Behavioural Public Policy (2017), 1: 1, 4–25
© *Cambridge University Press* doi:10.1017/bpp.2016.3

Nudges that fail

CASS R. SUNSTEIN*

Abstract: Why are some nudges ineffective, or at least less effective than choice architects hope and expect? Focusing primarily on default rules, this essay emphasizes two reasons for this. The first involves strong antecedent preferences on the part of choosers. The second involves successful "counternudges," which persuade people to choose in a way that confounds the efforts of choice architects. Nudges might also be ineffective, and less effective than expected, for five other reasons: (1) some nudges produce confusion in the target audience; (2) some nudges have only short-term effects; (3) some nudges produce "reactance" (though this appears to be rare); (4) some nudges are based on an inaccurate (though initially plausible) understanding on the part of choice architects of what kinds of choice architecture will move people in particular contexts; and (5) some nudges produce compensating behavior, resulting in no net effect. When a nudge turns out to be insufficiently effective, choice architects have three potential responses: (1) do nothing; (2) nudge better (or differently); and (3) fortify the effects of the nudge, perhaps through counter-counternudges, or perhaps through incentives, mandates, or bans.

Submitted 3 August 2016; accepted 11 October 2016

Recent years have seen a keen interest in "nudges," defined as approaches that steer people in particular directions, but that also allow them to go their own way (Sunstein & Thaler, 2008; Jones *et al.*, 2014; Executive Order, 2015; Halpern, 2015; Thaler, 2015; Alemanno & Sibony, 2016; Mathis & Tor, 2016). In daily life, a GPS device is an example of a nudge; so is an "app" that tells people how many calories they consumed during the previous day; and so is a text message, informing customers that a bill is due or that a doctor's appointment is scheduled for the next day. In government, nudges span an exceptionally wide range. They include graphic warnings for cigarettes (Jolls, 2015); labels containing information about energy efficiency or fuel economy; "nutrition facts" panels on food; default rules for public assistance

* Correspondence to: Cass R. Sunstein, Robert Walmsley University Professor, Harvard Law School, USA. Email: csunstei@law.harvard.edu

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

programs (as in "direct certification" of the eligibility of poor children for free school meals) (Conway *et al.*, 2015); texts and email messages; and even the design of government websites, which list certain items before others and in large fonts (Krug, 2014). In these cases, and many others, nudges work by altering the "choice architecture," understood as the background against which choices are made (Balz *et al.*, 2013).

The interest in such choice-preserving approaches stems from the belief – or the hope – that they can be highly effective in achieving important social goals (Pichert & Katsikopoulos, 2008; Bettinger *et al.*, 2009; Allcott, 2011; Chetty *et al.*, 2012; Benartzi & Thaler, 2013; Egebark & Ekstrom, 2013; Manoli & Turner, 2014; Conway *et al.*, 2015; Ebeling & Lotz, 2015; Halpern, 2015). Even if people are allowed to go their own way, default rules, warnings, reminders, invocations of social norms, and other forms of choice architecture can have significant consequences on behavior. Sometimes the magnitude of those consequences is unexpected (Thaler, 2015). But no one should deny that some nudges are ineffective or counterproductive (Executive Office of the President, 2015). For example, information disclosure might have little impact, certainly if it is too complicated for people to understand, and sometimes even if it is simple and clear (Golman *et al.*, 2014). If people are told about the number of calories in candy bars, they might not learn anything they do not already know, and even if they learn something, they might be unaffected. A reminder might fall on deaf ears; a warning might be ignored (or even make the target action more attractive). In some cases, a plausible (and abstractly correct) understanding of what drives human behavior turns out to be wrong in a particular context; once a nudge is tested, it turns out to have little or no impact.

In the terms made famous by Albert Hirschman, nudging might therefore be *futile* (Hirschman, 1991), or at least close to it. Alternatively, their effects might also be *perverse*, in the sense that they might have the opposite of the intended consequence – as, for example, when calorie labels increase caloric intake (Downs *et al.*, 2013). To complete Hirschman's trilogy, nudges may also *jeopardize* other important goals, as when a nudge, designed to reduce pollution, ends up increasing energy costs for the most disadvantaged members of society (Sunstein & Reisch, 2013). Hirschman's main goal was to explore "the rhetoric of reaction," not to suggest that futility, perversity, and jeopardy are inevitable or even likely. On the contrary, he saw them as predictable rhetorical moves, sometimes offered in bad faith. But there is no question that public-spirited reforms, including nudges, often run into each of these three objections. Futility, perversity, and jeopardy may reflect reality rather than rhetoric.

Of all of the tools in the choice architect's repertoire, default rules may be the most promising; they are almost certainly the most discussed (Johnson &

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

Goldstein, 2013). Whether the area involves savings behavior, poverty reduction, or the environment, default rules have had significant effects on outcomes (Johnson & Goldstein, 2003; Johnson & Goldstein, 2013), but sometimes they do very little, or at least far less than anticipated (Cain *et al.*, 2012; Willis, 2012; Bubb & Pildes, 2014; Fagerlin *et al.*, 2016). My principal goal here is to identify two reasons why this might be so. The first involves *strong contrary preferences on the part of the chooser*, who will therefore opt out. The second involves *countervailing behavior on the part of those whose economic interests are at stake*, who may be able to move choosers in their preferred direction (often with the assistance of behavioral insights). As we shall see, these two explanations help account for the potential ineffectiveness of many other nudges, not only default rules.

It is a useful simplification to posit that in deciding whether to depart from default rules, or to reject nudges of any kind, choosers consider two factors: the costs of decisions and the costs of errors. When it is not especially costly to decide to reject a nudge, and when choosers believe that doing so will reduce significant error costs, a nudge will be ineffective. We shall also see, though more briefly, other reasons as to why nudges might be ineffective; of these, the most important is the choice architect's use of a plausible (but ultimately mistaken) hypothesis about how choice architecture affects behavior.

If nudges do not work, there is of course the question of what to do instead. The answer depends on normative criteria. A nudge might turn out to be ineffective, or far less effective than expected, but that might be a good thing; it might explain why choice architects chose a nudge rather than some other instrument (such as a mandate). Suppose that choice architects care about social welfare, and that they want to increase it (Alder, 2011). If so, promoting welfare provides the right criterion, and effectiveness does not. By itself, the ineffectiveness of nudges – or, for that matter, their effectiveness – tells us little and perhaps even nothing about what has happened to social welfare. Imagine that 90% of a relevant population opts out of a default rule, so that the nudge is largely ineffective, or imagine that 10% opts out, or that 50% does. In all of these cases, choice architects must consider whether the result suggests that the nudge is, all things considered, a success or a failure, and in order to undertake that consideration, they must ask particular questions about the consequences for social welfare.

