Working Paper in Economics No. 754

A theoretical framework explaining the mechanisms of nudging

Åsa Löfgren and Katarina Nordblom

Department of Economics, March 2019

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

A theoretical framework explaining the mechanisms of nudging*

Åsa Löfgren† and Katarina Nordblom‡

March 8, 2019

WP 754, Dept of Economics, University of Gothenburg

Abstract

In this paper we develop a theoretical model to clarify the underlying mechanisms that drive individual decision making and responses to behavioral interventions, such as nudges. The contribution of the paper is threefold: First, the model provides a theoretical framework that comprehensively structures the individual decision-making process applicable to a wide range of choice situations. Second, we reduce the confusion regarding what should be called a nudge by offering a clear classification of behavioral interventions. We distinguish among what we label as pure nudges, preference nudges, and other behavioral interventions. Third, we identify the mechanisms behind the effectiveness of behavioral interventions based on the structured decision-making process. Hence, the model can be used to predict under which circumstances, and in which choice situations, a nudge is likely to be effective.

JEL classification: D11, D91. *Keywords:* Nudge, decision making, behavioral intervention

---

*We would like to thank Fredrik Carlsson, Andreas Nilsson, and seminar participants at University of Gothenburg and the CeCAR Workshop 2017 for their valuable comments and suggestions. Financial support from the UGOT Centre for Collective Action Research and MISTRA Carbon Exit is gratefully acknowledged.

†Corresponding author. Department of Economics, University of Gothenburg Box 640, 405 30 Göteborg, Sweden and Center for Collective Action Research (CeCAR) in Gothenburg. E-mail: asa.lofgren@economics.gu.se.

‡Department of Economics, University of Gothenburg, CeCAR and Uppsala Center for Fiscal Studies (UCFS) E-mail: katarina.nordblom@economics.gu.se.

1 Introduction

In 2017 Richard H. Thaler, who, together with Cass R. Sunstein, wrote the book *Nudge* in 2008, was rewarded the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel. The prize boosted an already increasing interest in nudges, and a search on "nudge" on Google Scholar shows a 50% increase in hits during 2009-2017 as compared with 2000-2008, with the majority of the papers having been published during the most recent years. The use of behavioral interventions, i.e. to alter people's behavior without changing economic incentives, has gained significant interest over time among both practitioners and academics. The nudge literature has so far been mainly empirical, and very little theoretical work has been conducted on the issue. One exception is Farhi and Gabaix (2018), who analyze nudges in relation to optimal taxation. While their analysis has a normative focus, we develop a theoretical model to clarify the underlying mechanisms that drive individual decision making and responses to behavioral interventions, such as nudges. The contribution of the paper is threefold: First, we provide a theoretical framework that comprehensively structures the individual decision-making process applicable to a wide range of choice situations. Second, we offer a classification of behavioral interventions based on the decision-making process, thereby reducing the confusion regarding what should be considered a nudge and what should not. We suggest a clear distinction among what we label as *pure nudges*, *preference nudges*, and other behavioral interventions. Third, we identify the mechanisms behind the effectiveness of behavioral interventions based on the structured decision-making process. Hence, the model can be used to predict under which circumstances, and in which choice situations, a nudge (or some other kind of behavioral intervention) is likely to be effective. This is critical information both for policy makers who are considering implementing a nudge and for researchers working on the design and evaluation of behavioral interventions. There are also many normative considerations when implementing nudges that have been discussed in the literature on libertarian paternalism (Hausman and Welch, 2010). Even if our analysis is purely positive and does not attempt to provide an answer to whether a nudge *should* be implemented, the model may be used as a basis for normative discussions since it clarifies important aspects of the decision-making process when a nudge is implemented.

Going back to the work by Thaler and Sunstein, the following is perhaps the

most widely recognized definition of a nudge:

> A nudge ... is any aspect of the choice architecture that alters people's behavior in a predictable way without ... significantly changing their economic incentives. (2008, 6)

While at first glance this seems to be a straightforward definition, the devil is in the details, and it is crucial how "without ... significantly changing their economic incentives" is interpreted for determining what constitutes a nudge. Just like Hansen (2016), we classify behavioral interventions as nudges or not based on the decision-making process, rather than the motive behind the intervention. In our model, we make a distinction between *attentive* and *inattentive* choices, and in our strict definition of nudges, we disqualify anything that affects the attentive choice as a nudge. In the terminology of Kahneman (2003), System 1 and System 2 are two cognitive systems guiding individual decision making. They refer to intuition and reasoning, respectively. Applying this in our model, we think of individuals as making their choices in two steps. In the first step (which is likely to be instantaneous), the individual chooses whether System 1 or System 2 should be decisive when making the actual choice in the second step. If System 1 is decisive, we refer to this as an inattentive choice, which is made according to some simplified choice rule and not based on utility maximization. If System 2 is decisive, we instead refer to this as an attentive choice where the individual acts as *Homo economicus*.

