomia e Transparência

FRANCESCO FERRARO

VOL. 8 N.º 2 SETEMBRO 2021

WWW.E-PUBLICA.PT



COM O APOIO DE:

FCT Fundação
para a Ciência
e a Tecnologia

ISSN 2183-184x

THE PROBLEM WITH NUDGES: PATERNALISM, AUTONOMY, AND TRANSPARENCY

O PROBLEMA COM OS NUDGES: PATERNALISMO, AUTONOMIA E TRANSPARÊNCIA

FRANCESCO FERRARO

University of Milan

Dipartimento di Scienze Giuridiche "Cesare Beccaria", Sezione di Filosofia e Sociologia del Diritto

Via Festa del Perdono 7, Milano. Itália

Francesco.Ferraro@unimi.it

Abstract: Nudging as a tool of public policies has often been criticized on account of two assumptions: 1) it is more dangerous than traditional command & control regulation for human dignity and autonomy; 2) when it is justified on paternalistic grounds, it is even more problematic than in its non-paternalistic applications. This paper seeks to challenge both assumptions, by showing that the main problem with nudging is one of transparency, but this can be obtained without impairing its effectiveness; moreover, paternalistic nudging is means paternalism, not ends paternalism, which means that it is not a form of perfectionism superimposed on the people's individual ends.

Keywords: Nudging; Autonomy; Means Paternalism; Libertarian Paternalism; Perfectionism.

Summary: 1. Introduction. 2. Paternalism as a complex notion. 3. Nudges as a threat to liberty and autonomy. 4. Is paternalistic nudging more problematic?

Resumo: O *nudging* como um instrumento de políticas públicas tem sido frequentemente criticado com base em dois pressupostos: 1) é mais perigoso para a dignidade humana e autonomia do que a tradicional regulação de “comando e controlo”; 2) é ainda mais problemático quando assenta em razões paternalistas, quando comparado com as suas aplicações não paternalistas. Este artigo procura desafiar ambos os pressupostos, mostrando que o principal problema do *nudging* é o da transparência, sendo que a mesma pode ser garantida sem prejudicar eficácia do *nudging*; além disso, o *nudging* paternalista corresponde a um paternalismo *de meios* e não a um paternalismo *de fins*, o que significa que não é uma forma de perfeccionismo que se sobrepõe aos fins individuais de cada pessoa.

Palavras-chave: *Nudging*; Autonomia; Paternalismo *de meios*; Paternalismo Liberal; Perfeccionismo.

Sumário: 1. Introdução. 2. Paternalismo como uma noção complexa. 3. *Nudges* como uma ameaça à liberdade e à autonomia. 4. Será o *nudging* paternalista mais problemático?

1. Introduction

Nudges as a tool of public policy are usually looked upon with suspicion. In comparison with other kinds of policies, like the implementation of traditional “command & control” regulation, they are regarded as somehow more dangerous for the protection of basic rights and liberties. Why is it so? What is the problem with nudging by public powers? The main concerns regarding public nudging under constitutional democracies have been summarized by Anne van Aaken¹. In the first place, she claims that nudging is a potential threat to human dignity and autonomy, the protection of which is obviously prescribed by, and entrenched in, both national constitutions and international charters; the threat it poses is supposedly greater than command & control state regulation. In the second place, according to van Aaken *paternalistic* nudging is subject to a heavier justificatory burden and legal constraints than its non-paternalistic instances. Van Aaken’s concerns, of course, are shared by many others, often with reference to nudges’ alleged manipulative tendency².

In what follows, I will challenge both of the above assumptions: I will try to show that neither is nudging more of a threat to dignity and autonomy than traditional regulation, nor do its paternalistic applications carry a heavier *prima facie* duty of justification. However, I will first show how paternalism itself is a complex notion, difficult to capture with an all-comprehensive definition, and I will inquire into the possibility of considering its “libertarian” version – which is often associated with the use of nudging as a public policy tool, in order to defend its acceptability – as full-fledged paternalism.

2. Paternalism as a complex notion

Of course, not all nudges are paternalistic, although Richard Thaler and Cass Sunstein seem to consider them as the most typical instrument of “libertarian paternalism”.³ Many nudges do not pursue paternalistic ends, because they are not aimed at the good of the addressed individual(s); they are aimed at some other ends, like environmental preservation, availability of organs for transplants, etc. Nonetheless, nudges seem to be particularly apt for paternalistic uses, as in the classical examples of retirement plans, healthier eating, etc. To understand what it means for nudging to be paternalistic, of course, one must first answer the question concerning what we mean by “paternalism”. This can prove challenging, since a definition capable of capturing the different uses of the term in ordinary language, while tracing at the same time conceptual boundaries precise enough

1. A. VAN AAKEN, *Constitutional Limits to Paternalistic Nudging: A Proportionality Assessment*, in A. KEMMERER, C. MÖLLERS, M. STEINBEIS, G. WAGNER (eds.), Baden-Baden, Hart-Nomos, 2016, pp. 161-195.