One answer is that if a nudge is ineffective, or less effective than expected, it is because it is not a good idea for those who were unaffected by it. Its failure is instructive and on balance should be welcomed, in the sense that if choosers ignore or reject it, it is because they know best. That answer makes sense if ineffectiveness is *diagnostic*, in the sense that it demonstrates that people are acting in accordance with their (accurate) sense of what will promote their welfare.

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

Sometimes that conclusion is correct, but for good behavioral reasons (Sunstein & Thaler, 2008), it sometimes is not.

A second answer is to try a different kind of nudge. It is important to test behaviorally informed interventions and to learn from those tests (Halpern, 2015); what is learned might well point in the direction of other nudges. That answer might be best if the people's choices (e.g. to ignore a warning or to opt out) are based on confusion, bias, or misunderstanding, and if a better nudge might dispel one or all of these. Many nudges do not, in fact, raise hard normative questions. They are designed to promote behavior that is almost certainly in the interest of choosers or society as a whole (Executive Office of the President, 2015); the failure of the nudge is not diagnostic. In such cases, a better nudge may well be the right response.

A third answer is to undertake a more aggressive approach, going beyond a nudge, such as an economic incentive (a subsidy or a tax), or coercion. A more aggressive approach might make sense when the choice architect knows that the chooser is making a mistake (Conly, 2013), or when the interests of third parties are involved. Some nudges are designed to protect such interests; consider environmental nudges, or nudges that are intended to reduce crime. In such cases, choice-preserving approaches might well prove inadequate or at best complementary to incentives, mandates, and bans.

Preliminaries: why default rules stick

Default rules are my principal topic here, and it makes sense to begin by explaining why such rules have such a large effect on outcomes. A great deal of research has explored three reasons for this, which should now be familiar (Carroll *et al.*, 2009; Johnson & Goldstein, 2013; Sunstein, 2015). The first involves inertia and procrastination (sometimes described as "effort" or an "effort tax"). To alter the default rule, people must make an active choice to reject that rule. Especially (but not only) if they are busy, or if the question is difficult or technical, it is tempting to defer the decision or not to make it at all. In view of the power of inertia and the tendency to procrastinate, they might simply continue with the status quo. Attention is a scarce resource, and it is effortful to engage it; a default rule might stick because that effort does not seem to be worth undertaking.

The second factor involves what people might see as the informational signal that the default rule provides. If choice architects have explicitly chosen that rule, many people will believe that they have been given an implicit recommendation, and by people who know what they are doing (and who are not objectionably self-interested). If so, they might think that they should not depart from it and go their own way, unless they have private information that is

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

reliable and that would justify a change (Madrian & Shea, 2001; Finkelstein *et al.*, 2006). Going one's own way is risky, and people might not want to do so unless they are quite confident that they should.

The third factor involves loss aversion, one of the most important and robust findings in behavioral science: people dislike losses far more than they like corresponding gains (Zamir, 2014). For present purposes, the key point is that the default rule establishes the status quo; it determines the reference point for counting changes as losses or instead as gains. If, for example, people are not automatically enrolled in a savings plan, a decision to enroll might well seem to be a loss (of salary). If people are automatically enrolled, a decision to opt out might well seem to be a loss (of savings). The reference point is established by the default rule.

Strong antecedent preferences

To see why default rules may be ineffective, begin with the case of marital names (Emens, 2007). When people marry, all states in the United States have the same default rule: both men and women retain their pre-marriage surnames. In the overwhelming majority of cases, American men do stick with the default. Relatively few men change their names. By contrast, the overwhelming majority of American women do change their names – for college graduates, the rate of surname change amongst women is 80% (Emens, 2007). In that respect, the default rule seems to have relatively little impact on women. To be sure, the percentage of women who change their names might be even higher if they were defaulted into doing so. Nonetheless, it is revealing that most married women in the United States reject the default.

Why does the default rule not stick for American women? Four factors seem to be important. First and most important, many women strongly want to change their names, and their desire is not unclear. This is not a complex or unfamiliar area in which the default rule helps to construct people's preferences, in which people have vague or ambiguous preferences, or in which people have to work to ascertain their preferences. True, many women are undoubtedly affected by social norms, which some of them may wish to be otherwise (Lessig, 1995); but with those norms in place, their preference is not unclear. When a social norm is strong, it may overwhelm the effect of the legal default rule; in fact, it might operate as the functional equivalent of such a rule – a point with general implications (Huh *et al.*, 2014).

Second, the issue is highly salient to married women. It is not exactly in the background. Because marriage is a defined and defining event, the timing of the required action is relatively clear. Procrastination and inertia are therefore less important; the effort tax is well worth incurring.

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

Third, the change of name is, for some or many of those who do it, a kind of celebration. It is not the sort of activity that most women seek to defer, or see as an obligation or as a way of helping their future selves. If people affirmatively like to choose – if choosing is fun or meaningful – a supposed "effort tax" from the default rule is nothing of the sort. Its sign changes; it may even be a kind of "effort subsidy." Sometimes choosing is a benefit rather than a burden.

Fourth, keeping one's name can impose costs, especially (but not only) if one has children. If a wife has a different name from her husband, or vice versa, it might be necessary to offer explanations and to dispel confusion. With some private and public institutions, offering those explanations might be burdensome and time-consuming. For some women, life is made more difficult if they do not have the same name as their husband. Social practices thus create a strong incentive to overcome the default. When the relevant conditions are met – strong preferences, clear timing, positive feelings about opting in, and greater ease and simplicity from opting in – the default rule is unlikely to matter much.

For present purposes, the central point is that strong preferences are likely to be sufficient to ensure that the default rule will not stick. In such cases, inertia will be overcome. People will not be much moved by any suggestion that might be reflected in the default rule (and in the particular context of marital names, the default probably offers no such suggestion) (Ditto & Tannenbaum, n.d.). Loss aversion will be far less relevant, in part because the clear preference and accompanying social norms, rather than the default rule, define the reference point from which losses are measured.

