



Keeling, G. (2017). Autonomy, Nudging and Post-Truth Politics. *Journal of Medical Ethics*. <https://doi.org/10.1136/medethics-2017-104616>

Peer reviewed version

License (if available):
Other

Link to published version (if available):
[10.1136/medethics-2017-104616](https://doi.org/10.1136/medethics-2017-104616)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via BMJ at <http://dx.doi.org/10.1136/medethics-2017-104616> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/pure/user-guides/explore-bristol-research/ebr-terms/>

Autonomy, Nudging and Post-Truth Politics

Geoff Keeling

Abstract: In his excellent essay, 'Nudges in a post-truth world', Neil Levy argues that 'nudges to reason', or nudges which aim to make us more receptive to evidence, are morally permissible. A strong argument against the moral permissibility of nudging is that nudges fail to respect the autonomy of the individuals affected by them. Levy argues that nudges to reason do respect individual autonomy, such that the standard autonomy objection fails against nudges to reason. In this paper, I argue that Levy fails to show that nudges to reason respect individual autonomy.

Keywords: Ethics, Public Policy, Political Science

I begin with a statement of the problem [1]. First, according to Levy's *descriptive commitment*, we live in a 'post-truth' era where many political actors, including voters and politicians, are unreceptive to evidence against their political views. Second, according to Levy's *prescriptive commitment*, it would be a good thing if political actors were more receptive to evidence against their political views. It would, at least, enable people to make more informed political decisions. Third, according to our best psychological evidence, when we present people with evidence against their deeply-held commitments, this often causes them to hold these commitments more firmly. This is called the *backfire effect* [2,3].

Here is the problem that Levy addresses:

(Q) How can we make people more receptive to evidence against their political views conditional on the existence of the backfire effect?

Levy's aim is to provide an answer to (Q), and explain why his answer is morally permissible. In what follows, I first explain Levy's answer to (Q). I then explain Levy's argument for the moral permissibility of his answer, and argue that his argument fails to support its conclusion.

Levy's answer to (Q) draws on a mounting body of psychological evidence in favour of the claim that the *way* in which evidence is presented to people affects their receptiveness to that evidence [4]. For example, if someone from our own political standpoint presents evidence against one of our political beliefs, we are more likely to be receptive to this evidence than if someone from a contrary political standpoint presents that same evidence. Levy proposes that policymakers should exploit this phenomenon. On Levy's view, policymakers should develop 'nudge' policies which manipulate the sources through which individuals receive evidence against their political beliefs in order to make them more receptive to this evidence.

What makes nudge policies different from coercive policies is that nudges do not limit the set of options available to the agent, but instead exploit known cognitive heuristics and biases to present choices in a way that increases the probability of the individual choosing an option that is in their best interests [5]. A standard

example is presenting healthier food at eye-level in a food store to ‘nudge’ individuals towards buying healthier foods. Levy’s answer to (Q) achieves two aims: first, it encourages individuals to engage with evidence against their political views whilst minimising the chances of a backfire effect. Second, it does so in a way that does not involve policymakers coercing people to form certain views, as the recipients of the nudge are left to make an informed decision on the basis of *more*, and not *less* evidence.

I have presented Levy’s answer to (Q). I now explain Levy’s reasons for thinking that it would be morally permissible for policymakers to implement these policies. A strong argument against the moral permissibility of nudging is that nudges fail to respect individual autonomy. The roots of this objection trace back to Kant. The basic idea is that individuals are practical reasoners who make decisions based on reasons with the aim of satisfying their preferences. The capacity to reason, and to act on reasons, seems to demand a kind of respect, according to which individuals cannot be used merely as a means to satisfy another’s preferences. It has been argued that nudges fail to respect individual autonomy insofar as they exploit our cognitive weaknesses to influence our decision, which gives insufficient moral weight to our status as practical reasoners. Levy’s argument for the moral permissibility of ‘nudges to reason’, that is, nudges which aim to make us more receptive to evidence, is that these nudges *do* respect individual autonomy so understood.