Homo economicus bases decisions on economic incentives only and is therefore, by definition, not nudgeable. One important reason most people are different from *Homo economicus* is that decision making and optimization are complicated, and the outcome of a decision is often subject to uncertainty. Hence, making a well-informed, utility-maximizing choice often comes at a cognitive cost. To reduce or avoid this cost, individuals may use simplified choice rules such as basing decisions on intuition or habit, choosing the first or last option on a list (Murphy et al., 2006), choosing a pre-selected option, or even a random draw to determine the choice. Salant (2011) shows that any choice rule used to process information that is less complicated than utility maximization is subject to framing effects. We argue that it is when we make this kind of inattentive or intuitive choice instead of acting as *Homo economicus* that we are nudgeable. Or, as Thaler and Sunstein express it:

> ... a nudge is any factor that significantly alters the behavior of Humans although it would be ignored by Econs. (2008, 9)

Think of the case with organ donations. Making the choice whether to donate one's organs may give rise to emotions and existential questions that are uncomfortable to some. Therefore, one may might prefer to make no choice at all. In such a case, the default option becomes crucial. Changing the default option from no donation to donation has been shown to have a huge effect on the decision to donate (see, e.g., Johnson and Goldstein, 2003; Abadie and Gay, 2006; Li et al., 2013).

Another, completely different decision is the one we make every day about what to have for lunch. That decision rarely gives rise to the kind of dilemma connected with organ donations, but one still may rationalize in order to reduce the cost of decision making. One may, for example, always choose the first dish that is listed on the menu. This choice rule also reduces the cognitive cost of making the decision (although of another magnitude than in the organ donation case). In a recent experiment, changing the order on the menu (from having a meat dish on top to a vegetarian option first) turned out to increase the share of vegetarian meals significantly (Kurz, 2017). When the nudge was removed, however, behavior returned to the prenudge state.

To see how we reason about the two steps in the decision-making process, think of choosing a snack when the available options are an apple and a candy bar. *Ex ante*, before you are put in the choice situation, you may list preferential attributes that are important to your choice of snack, such as price, nutrition and taste. When you are about to make your choice, it is made in two steps. In the first step, you decide whether to gather all information about price, nutrition, and taste or whether you should simply pick one of the options—that is whether you should let System 2 make an informed and rational choice or whether System 1 should just make an inattentive choice without any reasoning. If you go for the former, you will choose the preferred option in the second step, but this requires you to exert some effort. If you do not think it is worth the effort, you just pick one option without thought. It is in this situation that you are nudgeable. If the apple is put at eye level and the candy bar close to the floor, you may be nudged to choose the apple if you just pick the option closest to you. This is what we refer to as a *pure nudge*. A *preference nudge*, on the other hand could be a notice stating "The apple is better

for you". However, whether that counts as a nudge or not may differ across people. If nutrition is an important attribute and you are not really aware of the relative nutrition, you may learn something from the notice, which we would not consider a nudge, but rather information. Information is not a nudge, as it would also affect the attentive choice. If the notice ("The apple is better for you") instead reminds you of something you already know (it is healthier to eat an apple than a candy bar), we classify it as a *preference nudge*. The nudge is not preference irrelevant as is the *pure* nudge, but the choice is still made inattentively. Our model also indicates when a nudge is likely to fail. If you have a strong habit of eating chocolate, you will inattentively choose the chocolate disregarding where it is placed and what notices may be displayed.

In Section 2, we present in detail how the first step in the decision-making process is made, and in Section 3, we specifically study the inattentive choice and the two kinds of nudges. We show that the effectiveness of nudges crucially hinges on the importance and complexity of the choice, and how confident the decision maker is in the particular choice situation. Our predictions are formulated in terms of several claims, which are used in Section 4 to categorize various choice situations and examine the policy implications and in Section 5 to systematically explore a wide range of existing empirical nudge studies. Section 6 concludes.

2 The model

In standard utility theory, individuals are assumed to make an informed choice resulting in the optimal outcome. However, there are numerous examples indicating that humans are not that rational but sometimes make choices without much thought. We suggest a model where optimizing comes at a cognitive cost and the decision consists of two steps (of which at least the first may be instantaneous): first deciding whether it pays off to exert effort to make an informed (*attentive*) choice, and second, making the choice based on the first step—that is, either an attentive or an *inattentive* choice. The inattentive choice is based on some simple choice rule rather than on optimization and may therefore lead to a mistake. This is illustrated in Figure 1. Hence, there is a trade-off between the cost of making an attentive choice and the consequences of a potential mistake from making an inattentive one.