Consider four other examples, reflecting the importance of strong antecedent preferences:

1. A study in the United Kingdom found that most people opted out of a savings plan with an unusually high default contribution rate (12% of pretax income) (Beshears *et al.*, 2010). Only about 25% of employees remained at that rate after a year, whereas about 60% of employees shifted to a lower default contribution rate. The default contribution rate was not entirely ineffective (25% is a large number), but it was far less effective than it would have been if the savings plan had a default contribution rate that fit more closely with people's preferences.
2. Many workers opt out if a significant fraction of their tax refund is defaulted into U.S. savings bonds. In large numbers, they reject the default, apparently because they have definite plans to spend their refunds and do not have much interest in putting their tax refunds into savings bonds (Bronchetti *et al.*, 2011). Their preferences are strong, and they predate the default rule.
3. Consistent with the standard finding of the effectiveness of defaults, a change in the default thermostat setting had a major effect on employees at the

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

Organization for Economic Co-operation and Development (OECD) (Barascud *et al.*, 2013). During winter, a 1°C decrease in the default caused a significant reduction in the average setting. But when choice architects reduced the default setting by 2°C, *the reduction in the average setting was actually smaller*, apparently because sufficient numbers of employees thought that it was too cold, and returned the setting to the one that they preferred. In the face of clear discomfort, inertia is overcome. From this finding, we might venture the following hypothesis, taken very broadly: people will be inclined to change a default rule if it makes them uncomfortably cold. To be sure, strong social norms or feelings of conscience (Manganari & Theotokis, 2014) might counteract this effect – but they will have to be quite strong.

4. A great deal of work shows that the placement of food in cafeterias and grocery stores can have a large effect on what people choose (Wansink, 2013), but there are limits to how much the placement of food can influence choices (de Wijk *et al.*, 2016). Rene A. de Wijk *et al.* (2016) sought to increase consumption of whole-grain bread, which is generally healthier than other kinds of bread. For several weeks, they placed the whole-grain bread at the entrance to the bread aisle (the most visible location) and for different weeks, they placed it at the aisle's exit (the least visible location). The behavioral prediction is that when whole-grain bread is more visible, more people will buy it. But there was no such effect. Whole-grain bread accounted for about a third of the bread sold – and it did not matter whether it was encountered first or last. As the authors suggest, the best explanation for this finding is that people know what bread they like, and they will choose it, whatever the architecture of the supermarket. Strong antecedent preferences trump the effect of the default rule. Note, too, the finding that while school children could well be nudged (through the functional equivalent of default rules) into healthier choices, researchers were not able to counteract the children's strong preference for (unhealthy) French fries (Just & Wansink, 2009).

In all of these cases, it is a useful oversimplification to hypothesize that choosers are making a rational decision about whether to depart from a default, based on the costs of decisions and the costs of errors. When a preference does not exist at all, or when it is weak, choosers will accept the default rule or decline to focus on it; attention is a limited resource, and people ought not to use it when they have no reason to do so. When a preference does not exist at all, or when it is weak, the informational signal contained in the default rule justifiably carries weight. But when choosers have a strong contrary preference, the cost of making the relevant decision is lower (because people already know what they think) and the cost of sticking with the

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

default is higher (because people know that it points them in the wrong direction).

Of course, any such analysis of costs and benefits is often intuitive and automatic rather than deliberative and reflective, and it might well involve heuristics and biases. For many choosers, inertia might well have more force than it should on the basis of a fully rational weighing of costs and benefits. Moreover, recent events in someone's life – involving, for example, a lack of available income in a certain period, or a bad experience with a certain food – might trigger the availability heuristic and lead people wrongly to reject a default or any kind of choice architecture. The point is not that the decision to reject a default reflects an accurate calculation, but that people may make either an intuitive (and fast) or a deliberative (and slow) judgment about whether to reject nudges.

Counternudges: prompting people to opt out

Suppose that self-interested actors have a strong incentive to convince people (say, their customers) to opt in or out. If so, they might be able to take effective steps to achieve their goals. They might be able to persuade people to choose and thus to overcome the default, rendering it ineffective (Willis, 2014). They might "phish" people, whom they consider "phools" (Akerlof & Shiller, 2015). They undertake effective counternudges.

Consider the regulatory effort in 2010 by the Federal Reserve Board to protect consumers from high bank overdraft fees (Requirements for Overdraft Services, 2010). To provide that protection, the Board did not impose any mandate, but instead regulated the default rule. It said that banks could not automatically enroll people in overdraft "protection" programs; instead, customers had to sign up to them. More specifically, the Board's regulation forbids banks from charging a fee for overdrafts from checking accounts unless the account holder has explicitly enrolled in the bank's overdraft program (Willis, 2012). One of the goals of the non-enrollment default rule is to protect customers, and especially low-income customers, from taking the equivalent of extraordinarily high-interest loans, such as loans with interest rates of up to 7000%. The central idea is that many people end up paying large fees essentially by inadvertence. If the default rule is switched so that consumers end up in the program only if they really want to be involved, then they will benefit from a safeguard against excessive charges.

In principle, the regulation should have had a very large effect, and indeed, an understanding of the power of default rules helped to motivate its promulgation. The Board explicitly observed that "studies have suggested [that] consumers are likely to adhere to the established default rule, that is, the outcome

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

that would apply if the consumer takes no action." The Board also referred to research on the power of automatic enrollment to increase participation in retirement savings plans. It emphasized the phenomenon of unrealistic optimism, suggesting that consumers might well think, unrealistically, that they would not overdraw their accounts. No one argues that a default rule can entirely cure the problem of unrealistic optimism, but it can provide a remedy against its most serious harmful effects, at least if the default is sticky.

As Lauren Willis (2012) shows in an important article, the effect of the regulation has not been nearly as large as might have been expected. The reason is that people are opting into the program, and thus rejecting the non-enrollment default, in large numbers. The precise figures remain unclear, but the overall level of opting in seems to be around 15%, and at some banks, it is as high as 60%. Here is the most striking finding: among people who exceed the amount in their checking account more than 10 times per month, the level appears to be over 50%.

A central reason is that many banks *want* to be able to charge overdraft fees and hence use a number of behaviorally informed strategies to facilitate opting in. For those who believe that opting in is a bad idea for many or most customers, this is a clear case of phishing. As Willis demonstrates, banks have taken steps to make opting in as easy as possible – for example, simply by pushing a button on an ATM. They have also engaged in active marketing and created economic incentives in order to persuade people to opt in. They have cleverly exploited people's belief, which is often inaccurate, that it is costly not to be enrolled in the program. For example, they have sent materials explaining, "You can protect yourself from … fees normally charged to you by merchants for returned items," and "The Bounce Overdraft Program was designed to protect you from the cost … of having your transactions denied." They have sent their customers a great deal of material to persuade them that enrollment is in their economic interest.

Consider the following excerpt from one bank's marketing materials, explicitly exploiting loss aversion (Willis, 2012):

> Yes: Keep my account working the same with Shareplus ATM and debit card overdraft coverage.
> No: Change my account to remove Shareplus ATM and debit card overdraft coverage.

There is a large contrast here with the retirement context, in which providers enthusiastically endorse automatic enrollment and have no interest in getting people to opt out. Those who run retirement plans are quite happy if more people are participating, and hence they are glad to work with employers, or the government, to promote enrollment. By contrast, the Federal Reserve

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

Board called for a default that banks dislike, and at least to some extent, the banks have had their revenge. In the savings context, the default rule would not be nearly as sticky if providers wanted people to opt out.

