

# The distributional effects of nudges

Nudges are tools to achieve behavioural change. To evaluate nudges, it is essential to consider not only their overall welfare effects but also their distributional effects. Some nudges will not help, and might hurt, identifiable groups. More targeted, personalized nudging may be needed to maximize social welfare and promote distributive justice.

Cass R. Sunstein

In recent years, both public and private institutions have shown a great deal of interest in ‘nudges’, understood as interventions that preserve freedom of choice but also steer people in particular directions<sup>1,2</sup>. Nudges fall into two categories: the educative and the architectural. Educative nudges include warnings, reminders and disclosure of information (such as calorie labels, allergy warnings and fuel economy labels). Architectural nudges include automatic enrolment, mandatory choice, simplification or ‘sludge reduction’<sup>3</sup>, and design of websites, forms or in-person shops to highlight and draw attention to certain options. Whether educative or architectural, nudges have often been found to have substantial effects on outcomes and to be highly cost-effective. For example, a shift from an opt-in to an opt-out design can greatly increase participation rates in relevant programmes (for example, involving retirement savings or consumer protection)<sup>4</sup>.

## Disaggregating the effects of nudges

It is increasingly clear, however, that such aggregate effects will not tell public and private institutions everything that they need to know. To evaluate an increase in participation rates, we need to know whether participation is actually beneficial for the relevant population. Increased savings could reduce welfare if people need the money now. And without some kind of disaggregation, we might not know whether a nudge is helping or hurting identifiable groups, whether it is causing serious unintended harm to some or many, or whether it should be redesigned in some important way.

As an example, consider a case in which numerous workers have been automatically enrolled in a retirement plan with a high default contribution rate, and in which — as a result of automatic enrolment — participation rates have become very high, especially among low-wage workers. It is possible that with automatic enrolment, low-wage workers have been made worse off. They might benefit from having the

money now, and despite that fact they might not opt out (perhaps because of inattention and inertia).

Whenever automatic enrolment increases participation rates across the board, there is at least a risk that some people will be hurt<sup>5</sup>. This risk is especially troubling if an architectural nudge causes serious harm to identifiable categories of people: for example, people who are poor, elderly individuals, women, people of colour or people suffering from physical or mental disabilities. An appropriate response might be a more targeted approach, automatically enrolling only those groups whose members are likely to benefit.

For educative nudges, there are corresponding risks. Some evidence<sup>6</sup> suggests that people with a good deal of self-control, and perhaps without much need to lose weight, value calorie labels and benefit from them — but that people who tend to lack self-control, and who may perhaps have a real need to lose weight, are less likely to value those labels at all and less likely to benefit from them (and may in fact be harmed by them). To that extent, it is at least possible that calorie labels may not be justified on welfare grounds; it is also possible that such labels, and other educative nudges, will have undesirable distributional effects, hurting the people they are intended to help<sup>7</sup>.

Although there is limited evidence on this point, some educative nudges could increase the behaviour that they are designed to promote and for that reason harm members of identifiable groups. It is plausible to think that if people are informed that their behaviour will be observed and thus that they might receive reputational rewards, they will act in a more prosocial manner (not polluting, not driving while drunk or giving to charity), which suggests that observability can be an effective nudge<sup>1</sup>. At the same time, an observability nudge has been found in some circumstances to be counterproductive<sup>8</sup>, perhaps because of inequality concerns: if people are less well-off, knowledge that their behaviour is being observed might make them less likely, not more likely, to give to charity.

It would be valuable to learn more about the circumstances in which an educative nudge reduces the behaviour it is meant to encourage (as, for example, when a warning backfires), has unintended and unwelcome distributional effects, or makes people less likely to engage in action that increases collective welfare.

## Nudges can help those most in need

Importantly, many nudges turn out to be more, rather than less, desirable once we disaggregate their effects. Focusing on architectural nudges, Mrkva et al.<sup>9</sup> find that nudges markedly reduce barriers that contribute to inequality; the relevant areas include COVID-19 health decisions, retail purchases and financial decisions. In this light, we can consider a situation in which an employer or an educational institution is examining the appropriate choice architecture for decisions involving, for example, health insurance or retirement. Smart defaults, nudging people towards options that are most likely to suit their situations, can have especially large benefits for low-income choosers, for choosers with less domain-relevant expertise and for choosers with lower numeracy<sup>9</sup>. Another example might be a case in which a government agency decides to automatically enrol homeless and migrant children in order to increase participation in a programme designed to provide free meals to children who are poor; on distributional grounds, such a nudge should be enthusiastically welcomed. More broadly, ‘sludge reduction’, in the form of efforts to reduce administrative burdens and barriers, can have substantial benefits for people who are old, sick, poor or otherwise vulnerable<sup>3</sup>.