Here is Levy’s argument. Levy first challenges the distinction between our ‘affective’ and ‘deliberative’ cognitive apparatus. He does this by highlighting the role that emotions play in our deliberations: it would be difficult for us to make decisions in the absence of heuristics and biases. Levy goes on to argue that our heuristics and biases plausibly evolved to enable *better*, more *truth conducive* decision-making. As such, it is false to say that nudges to reason bypass our deliberative mechanisms in order to influence our beliefs. The affective mechanisms ‘exploited’ by nudges to reason are just as much a part of our deliberative mechanisms as our ‘cognitive’ decision-making apparatus. On these grounds, Levy claims that nudges to reason are more accurately construed as an appeal to our deliberative faculties, as opposed to an attempt to subvert them. Levy concludes that nudges to reason do not, therefore, fail to respect individual autonomy.

I believe that Levy’s argument is unsuccessful: he does not establish that nudges to reason respect individual autonomy. In short, his mistake is this. It is true that the autonomy objection to nudging is sometimes reliant on the claim that nudges subvert an individual’s deliberative apparatus. But this is only used as evidence for the claim that nudges involve a kind of *deceit* which is morally inappropriate given the kind of respect which is owed to practical reasoners. Levy might be correct that nudges to reason do not subvert an individual’s deliberative cognitive mechanisms. But it does not follow that nudges to reason do not deceive individuals in the relevant sense. It is possible to deceive an agent in a morally inappropriate way even if the deceit does not subvert the agent’s deliberative mechanisms. For example, suppose you must choose between *A* or *B*. There are exactly two pieces of evidence relevant to your decision, call them *E1* and *E2*. Suppose you would have overriding reasons to choose *A* if you knew about exactly one of *E1* or *E2*, but decisive reasons to choose *B* if you knew about both *E1* and *E2*. If I omit to tell you about exactly one piece of evidence before you make the decision, then I have deceived you in a way that shows insufficient respect to your status as a practical reasoner. But I did not subvert your deliberative mechanisms. I *exploited* them so

that you would choose *A* over *B*. Thus, even if Levy is correct that nudges to reason do not bypass an individual's cognitive mechanisms, it does not follow that nudges to reason respect individual autonomy.

Even if Levy's argument fails, it remains to be shown that nudges to reason *actually do* deceive individuals in a way which fails to respect their autonomy. I shall offer some brief motivation for this point. Let us suppose that a newspaper, *N*, from a political standpoint, *S*, is legally required to publish at least some arguments in favour of policies grounded in a different political standpoint, *not-S*. It strikes me that such a policy fails to respect the reader's autonomy in the following respect: according to the Kantian considerations that motivate the autonomy objection, we ought to treat the readers of *N* as though they have weighed-up the relevant epistemic reasons in favour of and against reading *N*, and chosen to take *N* as their source of political information for good reasons. To implement a policy which requires the newspaper to present evidence in favour of some opposing political standpoint, *not-S*, with the aim of making the readers receptive to other sources of evidence, fails to take seriously those readers as individuals capable of making informed decisions about which political evidence they give most weight to. In essence, the failure to respect autonomy lies in the assumption that *but for* the nudge, the reader would be incapable of making an informed judgement about which information sources she has best epistemic reason to use.

Levy might respond that we have good empirical evidence in favour of the conclusion that many newspaper readers do not weigh-up the epistemic reasons for and against using their preferred newspaper as a source of political information. But the autonomy objection is concerned with how individuals ought to be treated in virtue of their capability to act as practical reasoners. The fact that individuals are not responsive to reasons in all cases does not undermine the moral importance of treating them *as if* they had considered all the relevant epistemic considerations.

In conclusion, Levy argues that nudges to reason are morally permissible because they respect individual autonomy. I have argued that this is untrue: nudges to reason do not respect individual autonomy.

Bibliography

- [1] Levy, N. Nudges in a post-truth world. *J Med Ethics* 2017; 43:495-500.
- [2] Nyhan B, Reifler J, Richey S, et al. Effective messages in vaccine promotion: a randomized trial. *Pediatrics* 2014; 133:e835-42.
- [3] Nyhan B, Reifler J. Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine* 2015; 33:459-64.
- [4] Harris P. Trusting what you're told: how children learn from others. Cambridge, MA: Harvard University Press, 2012
- [5] Thaler HT, Sunstein CR. Nudge. New Haven, CT: Yale University Press, 2008.