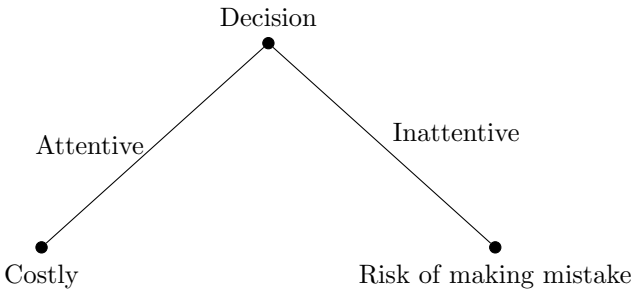

Figure 1: Decision-making process

Consider an individual who should make the dichotomous choice between x and y, but who does not have full information about all relevant attributes of the two alternatives. Hence, the true utilities of the two goods are not known *ex ante*. Making a rational choice—that is, actively optimizing—between x and y is therefore assumed to require some cognitive effort ς. If effort is exerted, the relevant attributes will become known, and the attentive choice will be based on which of the certain utilities $U(x)$ and $U(y)$ is the greatest.[1] Hence, total utility from the attentive choice will *ex post* turn out to be

$$U_{att} = \max\big\{U(x), U(y)\big\} - \varsigma \tag{1}$$

However, since the true utilities are not known without exerting effort, the expected utility from the attentive choice net of the effort cost is *ex ante*

$$E[U_{att}] = E[\max\big\{U(x); U(y)\big\}] - \varsigma. \tag{2}$$

To determine whether it pays off to exert the effort, ς, the expected utility in (2) is compared with the alternative, the expected utility of an inattentive choice, which we denote $E[U_{in}]$. The inattentive choice is made according to some simplified choice rule without exerting effort and may therefore result in a mistake. The consequence of such mistake depends on the importance of the choice. θ is the individual's confidence about the likelihood of making the preferred choice without

[1]That ς is an information cost that eliminates uncertainty is, of course, a simplification. However, the reasoning is analogous if the effort is physical or emotional instead. In our model, we assume that the individual makes a dichotomous choice and either exerts enough effort to enable a fully informed choice as *Homo economicus* or does not exert any effort at all. This is consistent with the notions of Reis (2006) but makes our approach a bit different from that of Gabaix (2014), who lets the agent continuously choose the level of attention. However, qualitatively, our results become very similar.

optimization—that is, picking the option that *ex post* turns out to yield the highest utility. Hence, *ex ante*

$$E[U_{in}] = \theta E[\max\left\{U(x), [U(y)\right\}] + (1-\theta)E[\min\left\{U(x), U(y)\right\}]. \tag{3}$$

The individual will make the choice inattentively iff $E[U_{in}] > E[U_{att}]$—that is, iff

$$\varsigma > (1-\theta)\underbrace{\left|E[U(x)] - E[U(y)]\right|}_{\Delta E[U]}. \tag{4}$$

This implies that iff the cost of exerting effort exceeds the expected benefit of making a rational attentive choice, the individual will instead make an inattentive choice. Handel and Schwartzstein (2018) refer to this as *frictions*, where individuals abstain from using information because it is too costly and instead make an inattentive choice. Condition (4) can be divided into three components, which may explain why different choice situations are more or less likely to be solved by optimization: (i) The higher the required effort, ς, the more likely the inattentive choice is.[2] The right-hand side of (4) captures the expected utility loss from following some simplified choice rule rather than optimizing: (ii) $1-\theta$ is the subjective probability of making a mistake in the inattentive choice. The lower it is, the more likely is an inattentive choice. Someone who *ex ante* is confident in making the right choice (high θ) without reasoning is thus less likely to exert effort to make an attentive choice. (iii) If the choice is considered to be unimportant (the expected utility consequence of making a mistake is small), the probability is also higher that the choice will be inattentive. This is captured by the magnitude of the absolute value of the difference in expected utilities, which we simply denote $\Delta E[U]$.[3]

Hence, required effort in relation to confidence and importance determines whether effort is exerted or not—that is, whether System 2 will be allowed to make an atten-

[2]Note that individuals may differ in their efficiency in their exerted effort.

[3]It can be noted that an unimportant choice in this context can refer to a choice between two goods each corresponding to high utilities, which might seem counterintuitive. However, if the difference in utilities between two such choices is close to zero, the cost of making a mistake is still very low.

tive choice or whether System 1 should make an inattentive one instead.[4] Different combinations of importance and confidence are illustrated in Figure 2.