In the same vein, some educative nudges could disproportionately benefit people at the bottom of the economic ladder, especially in light of the challenge of cognitive scarcity<sup>10</sup>; for example, warnings or reminders designed to help people to avoid late fees or overuse fees, or not to miss medical appointments. To the extent that warnings or reminders target people who are busy, occupied or inattentive, their

distributional effects might turn out to count in their favour<sup>11</sup>. Indeed, regulation of the credit card market, motivated in large part by behavioural findings, has been found to provide special help to people with poor credit ratings<sup>12</sup>.

### Achieving distributive justice

For some behaviourally informed interventions, normative assessment of their distributional effects is not at all straightforward even if the evidence is clear. Let us return in this light to architectural nudges and in particular to automatic enrolment in savings plans, supposing once more (just for purposes of analysis) that one of its consequences is to produce especially high participation rates among low-income employees. It would not be obvious how to evaluate any such finding. A low opt-out rate provides at least some assurance that participation is in the interest of low-income employees, or at least not contrary to their interest. Because such employees may have a special need for retirement savings, a large increase in participation rates might suggest that, on distributional grounds, automatic enrolment is a substantial success.

To evaluate this suggestion, it would be valuable to assemble evidence about the full set of consequences on the behaviour and welfare of those who end up with less take-home pay<sup>13</sup>. For both architectural nudges (such as automatic enrolment in green energy)<sup>14</sup> or educative nudges (such as reminders or health warnings), it would

also be valuable to be able to predict those consequences in advance, so as to increase the likelihood that nudges will be used only when we have reason to be confident that their effects will be beneficial. An appreciation of the lessons from the research outlined here, alongside an understanding of the relevant population (is it diverse and, if so, along what dimensions?) and the particular nudge at issue (is it architectural or educative?), can provide valuable insights.

The broadest lesson is that in evaluating nudges, it is essential for researchers, and for those in the public and private sectors, to consider their distributional effects and to focus in particular on questions of distributive justice. They should ask specific questions: (1) who is likely to be helped, and who is likely to be hurt? (2) What are the expected effects on the least well-off? Will the relevant nudges benefit those who are most in need of help? (3) Do the benefits to those who are helped exceed the costs to those who are hurt? In some cases, answers to those questions will actually strengthen the argument for nudges; in other cases, such answers will weaken that argument; and in still other cases, such answers will raise a set of new issues, normative and empirical. The most important implication might well involve the value of shifting towards more targeted or personalized nudging, which can often produce higher welfare benefits, and be far better on distributional grounds, than ‘mass’ approaches<sup>9,15</sup>.

Cass R. Sunstein  

Harvard University, Cambridge, MA, USA.

✉ e-mail: [csunstei@law.harvard.edu](mailto:csunstei@law.harvard.edu)

Published online: 8 November 2021  
<https://doi.org/10.1038/s41562-021-01236-z>

### References

1. Thaler, R. H. & Sunstein, C. *Nudge: The Final Edition* (Penguin, 2021).
2. Beshears, J. & Kosowsky, H. *Organ. Behav. Hum. Decis. Process.* **161** (Suppl.), 3–19 (2020).
3. Sunstein, C. *Sludge: What Stops Us From Getting Things Done And What To Do About It* (MIT Press, 2021).
4. Jachimowicz, J. M., Duncan, S., Weber, E. U. & Johnson, E. J. *Behav. Public Policy* **3**, 159–186 (2019).
5. Roberts, J. *Mich. Law Rev.* **116**, 1045–1066 (2018).
6. Thunström, L. *Judgm. Decis. Mak.* **14**, 11–25 (2019).
7. Thunström, L., Gilbert, B. & Ritten, C. J. *J. Econ. Behav. Organ.* **153**, 267–282 (2018).
8. Bolton, G., Dimant, E. & Schmidt, U. *J. Econ. Behav. Organ.* **191**, 946–964 (2021).
9. Mrkva, K., Posner, N. A., Reeck, C. & Johnson, E. J. *J. Mark.* **85**, 67–84 (2021).
10. Mullainathan, S. & Shafir, E. *Scarcity: Why Having Too Little Means So Much* (Times Books, 2013).
11. Bocoum, F., Grimm, M., Hartwig, R. & Zongo, N. *Soc. Sci. Med.* **220**, 102–111 (2019).
12. Agarwal, S., Chomsisengphet, S., Mahoney, N. & Stroebel, J. Q. *J. Econ.* **130**, 111–164 (2015).
13. Beshears, J., Choi, J., Laibson, D., Madrian, B. C. & Skimmyhorn, W. J. *Finance*, <https://doi.org/10.1111/jofi.13069> (2021).
14. Liebe, U., Gewinner, J. & Diekmann, A. *Nat. Hum. Behav.* **5**, 576–585 (2021).
15. Mills, S. *Behav. Public Policy*, <https://doi.org/10.1017/bpp.2020.7> (2020).

### Competing interests

The author declares no competing interests.

### Additional information

**Peer review information** *Nature Human Behaviour* thanks Linda Thunström and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.