From this illuminating tale, we can draw a general lesson, which is that if regulated institutions are strongly opposed to the default rule and have easy access to their customers, they may well be able to use a variety of strategies, including behavioral ones, to encourage people to move in their preferred directions – and thus to abandon the default. In such cases, the default is ineffective not because choosers independently dislike it, but because companies and firms convince them to choose to reject it. Here, too, it is useful to see people as making a slow or fast judgment about the costs of decisions and the costs of errors, with the important twist that the targets of the default (for the most part, affected companies) work to decrease the (actual and perceived) costs of decisions, and to increase the perceived costs of errors, in order to render the nudge ineffective.

Consider the domain of privacy, where many people also favor a particular default rule, which would safeguard privacy unless people voluntarily chose to give it up (Willis, 2012). If companies want people to relinquish privacy, they might well be able to convince them to do so – perhaps by using tactics akin to those of financial institutions in the Federal Reserve case, or perhaps by denying people services or website access unless they waive their privacy rights (Willis, 2012). Willis herself gives the example of The Netherlands, which created a "Don't-Track-Me" default. As she explains, "Virtually all firms (and even nonprofits) responded to the law by requiring consumers to opt out as a condition of using the firm's website, not by competing on promises not to track" (Willis, 2012).

The general point is plain: when default rules or other forms of choice architecture are ineffective, it is often because self-interested actors have the incentive and the opportunity to impose some kind of counternudge, leading people to choose in their preferred way.

What should be done?

If a default rule proves ineffective, there are three possible responses. The first is to rest content on the grounds that freedom worked. The second is to learn from the failure and to try a different kind of nudge, with continuing testing. The third is to take stronger measures – counter-counternudges, or even mandates and bans – on the grounds that freedom caused a problem.

As we have seen, evaluation of these responses depends on normative considerations; the fact that a nudge has proved insufficiently effective is not, by itself, decisive about what should be done. To be sure, ineffectiveness can be taken to

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

suggest that, on welfare grounds, there is no problem at all; people are going their own way, and that is fine. But many nudges are not controversial; they are designed (for example) to encourage people to take very low-cost steps to improve their lives, and if they fail, people's lives will be worse (Thaler, 2016). If choice architects have good reason to think that people's choices are not promoting their welfare, it is worth considering alternatives, and if the interests of third parties are at stake, the ineffectiveness of nudging may be a reason to consider mandates and bans.

One clarification is necessary before we begin: it is important to consider the possibility of heterogeneous responses within the affected population, and the aggregate effect may tell us far less than we need to know. In these instances, sub-analyses can reveal that the nudges are highly effective on distinct subpopulations, during distinct time periods, or in specific contexts. Suppose, for example, that calorie labels have little or no effect on a large population. Even if this is so, it might be effective, in a relevant sense, if it has a significant impact on people with serious weight problems. Or suppose that 40% of people opt out of a default rule – say, a rule calling for enrollment in pension or wellness plans. Those who opt out may be precisely the group that ought to opt out, given their preferences and situations. Alternatively, the aggregate numbers may disguise a serious problem, as when the population that is unaffected by a nudge (say, by a warning or a reminder) includes those who would most benefit from it.

Freedom worked

Suppose that default rules fail to stick because people do not like them. In such cases, the fact that they are mere defaults, rather than mandates, might be both good and important. Any default rule might be ill-chosen, or it might not fit individual circumstances. If so, the fact that people can reject it is a valuable safeguard. In the case of very large default contribution rates, that claim seems quite plausible. Or consider a situation in which an institute adopts a default rule in favor of double-sided printing and many people switch to single-sided printing. If the latter fits their situations, there is no problem.

Something similar might well be said if and when self-interested institutions that are burdened by a default rule use counternudges to convince people to reject it. The tale of overdraft protection seems to be one of regulatory failure or of at least incomplete success, but things are not so clear. Recall that many people (perhaps as high as 85%) do not opt in to the program. Recall, too, that the largest proportion of people who do opt in are those who actually go over their checking limits. For such people, it is not implausible to think that opting in is a good idea. At least some of them, or perhaps the

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

majority, might well be rational to make that choice. If they cannot borrow from their bank – and overdraft protection is a form of borrowing – they might have to borrow from someone else, which would mean a level of inconvenience and at least potentially even higher interest rates. If so, many people might have to resort to payday lenders, whose rates may or may not be lower.

Because inconvenience can be a real problem and because higher rates might hit people especially hard, overdraft protection might well be in the interest of many or most of the people who end up opting in. Note in this regard that state-level regulation of payday lenders has led consumers to have to resort to equally expensive sources of credit (Homonoff, 2013). This finding strongly suggests that if people cannot use overdraft protection, they might simply go elsewhere.

With this point in mind, we might even think that the Federal Reserve's policy has been a significant success. People are no longer automatically enrolled in overdraft protection, and the vast majority of customers no longer have such protection, which may well be saving them money. At the same time, those who want such protection, or need it, have signed up for it. That sounds like an all-things-considered success for nudging. Is there really a problem? That question can be asked whenever institutions succeed in convincing people to opt out of a default rule. Of course, such institutions might be self-interested, but they might also be producing mutually advantageous deals.

The same points might be made if people reject a default rule in favor of protection of privacy in online behavior. Perhaps that form of privacy does not much matter to people. Perhaps those who want them to waive it can offer something to make doing that worth their while. A default rule gives people a kind of entitlement, and entitlements have prices. If people are willing to give up an entitlement for what seems to be a satisfactory price, where is the objection? A counternudge might be quite welcome.

Nudge better

Suppose that (for example) information disclosure fails to change behavior. If choice architects have learned that fact, they have gained something important: knowledge (Halpern, 2015). A primary goal of the enterprise is to test hypotheses. If one hypothesis turns out to be false, that is a victory.

Again, the failure of the hypothesis does not, by itself, tell us whether something else should be done. Perhaps people are given a warning about the risks associated with some anti-cancer drug; perhaps they act in spite of the warning. If so, there might be no problem at all. The goal of the warning is to ensure that choices are informed, not that they are different. If people's behavior does not

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

change after they receive information about the risks associated with certain activities (say, football or boxing), nothing might have gone wrong.

Suppose, however, that the underlying problem is significant. Perhaps people are not applying for benefits for which an application would almost certainly be in their interest. Perhaps they are failing to engage in behavior that would much improve their economic situation or their health (say, taking prescription medicine or seeing a doctor) (Executive Office of the President, 2015). If so, then other nudges might be tried and tested instead – for example, a clearer warning, uses of social norms, or a default rule. By itself, information might not trigger people's attention, and some other approach might be better, and if a default rule fails, it might make sense to accompany it with information or with warnings. There may well be no alternative but to experiment and to test (Halpern, 2015). One general lesson is that if the goal is to change behavior, choice architects should "make it easy" (Halpern, 2015); we might say that in the case of an ineffective nudge, the next step is "make it easier." Another lesson is that people's acts often depend on their social meaning, which can be a subsidy or a tax; if a nudge affects meaning, it can change a subsidy into a tax, or vice versa (Lessig, 1995).