2. On the relationship between nudges and manipulation and the related ethical concerns see T. M. WILKINSON, “Nudges and Manipulation”, *Political Studies*, 61, 2, 2013, pp. 341-355.

3. R. H. TALER and C. R. SUNSTEIN, *Nudge. Improving decisions about health, wealth and happiness*, Revised edition, London, Penguin Books, pp. 5-6.

to make it amenable to philosophical enquiry, is very difficult to attain. Gerald Dworkin has provided the following tentative definition.⁴

"I suggest the following conditions as an analysis of *X acts paternalistically towards Y by doing (omitting) Z*:

1. *Z* (or its omission) interferes with the liberty or autonomy of *Y*.
2. *X* does so without the consent of *Y*.
3. *X* does so only because *X* believes *Z* will improve the welfare of *Y* (where this includes preventing his welfare from diminishing), or in some way promote the interests, values, or good of *Y*."

As useful as this definition may be as a starting point for further elaboration, all the conditions it puts forward are problematic. The first condition, for instance, makes reference to the interference of the paternalistic agent *X* with the liberty or autonomy of another individual *Y*. There are two problems with this condition:

1. In the first place, interference could even be aimed at enhancing the freedom of *Y*: for instance, by presenting choices or opportunities which *Y* had not even considered before.⁵ *X* can be a parent who invites to her daughter's birthday party some kids she does not usually hang with, without consulting her, just to give her more possibilities of making new friends. Thence, paternalism is not necessarily autonomy- or liberty-reducing.
2. In the second place, actual interference by action or omission is not required. For instance, *X* can deny help requested by *Y*, on account of the fact that helping *Y* will not do him good, maybe because *Y* needs to overcome his laziness or insecurity and stop asking for help.⁶

The second one of Dworkin's conditions states that *X* acts, or omits to, without the consent of *Y*. Now, think of the case of the "happy paternalist", say, the Victorian father and husband whose paternalistic intervention is accepted by his wife and children, whose wish is actually that he behave paternalistically towards them. Here, the subjects of paternalistic intervention give their consent to it. Therefore, Shane Ryan has proposed to reformulate this condition as follows: *Z* acts *irrespective of* the wishes of *Y* or his consent. This reformulation makes the first of Dworkin's conditions redundant: actual interference is not required,

4. G. DWORIN, *Paternalism*, in E. N. ZALTA (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), <https://plato.stanford.edu/archives/fall2020/entries/paternalism>

5. S. SHIFFRIN, "Paternalism, Unconscionability Doctrine, and Accommodation", *Philosophy and Public Affairs*, 29, 3, 2000, pp. 205-250, at 214.

6. SHIFFRIN, *PPA*.

but the very fact that paternalist action takes place irrespective of the wishes and consent of those addressed makes it inherently threatening for autonomy and/or liberty⁷.

Even the third condition, the “condition of benevolence” (i.e. *X* acts because she thinks that she will somehow improve *Y*’s welfare or enhance his interests, values, or good in some sense), has been questioned (for instance, to make room for paternalistic intervention which does not target *Y*’s good, but someone else’s, as in the case of epistemic paternalism)⁸. Even if we retain this last condition, as I think we should, its interpretation is open to a number of possibilities. These include the *moralistic* version of paternalism or moral paternalism, which conceives of *Y*’s good in moral terms and, conversely, endorse the idea of a possible “moral harm” to *Y* that *X*’s paternalist intervention can be aimed to prevent. Moral paternalism is usually contrasted with *welfare paternalism*, in the sense that with moral paternalism *Y*’s welfare (the satisfaction of his preferences) is not taken into account: actually, even *Y*’s informed preferences and deepest wishes can be incompatible with his moral good.