Alternatively, the particular nudge might fall in the right general category, but it might need to be refined. Information disclosure might be ineffective if it is complex, but succeed if it is simple. Social norms might move behavior, but only if they are the norms in the particular community, not in the nation as a whole (Halpern, 2015). A reminder might fail if it is long and poorly worded, but succeed if it is short and simple.

Freedom failed

Altering rules, framing rules, and counter-counternudges

In some cases, freedom of choice itself might be an ambiguous good. For behavioral or other reasons, an apparently welcome and highly effective counternudge might turn out to be welfare-reducing – a form of "phishing." People might suffer from present bias, optimistic bias, or a problem of self-control (Bar-Gill, 2012). The counternudge might exploit a behavioral bias of some kind (Bar-Gill, 2012; Ru & Schoar, 2016). What might be necessary is some kind of counter-counternudge – for example, a reminder or a warning to discourage people from opting in to a program that is generally not in their interest.

In the case of overdraft protection programs, some of those who opt in, and who end up receiving that protection, are probably worse off as a result. Perhaps they do not understand the program and its costs; perhaps they were duped by a behaviorally informed messaging campaign. Perhaps they

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

are at risk of overdrawing their accounts, not because they need a loan, but because they have not focused on those accounts, and on how they are about to go over. Perhaps they are insufficiently informed or attentive. To evaluate the existing situation, we need to know a great deal about the population of people who opt in. In fact, this is often the key question, and it is an empirical one. The fact that they have opted in is not decisive.

The example can be taken as illustrative. If a default rule or some other nudge is well-designed to protect people from their own mistakes, and if it does not stick, then its failure to do so is nothing to celebrate. The fact of its ineffectiveness is a tribute to the success of a self-interested actor seeking to "phish" for "phools" (Akerlof & Shiller, 2015). The counternudge is a form of manipulation or exploitation, something to counteract rather than to celebrate (Ru & Schoar, 2016).

The same point applies to strong antecedent preferences, which might be based on mistakes of one or another kind. A GPS device is a defining nudge, and if people reject the indicated route on the grounds that they know better, they might end up getting lost. The general point is that if the decision to opt out is a blunder for many or most, then there is an argument for a more aggressive approach. The overdraft example demonstrates the importance of focusing not only on default rules, but also on two other kinds of rules as well, operating as counter-counternudges: *altering rules* and *framing rules* (Willis, 2012).

Altering rules establish how people can change the default. If choice architects want to simplify people's decisions, and if they lack confidence about whether a default is suitable for everyone, they might say that consumers can opt in, or opt out, by making an easy phone call (good) or by sending a quick email (even better). Alternatively, choice architects, confident that the default is right for the overwhelming majority of people, might increase the costs of departing from it. For example, they might require people to fill out complex forms or impose a cooling-off period. They might also say that even if people make a change, the outcome will "revert" to the default after a certain period (say, a year), requiring repeated steps. Or they might require some form of education or training, insisting on a measure of learning before people depart from the default.

Framing rules establish and regulate the kinds of "frames" that can be used when trying to convince people to opt in or opt out. We have seen that financial institutions enlisted loss aversion in support of opting in. Behaviorally informed strategies of this kind could turn out to be highly effective. But that is a potential problem. Even if they are not technically deceptive, they might count as manipulative, and they might prove harmful. Those who believe in freedom of choice, but seek to avoid manipulation or harm, might want to

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

constrain the permissible set of frames – subject, of course, to existing safeguards for freedom of speech. Framing rules might be used to reduce the risk of manipulation.

Consider an analogy: if a company says that its product is "90% fat-free," people are likely to be drawn to it, far more so than if the company says that its product is "10% fat." The two phrases mean the same thing, and the "90% fat-free" frame is legitimately seen as a form of manipulation. In 2011, the American government allowed companies to say that their products are 90% fat-free – but only if they also say that they are 10% fat. We could imagine similar constraints on misleading or manipulative frames that are aimed at getting people to opt out of the default. Alternatively, choice architects might use behaviorally informed strategies of their own, supplementing a default rule with (for example) uses of loss aversion or social norms to magnify its impact (Keller *et al.*, 2011).

To the extent that choice architects are in the business of choosing among altering rules and framing rules, they can take steps to make default rules more likely to stick, even if they do not impose mandates. They might conclude that mandates and prohibitions would be a terrible idea, but that it makes sense to make it harder for people to depart from default rules. Sometimes that is the right conclusion. The problem is that when choice architects move in this direction, they lose some of the advantages of default rules, which have the virtue of easy reversibility, at least in principle. If the altering rules are made sufficiently onerous, the default rule might not be all that different from a mandate.

There is another possibility: choice architects might venture a more personalized approach. They might learn that one default rule suits one group of people, and that another suits a different group; by tailoring default rules to diverse situations, they might have a larger effect than they would with a "mass" default rule (Porat & Strahilevitz, 2014). Or they might learn that an identifiable subgroup is opting out, either for good reasons or for bad ones. (Recall that aggregate effectiveness data might disguise very large effects or very small ones for relevant subgroups.) If the reasons do not seem good, choice architects might adopt altering rules or framing rules as safeguards, or they might enlist (for example) information and warnings. If they can be made to work well, more personalized approaches have the promise of preserving freedom of choice while simultaneously increasing effectiveness.

Incentives, mandates, and bans

However, preserving freedom of choice might not be a good idea. Indeed, we can easily imagine cases in which a mandate or ban might be justified on

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

behavioral or other grounds (Conly, 2013; Bubb & Pildes, 2014). Most democratic nations have mandatory social security systems, based in part on a belief that "present bias" is a potent force and a conclusion that some level of compulsory savings is justified on welfare grounds. Food safety regulations forbid people from buying goods that pose risks that reasonable people would not run. Such regulations might be rooted in a belief that consumers lack relevant information (and it is too difficult or costly to provide it to them), or they might be rooted in a belief that people suffer from limited attention or optimistic bias. Some medicines are not allowed to enter the market, and for many others, a prescription is required; people are not permitted to purchase them on their own (Conly, 2013).