Regarding this latter condition of benevolence, a further point should be made, regarding the problem of paternalism by public institutions, such as legislatures and governments. The assumption that all the members of such institutions, or even their majority, share the same purposes – in this case, paternalistic purposes concerning intervention in the interest of those acted upon – is obviously unwarranted or, better said, unrealistic. Therefore, how can we possibly characterize certain public policies as paternalistic? This is what Kalle Grill has called the “diversity problem”⁹. In legislation, speaking of paternalism is problematic also because laws often outlive their creators and their persistence in time can be due to very different reasons from those that prompted their issuing. Could we use this as a rebuttal to classic concerns about paternalistic policies? The answer is negative, because explaining away legislative and governmental paternalism in such a way is hardly a convincing response to those worried by paternalistic public action, since they are often reasonably driven by concerns about autonomy and liberty. Furthermore, Douglas Husak has put forward a solution to the diversity problem which consists into taking into account only the “best rationale” for public measures, i.e. their most convincing or plausible justification: if such justification is based upon the interests of those acted upon, then – if other conditions for speaking of paternalism are fulfilled, like lack of consideration for the consent of those addressed – the measures in question are paternalistic, regardless of what flesh-and-blood legislators actually had in mind¹⁰.

7. S. RYAN, “Paternalism: An Analysis”, *Utilitas*, 28, 2, 2016, pp. 123-135.

8. See, for instance, K. AHLSTROM-VII, *Epistemic Paternalism: A Defence*, London, Palgrave Macmillan, 2013, pp. 39-64.

9. K. GRILL, *Paternalism By and Towards Groups*, in K. GRILL and J. HANNA (eds.), *The Routledge Handbook of the Philosophy of Paternalism*, London-New York, Routledge, pp. 46-58.

10. D. N. HUSAK, *Legal Paternalism*, in H. LAFOLLETTE (ed.), *The Oxford Handbook of Practical Ethics*, Oxford, Oxford University Press, pp. 387-412.

Ryan's proposal to reframe the definition of paternalism allows to include also its "libertarian" version as full-fledged paternalism (while for Gerald Dworkin it does not qualify as such). Even if an intervention does not qualify as an interference with someone else's liberty or autonomy (as in the intentions of libertarian paternalists), it can still be made irrespective of the wishes or consent of those who receive it, with the intent of improving their welfare or good otherwise intended¹¹.

3. Nudges as a threat to liberty and autonomy

Moreover, for Ryan libertarian paternalism is *hard* paternalism, or at least it is compatible with hard paternalism, to the extent that it is not merely aimed at checking whether someone else is acting voluntarily and knowledgeably – as in *soft* paternalism – but it can also be directed at obtaining some specific outcome, which is considered to be good for those acted upon. However, if we have to take seriously Thaler and Sunstein's contention that libertarian paternalism aims at making choosers better off, *as judged by themselves*¹², then nudges prompted by this form of paternalism actually seek to respect people's preferences, to avoid curtailing their liberty or autonomy and restricting the options available to them. So-called "restrictive nudges" are not an expression of libertarian paternalism as seen by its proponents. The aspect of welfare-maximization is entirely subordinate to respecting the preferences of the nudged, or – better said – welfare is entirely interpreted in terms of preference satisfaction.

Now, we could hold that autonomy and/or liberty, understood as "self-rule", cannot be reduced to unrestricted freedom of choice, as per the libertarian tradition. There are more demanding conceptions of both autonomy and liberty, which conceive of them as freedom from "alien control" (as in the republican tradition) or from subjection to someone else's will (which is part of the Kantian conception of external freedom)¹³. For these "relational" conceptions, autonomy and liberty include not being ruled by others, i.e. retaining a measure of *control* – actual or, at least, virtual – over external interferences. As said above, for some, nudges do not allow those interfered with to retain such control, therefore they are incompatible with autonomy and/or liberty. Why so?

Nudges – so goes the argument – are not transparent enough for people to be aware they are being interfered with and to exercise some measure of control over them. If they are made transparent, they become ineffective. However, to support this argument, empirical evidence should be provided. Actually, the available evidence seems to show that many nudges work even if we are aware of them: for instance, reduced size of food containers can impact on the quantity

11. S. RYAN, "Libertarian paternalism is hard paternalism", *Analysis*, 78, 1, 2018, pp. 65-73.

12. THALER and SUNSTEIN, *Nudge*, 5.

13. I. KANT, *The Metaphysics of Morals*, Cambridge, Cambridge University Press, 1996; see also C. F. ROSTBØLL, "Kant, Freedom as Independence, and Democracy", *The Journal of Politics*, 78, 3, 2016, pp. 792-805.

of food we eat even if we are aware of this (very simple) nudge.¹⁴ After all, sometimes we even nudge ourselves in order to control some aspects of our behaviour. Now, what can happen once nudges are detected by those addressed? If they recognize that the nudge in question favours certain results against their preferences, they can simply decide to pursue their preferences as these would be without the nudge: for instance, they can still have chips with their burger instead of the salad they are being nudged to buy. If, of the contrary, the nudge goes in the same direction as their preferences, they can experience *preference effects* or *reactance effects*. The first, which usually prevail, make people act according to the nudge, because they recognize it favours their preferences. The second make people react unfavourably and resist the nudge even if it fulfils their initial preferences – people do not like to feel that they are being manipulated – but reactance effects are more unusual and can easily be contrasted by means of dialogue and explanation¹⁵.