Many occupational safety and health regulations ultimately have a paternalistic justification, and they take the form of mandates and bans, not nudges. In the domains of fuel economy and energy efficiency, existing requirements are difficult to defend without resort to behavioral arguments (Allcott & Sunstein, 2015). To be sure, they do reduce externalities in the form of conventional air pollutants, greenhouse gases, and energy insecurity, but if we consider only those externalities, the benefits of those requirements are usually lower than the costs. The vast majority of the monetized benefits accrue to consumers, in the form of reduced costs of gasoline and energy. On standard economic grounds, those reduced costs should not count, because consumers can obtain them through their own free choices; if they are not doing so, it must be because the relevant goods (e.g. automobiles, refrigerators) are inferior along some dimension. The U.S. government's current response is behavioral; it is that in the domain of fuel economy and energy efficiency, consumers are making some kind of mistake, perhaps because of present bias, perhaps because of a lack of sufficient salience. The argument is of course contested (Allcott & Sunstein, 2015), but if it is correct, the argument for some kind of mandate is secure on welfare grounds.

The same analysis holds, and more simply, if the interests of third parties are involved. Default rules are typically meant to protect choosers, but in some cases, third parties are the real concern. For example, a green default rule, designed to prevent environmental harm, is meant to reduce externalities and to solve a collective action problem, not to protect choosers as such (Ullmann-Margalit, 1977). A nudge, in the form of information disclosure or a default rule, is not the preferred approach to pollution (including carbon emissions). If a nudge is to be used, it is because it is a complement to more aggressive approaches, or because such approaches are not feasible (perhaps for political reasons). But if a default rule proves ineffective – for one or another of the reasons sketched here – there will be a strong argument for economic incentives, mandates, and bans (Sunstein & Reisch, 2013).

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

Conclusion and further considerations

Default rules are often thought to be the most effective nudges, but for two reasons, they might not have the expected impact. The first involves strong antecedent preferences. The second involves the use counternudges by those with an economic or other interest in convincing choosers to opt out.

These two reasons help account for the potential ineffectiveness of nudges in many contexts. Information, warnings, and reminders will not work if people are determined to engage in the underlying behavior (e.g. smoking, drinking, texting while driving, eating unhealthy foods). If (for example) cigarette companies and sellers of alcoholic beverages have opportunities to engage choosers, they might be able to weaken or undo the effects of information, warnings, and reminders (Akerlof & Shiller, 2015).

It is important to observe that nudges may be ineffective for independent reasons. Consider the following five reasons.

(1) If a nudge is based on a plausible but inaccurate understanding of behavior, and of the kinds of things to which people respond, it might have no impact. This is an especially important point, and it suggests the immense importance of testing apparently reasonable behavioral hypotheses. We might believe, for example, that because people are loss averse, a warning about potential losses will change behavior. But if such a warning frightens people or makes them think that they cannot engage in self-help, and essentially freezes them, then we might see little or no change (Sharot, 2017). We might believe that people are not applying for important services because of excessive complexity, and that simplification will make a large difference. But perhaps it will not; skepticism, fear, or inertia might be the problem, and simplification might not much help. Or we might hypothesize that an understanding of social norms will have a large effect on what people do. But if the target audience is indifferent to (general) social norms, and is happy to defy them, then use of social norms might have no impact. Some nudges seem promising because of their effects on small or perhaps idiosyncratic populations. Whether their impact will diminish, or be eliminated, when applied elsewhere is a question that needs to be tested. Of course, it is true that the failure of a behavioral hypothesis should pave the way to alternative or more refined hypotheses, including a specification of the circumstances in which the original hypothesis will or will not hold (Griskevicius *et al.*, 2012).

(2) If information is confusing or complex to process, people might be unaffected by it. There is a lively debate regarding the general effectiveness of disclosure strategies; some people are quite skeptical (Griskevicius *et al.*, 2012; Bubb, 2015). What seems clear is that the design of such strategies is

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

extremely important. Disclosure nudges, or educative nudges in general, may have far less impact than one might think in the abstract (Willis, 2011).

(3) People might show *reactance* to some nudges, rejecting an official effort to steer *because* it is an official effort to steer (Arad & Rubinstein, 2015; Duncan *et al.*, 2016; Hedlin & Sunstein, 2016). Most work on the subject of reactance explores how people rebel against mandates and bans, because they wish to maintain control (Brehm & Brehm, 1981; Pavey & Sparks, 2009). Default rules are not mandates, and hence it might be expected that reactance would be a nonissue. But as for mandates, so for defaults: they might prove ineffective if people are angry or resentful that they have been subjected to it. So, too, an effort to invoke social norms might not work if people do not care about social norms, or if they want to defy them. We are at the early stages of learning about the relationship between reactance and nudges, and thus far, it seems safe to say that for the most part, reactance is not likely to be an issue, simply because autonomy is preserved. But there are some interesting qualifications, and in some cases, they might prove to be important.

(4) Nudges might have *only a short-term effect* (Frey & Rogers, 2014). If people see a single reminder, they might pay attention to it – but only once. If people receive information about health risks, their behavior might be influenced; but after a time, that information might become something like background noise or furniture. It might cease to be salient or meaningful. Even a graphic warning might lose its resonance after a time. By contrast, a default rule is more likely to have a persistent effect, because people have to work to change it; but after a while, its informational signal might become muted, or inertia might be overcome. We continue to learn about the circumstances in which a nudge is likely to produce long-term effects (Allcott & Rogers, 2014).

(5) Some nudges might have an influence on the desired conduct, but also produce *compensating behavior, nullifying the overall effect*. Suppose, for example, that a cafeteria design works to induce high school students to eat healthy foods. Suppose, too, that such students eat unhealthy foods at snack time, at dinner, or after school hours. If so, the nudge will not improve public health. There is a general risk of a "rebound effect" – as, for example, when fuel-efficient cars lead people to drive more, reducing and potentially nullifying the effects of interventions designed to increase fuel efficiency. Perhaps a nudge will encourage people to exercise more than they now do, but perhaps they will react by eating more.

The general point is that any form of choice architecture, including the use of default rules, may have little or no net effect if people are able to find other domains in which to counteract it. The idea of compensating behavior can be

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

seen as a subset of the general category of strong antecedent preferences, but it points to a more specific case, in which the apparent success of the nudge is an illusion in terms of what choice architects actually care about (Hirschman, 1991).

What matters is welfare, not effectiveness (Sunstein, 2016). A largely ineffective nudge may have positive welfare effects; an effective nudge might turn out to reduce welfare. A strong reason for nudges, as distinguished from more aggressive tools, is that they preserve freedom of choice and thus allow people to go their own way. In many contexts, that is indeed a virtue, and the ineffectiveness of nudges, for some or many, is nothing to lament. But when choosers are making clear errors, and when third-party effects are involved, the ineffectiveness of nudges provides a good reason to consider stronger measures on welfare grounds.