Therefore, at most we can assume that some nudges might be less effective when made transparent. The trade-off between reduced effectiveness and enhanced transparency could be easily accepted¹⁶. Thence, control on the nudge by those who are nudged is feasible and their autonomy/liberty can be protected even under demanding conceptions of personal freedom. What kind of control should we require? *Type transparency*, i.e. disclosure of general nudge policies, is usually deemed insufficient, although it is usually the kind of transparency granted by democratic law-making. *Token transparency* seems to be required, in the sense of the recognizability of single concrete instances of nudging by those who are being nudged. Token transparency, however, should be interpreted in the sense of nudges being relatively easy to recognize, rather than being actually recognized by each and every subject: we should make reference to nudges easily detectable “in principle” by a “watchful person”; “[i]t should not be unreasonably costly to seek and find relevant information about the general types of nudging. Moreover, information about types should make it possible to recognize tokens of these policies through inference and without requiring unreasonable effort”¹⁷.

It appears, then, that the problem with nudges and their interference with our lives is one of transparency and control. In this sense, however, it should also be pointed out that non-nudging, “direct” strategies of command & control legislation often are no more transparent, and even less controllable, than nudges. As for transparency, think, for instance, of the extreme complexity of EU regulations concerning food: it is not easy at all for the typical European consumer to ascertain how these affect the availability of products and, therefore, her options of food consumption. As for control, coercive directives assisted by legal sanctions surely do not yield much control to those addressed. As we have seen, nudging is not necessarily autonomy-restrictive and, by all means,

14. B. WANSINK, *Mindless Eating: Why We Eat More Than We Think*, New York, Bantam Books, 2007, pp. 60-70.

15. A. T. SCHMIDT, “The Power to Nudge”, *American Political Science Review*, 111, 2, 2017, pp. 404-417, at 409.

16. SCHMIDT, *APSR*, p. 409.

17. SCHMIDT, *APSR*, p. 410. Schmidt calls this “reasonable token inference transparency”.

restrictions of autonomy by command & control strategies can be much more severe.

Lastly, it is worth reminding that different kinds of nudges interfere differently with the autonomy and liberty of the nudged; *qualitative* differences impact on the *quantity*, so to say, of the interference and on the possibility of transparency. In this sense, Robert Baldwin has distinguished between three “degrees” of nudging¹⁸. First Degree nudges are meant to enhance the individuals’ capacity of reflective decision-making, usually by providing information, reminders, etc. Therefore, they do not exploit those very same faulty mental processes and volitional limitations that can hinder satisfactory choices. Second Degree nudges, on the contrary, rely on those very limitations in order to push individuals towards certain desirable outcomes, although without altering the set of available choices, at least theoretically. Default rules with opt-out options are typical examples of Second Degree nudging, because they exploit human inertia. However, the addressed individuals would be capable, on reflection, to identify the nudge and to take a different course from the one which is being suggested, if they wish so. Lastly, Third Degree nudges are more “resistant to unpacking”, because even on reflection it is difficult to assess the extent of the nudging and its quality (since biases are exploited in a subtle way, like when loss aversion is targeted); or even because they have been intentionally devised to stand in the way of reflective thinking (as when emotions like fear are triggered). The higher the degree, the more problematic (i.e. difficult to justify) the nudge. However, once again the problem seems to be with token transparency: with Third Degree nudges, the targeted individuals cannot understand (with a reasonable effort) that they are being nudged, and how. But once the relevant information about the implementation of nudging policies is adequately spread through society, even token transparency should not be difficult to achieve: think, for instance, of scary graphic warnings on cigarette packages, which are supposedly a kind of Third Degree nudge targeting basic emotions like fear; their function is clear to everyone. Once an acceptable degree of transparency is achieved, the question of acceptability turns to whether the nudge respects the people’s preferences or, on the contrary, it manipulates them to change those preferences. This leads us to consider the specific issues raised by paternalistic nudging.