Acknowledgements

I am grateful to Hunt Allcott, Maya Bar-Hillel, George Loewenstein, Eric Maskin, Eric Posner, Lucia Reisch, and Maya Shankar for valuable comments on a previous draft, and to the Harvard Program on Behavioral Economics and Public Policy for support. In a few places, this article draws on previous work, above all Sunstein (2015).

References

Adler, M. (2011), *Welfare and Fair Distribution: Beyond Cost-Benefit Analysis*, New York, NY: Oxford University Press.

Akerlof, G. and R. Shiller (2015), *Phishing for Phools*, Princeton, NJ: Princeton University Press.

Alemanno, A. and A. L. Sibony (eds) (2016), *Nudge and the Law: A European Perspective*, Oxford: Hart Publishing.

Allcott, H. (2011), 'Social norms and energy conservation', *Journal of Public Economics*, **95**(9–10): 1082–1095.

Allcott, H. and T. Rogers (2014), 'The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation', *American Economic Review*, **104**(10): 3003–3037.

Allcott, H. and C. Sunstein (2015), 'Regulating Internalities', *Journal of Policy Analysis and Management*, **34**: 698–705.

Arad, A. and A. Rubinstein (2015), The People's Perspective on Libertarian-Paternalistic Policies, Unpublished manuscript, Retrieved from http://arielrubinstein.tau.ac.il/papers/LP.pdf.

Balz, J., C. Sunstein and R. Thaler (2013), Choice Architecture and Retirement Savings Plans, in: E. Shafir (ed.), *The Behavioral Foundations of Policy* (428–439), Princeton, NJ: Princeton University Press.

Bar-Gill, O. (2012), *Seduction By Contract*, Oxford, United Kingdom: Oxford University Press.

Barascud, F., Z. Brown, N. Johnstone, I. Hascic and L. Vong (2013), 'Testing the Effects of Defaults on the Thermostat Settings of OECD Employees', *Energy Economics*, **39**: 128–134.

Benartzi, S. and R. Thaler (2013), 'Behavioral Economics and the Retirement Savings Crisis', *Science*, **339**: 1152–1153.

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

Beshears, J., J. Choi, D. Liabson and B. Madrain (2010), The Limitations of Defaults, Unpublished manuscript, Retrieved from http://www.nber.org/programs/ag/rrc/NB10-02,%20Beshears,%20Choi,%20Laibson,%20Madrian.pdf

Bettinger, E., B. Long, P. Oreopoulos and L. Sanbonmatsu (2009), The Role of Simplification and Information in College Decisions: Results from the H&R Block FAFSA Experience (NBER Working Paper No. 15361), Retrieved from http://www.nber.org/papers/w15361.

Brehm, J. and S. Brehm (1981), *Psychological Reactance: A Theory of Freedom and Control*, New York, NY: Academic Press.

Bronchetti, E., T. Dee, D. Huffman and E. Magenheim (2011), When a Nudge Isn't Enough: Defaults and Saving Among Low-Income Tax Filers 28-29 (NBER, Working Paper No. 16887), Retrieved from http://www.nber.org/papers/w16887

Bubb, R. (2015), 'TMI? Why the Optimal Architecture of Disclosure Remains TBD', *Michigan Law Review*, **113**(6): 1021–1042.

Bubb, R. and R. Pildes (2014), 'How Behavioral Economics Trims Its Sails and Why', *Harvard Law Review*, **127**(6): 1593–1678.

Cain, D., G. Loewenstein and S. Sah (2012), 'The Unintended Consequences of Conflict of Interest Disclosure', *Journal of the American Medical Association*, **307**(7): 669–670.

Carroll, G., J. Choi, D. Liabson, B. Madrian and A. Metrick (2009), 'Optimal Defaults and Active Decisions', *Quarterly Journal of Economics*, **124**(4): 1639–1674.

Chetty, R., J. Friedman, S. Leth-Peterson, T. Nielsen and T. Olsen (2012), Active vs. Passive Decisions and Crowdout in Retirement Savings Accounts: Evidence from Denmark (NBER Working Paper No. 18565), Retrieved from http://www.nber.org/papers/w18565.

Conly, S. (2013), *Against Autonomy: Justifying Coercive Paternalism*, Cambridge, United Kingdom: Cambridge University Press.

Conway, K., A. Gothro, B. Kyler and Q. Moore (2015), Direct Certification in the National School Lunch Program Report to Congress: State Implementation Progress, School Year 2013–2014, Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Department, Office of Policy Support. Retrieved from http://www.fns.usda.gov/direct-certification-national-school-lunch-program-report-congress-state-implementation-progress.

de Wijk, R., N. Holthuysen, A. Maaskant, I. Polet, E. van Kleef and M. Vingerhoeds, (2016), 'An In-Store Experiment on the Effect of Accessibility on Sales of Wholegrain and White Bread in Supermarkets', *PLOS ONE*, **11**(3), Article e0151915. Retrieved from http://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0151915.

Ditto, P. and D. (-) 'Tannenbaum, Information Asymmetries in Default Options. Unpublished manuscript'. Retrieved from https://webfiles.uci.edu/dtannenb/www/documents/default%20information%20asymmetries.pdf.

Downs, J., G. Loewenstein, B. Wanskink and J. Wisdom (2013), 'Supplementing Menu Labeling With Calorie Recommendations to Test for Facilitation Effects', *American Journal of Public Health*, **103**(9): 1604–1609.

Duncan, S., J. Jachimowicz and E. Weber (2016), Default-Switching: The Hidden Cost of Defaults, Unpublished manuscript, Retrieved from http://papers.ssrn.com/sol3/cf_dev/AbsByAuth.cfm?per_id=46495#reg.

Ebeling, F. and S. Lotz (2015), 'Domestic Uptake of Green Energy Promoted By Opt-Out Tariffs', *Nature Climate Change*, **5**: 868–871. Retrieved from http://www.nature.com/nclimate/journal/v5/n9/full/nclimate2681.html?WT.ec_id=NCLIMATE-.

Egebark, J. and M. Ekstrom (2013), 'Can Indifference Make the World Greener?' *Journal of Environmental Economics and Management*, **76**: 1–13.

Emens, E. (2007), 'Changing Name Changing: Framing Rules and the Future of Marital Names'. *University of Chicago Law Review*, **74**(3): 761–863.

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

Executive Order No. 13707, 3 C.F.R. 56365–56367 (2015), Retrieved from https://www.whitehouse.gov/the-press-office/2015/09/15/executive-order-using-behavioral-science-insights-better-serve-american.

Executive Office of the President, National Science and Technology Council (2015), Social and Behavioral Sciences Team, Annual Report, Washington, D.C. Retrieved from https://www.whitehouse.gov/sites/default/files/microsites/ostp/sbst_2015_annual_report_final_9_14_15.pdf.

Fagerlin, A., S. Sah and P. Ubel (2016), 'Effect of Physician Disclosure of Specialty Bias on Patient Trust and Treatment Choice, PNAS: Proceedings of the National Academy of Sciences', 113(27), 7465–7469. Retrieved from http://www.pnas.org/content/early/2016/06/16/1604908113.full.pdf.

Finkelstein, S., M. Liersch and C. McKenzie (2006), 'Recommendations Implicit in Policy Defaults', *Psychological Science*, **17**(5): 414–420.

Frey, E. and T. Rogers (2014). 'Persistence: How Treatment Effects Persist After Interventions Stop', *Insights from Behavioral and Brain Sciences*, **1**: 172–179.

Golman, R., G. Loewenstein and C. Sunstein (2014). 'Disclosure: Psychology Changes Everything', *Annual Review of Economics* **6**: 391–419.

Griskevicius, V., D. T. Kenrick, Y. J. Li and S. L. Neuberg (2012), 'Economic Decision Biases and Fundamental Motivations: How Mating and Self-Protection Alter Loss Aversion', *Journal of Personality and Social Psychology*, **102**(3): 550–561.

Halpern,, D. (2015), *Inside the Nudge Unit*. London, United Kingdom: WH Allen.

Hedlin, S. and C. Sunstein (2016), 'Does Active Choosing Promote Green Energy Use? Experimental Evidence', *Ecology Law Quarterly*, **43**(1): 107–142.

Hirschman, A. (1991), *The Rhetoric of Reaction: Perversity, Futility, Jeopardy*. Cambridge, MA: Belknap Press of Harvard University Press.

Homonoff, T. (2013), Essays in Behavioral Economics and Public Policy, Unpublished doctoral dissertation, Princeton University, Retrieved from http://dataspace.princeton.edu/jspui/bitstream/88435/dsp01jw827b79g/1/Homonoff_princeton_0181D_10641.pdf.

Huh, Y. E., C. Mordwedge and J. Vosgerau (2014), 'Social Defaults: Observed Choices Become Choice Defaults', *Journal of Consumer Research*, **41**: 746–760.

Johnson, E. and D. Goldstein (2003), 'Do Defaults Save Lives?' *Science* **302**(5649): 1338–1339.

Johnson, E. and D. Goldstein (2013), Decisions by Default. In: E. Shafir (ed.), *The Behavioral Foundations of Policy* (417–418), Princeton, NJ: Princeton University Press.

Jolls, C. (2015), 'Debiasing Through Law and the First Amendment'. *Stanford Law Review* **67**(6): 1411–1446.

Jones, R., J. Pykett and M. Whitehead (2014), *Changing Behaviors: On the Rise of the Psychological State*, Cheltenham, United Kingdom: Edward Elgar.

Just, D. and B. Wansink (2009), 'Smarter Lunchrooms: Using Behavioural Economics to Improve Meal Selection', *Choices*, **24**(3), Retrieved from http://www.choicesmagazine.org/UserFiles/file/article_87.pdf.

Keller, P., B. Harlam, G. Loewenstein and K. Volpp (2011), 'Enhanced Active Choice: A New Method to Motivate Behavior Change', *Journal of Consumer Psychology*, **21**, 376–383.

Krug, S., (2014), *Don't Make Me Think*, Revisited. San Francisco, CA: New Riders.

Lessig, L. (1995), The Regulation of Social Meaning, *University of Chicago Law Review*, **62**(3): 943–1046.

Madrian, B. and D. Shea (2001), 'The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior', *Quarterly Journal of Economics*, **116**(4): 1149–1187.

Manganari, E. and A. Theotokis (2014), 'The Impact of Choice Architecture on Sustainable Consumer Behavior: The Role of Guilt', *Journal of Business Ethics*, **131**: 423–437.

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press

Manoli, D. and N. Turner (2014), Nudges and Learning: Evidence from Informational Interventions for Low-Income Taxpayers (NBER Working paper No. 20718). Retrieved from http://www.nber.org/papers/w20718.
Mathis, K. and A. Tor (eds.) 2016, *Nudging – Possibilites, Limitations and Applications in European law and Economics*, Switzerland: Springer.
Pavey, L. and P. Sparks (2009), 'Reactance, Autonomy and Paths to Persuasion: Examining Perceptions of Threats to Freedom and Informational Value', *Motivation and Emotion*, **33**(3): 227–290.
Pichert, D. and K. V. Katsikopoulos (2008), 'Green Defaults: Information Presentation and Pro-environmental Behaviour, *Journal of Environmental Psychology*, **28**(1): 63–73.
Porat, A. and L. Strahilevitz (2014), 'Personalizing Default Rules and Disclosure with Big Data', *Michigan Law Review*, **112**(8): 1417–1478.
Requirements for Overdraft Services, 45 C.F.R. 205.17 (2010).
Ru, H. and A. Schoar (2016), Do Credit Card Companies Screen for Behavioral Biases? (NBER Working Paper 22360), Retrieved from http://www.nber.org/papers/w22360.
Sharot, T. (forthcoming 2017), *The Influential Mind*, New York, NY: Henry Holt.
Sunstein, C. and L. Reisch (2013), 'Automatically Green', *Harvard Environmental Law Review*, **38**(1): 127–158.
Sunstein, C. (1996), 'Social Norms and Social Roles', *Columbia Law Review*, **96**(4): 903–968.
Sunstein, C. (2015), *Choosing Not to Choose*, New York, NY: Oxford University Press.
Sunstein, C. (2016), *The Ethics of Influence: Government in the Age of Behavioral Science*, New York, NY: Cambridge University Press.
Sunstein, C. and R. Thaler (2008), *Nudge*, New York, NY: Penguin Books.
Thaler, R. (2015), *Misbehaving*, New York, NY: W.W. Norton & Company, 1–10, 309–345.
Thaler, R. (2016), 'Behavioral Economics: Past, Present, and Future', *American Economic Review*, **106**(7): 1577–1600.
Ullmann-Margalit, E. (1977), *The Emergence of Norms*, Oxford, United Kingdom: Clarendon Press.
Wansink, B. (2013), *Slim By Design: Mindless Eating Solutions for Everyday Life*, New York, NY: William Morrow.
Willis, L. (2011), 'The Financial Education Fallacy', *American Economic Review*, **101**(3): 429–434.
Willis, L. (2012), 'When Nudges Fail: Slippery Defaults', *University of Chicago Law Review* **80**(3): 1155–1229.
Willis, L. (2014), 'Why Not Privacy by Default?', *Berkeley Technology Law Journal*, **29**(1): 61–134.
Zamir, E. (2014), *Law, Psychology, and Morality: The Role of Loss Aversion*, New York, NY: Oxford University Press.

https://doi.org/10.1017/bpp.2016.3 Published online by Cambridge University Press