4. Is paternalistic nudging more problematic?

The second thesis which I wish to address is the idea that paternalistic nudging should be regarded as *prima facie less acceptable* than non-paternalistic nudging and, therefore, that it is subject to a heavier burden of justification. I think that this thesis is based on two unwarranted assumptions. The first is a false dichotomy between autonomy and rationality. Van Aaken, for instance, ascribes to Thaler and Sunstein an objective conception of welfare that can conflict with

18. R. BALDWIN, “From Regulation to Behaviour Change: Giving Nudge the Third Degree”, *Modern Law Review*, 77, 6, 2014, pp. 831-857.

people's different conceptions of a "good life" and self-chosen lifestyles. In her view, assessing people's choices against a "full rationality standard" amounts to disregarding their autonomy and limiting their possibility to freely choose the values they want to pursue¹⁹. But this is false: as we have seen, these authors propound nudging as a means for allowing people to make choices that are better in their own eyes. Rationality, in their view – the rationality of what they call "Econs", i.e. perfect economic agents, as contrasted to Humans – is entirely subservient to maximizing preference satisfaction, not to tread on those preferences. Actually, enhancing rational decision or – as in other strategies – mimicking its results should strengthen, not weaken, autonomy. In some cases, as in "counter-nudging" strategies, public nudges inspired by libertarian paternalism can contrast the effects of nudging applied by private actors for profit, which, as a form of interference not subjected to control by those addressed, limits their autonomy and leads to choices which are typically regretted later on. Of course, it should be taken into account that temporary or *actual* preferences (for instance, as in short-term temptations) differ from more stable and informed *true* preferences, which are those that usually drive later remorse and repent for earlier, unsatisfactory choices. Alignment between choices and true preferences should be the purpose of libertarian paternalistic nudges²⁰. Now, it could very well be that even respecting true and stable preferences is a difficult task for nudging policies; sometimes, choice architects end up influencing an individual's actions in a direction which reflects what they "presume her to prefer or value but not necessarily what she prefers or values"²¹. As public nudges are usually part of broad and general policies targeting wide groups of individuals, this risk is always present. But, at least in principle, libertarian paternalist choice architects cannot be accused of superposing their own substantive view of rationality on individuals, thereby impairing their autonomy.

The second, unfounded assumption is that nudging in order to shape, or to change, the people's informed and stable true preferences, is compatible with libertarian paternalism. Unlike cognitive and "means" paternalism, preference or "end" paternalism is incompatible with the kind of paternalism advocated by Thaler and Sunstein²². Assuming that people do not know how to live a "good life" and interfering with the formation of their preferences can at best be a form of moral paternalism, which, of course, is something which liberal, pluralistic democracies should guard against. Moral paternalism is a kind of perfectionism: the above-mentioned "condition of benevolence", implied by paternalistic action, is respected, in that the paternalist agent interferes with another individual in order to promote her interest or good, but the good in question is interpreted in

19. VAN AAKEN, *Constitutional Limits to Paternalistic Nudging*, p. 170.

20. K. GRILL, "Respect for What? Choices, Actual Preferences, and True Preferences", *Social Theory and Practice*, 41, 4, Special Issue: Preference, Choice, and (Libertarian) Paternalism, 2015, pp. 692-715,

21. W. GLOD, "How Nudges Often Fail to Treat People According to Their Own Preferences", *Social Theory and Practice*, 41, 4, Special Issue: Preference, Choice, and (Libertarian) Paternalism, 2015, pp. 599-617.

22. C. R. SUNSTEIN, *Why Nudge? The Politics of Libertarian Paternalism*, New Haven and London, Yale University Press, 2014, p. 19.

a moral sense and, as said above, can be wholly indifferent to that individual's preferences, both actual and true. Moral paternalism often rests on the dubious notion of a "moral harm" which it tries to avert.²³ But it is unclear how moral paternalism could employ nudges: usually, stable preferences are shaped by education (mostly, though not exclusively, at a young age) or acted upon by means of coercive regulation and proper incentives, both positive and negative. At best, genuine nudges can act on temporary preferences or (better said) wishes. Stable preferences are those that libertarian paternalists seek to satisfy and to protect against bad choices, prompted by biases, weakness of will and similar shortcomings.

Therefore, the problem with nudges is not paternalism per se, but rather a certain kind of paternalism, which usually resorts to coercive methods or to education rather than nudging. Nor are nudges more problematic than other kinds of policies, at least when they are made sufficiently transparent: something which can be attained without impairing their effectiveness, at least not to an unacceptable degree.

23. J. FEINBERG, *The Moral Limits of Criminal Law, Volume Three: Harm to Self*, Oxford, New York, etc., Oxford University Press, 1986, p. xvii. See also G. DWORKIN, "Moral Paternalism", *Law and Philosophy*, 24, 3, 2005, pp. 305-319; P. DE MARNEFFE, *Moralism and Moral Paternalism*, in K. GRILL and J. HANNA (eds.), *The Routledge Handbook of the Philosophy of Paternalism*, London-New York, Routledge, pp. 35-45.