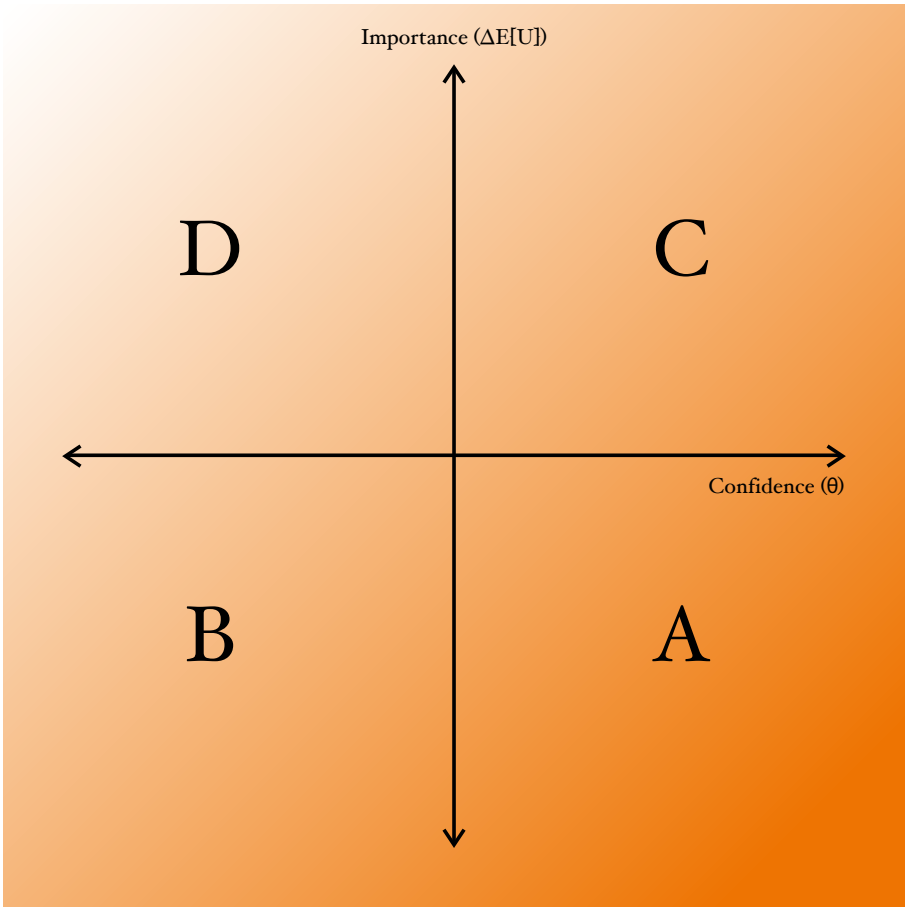

Figure 2: Attentiveness in the importance and confidence space

The intensity of the color reflects the likelihood of making an inattentive choice. According to (4), the product of unconfidence $(1-\theta)$ and importance $(\Delta E[U])$ has to be lower than the required effort, ς, for an inattentive choice to take place. In the top-left corner of Figure 2, an unconfident individual makes an important choice, which thus is the choice situation most likely to be made attentively. Only if ς is very high, a choice categorized in D would be made inattentively. In the bottom-right corner, A, we instead have the choices most likely to be made inattentively. If the choice is considered unimportant by a confident individual, the choice will be made inattentively also despite a very low cost of making the attentive choice. In

[4]In the terminology of Kahneman (2003), System 1 refers to intuitive decision making and System 2 to rational reasoning.

between, we have B, where an unconfident decision maker makes an unimportant choice, and C, where the decision maker is confident when making an important choice.

A critical question is what determines the individual's confidence about not making a mistake in the inattentive choice, θ. Without any information about the alternatives, a random draw between x and y could be made, and the probability of the preferred outcome is $\theta = 0.5$. Since the individual cannot have a prior belief that is worse than random, it follows that $\theta \geq 0.5$. The more confident the individual is about making the right choice also without optimization, the higher is θ. The actual choice rule may be decisive for the value of θ. If it is due to habit, the individual is probably confident about choosing the right option, and θ is close to one. If the choice rule is to choose whatever a friend, parent, or authority chooses or decides, the size of θ depends on how strong is the confidence that one has in that person. Hence, although one has a very vague idea about $U(x)$ and $U(y)$, one could still be very confident about picking the best option.

In Section 3, we explore the mechanisms behind inattentive choices in each of the four categories in Figure 2 and determine the likely nudgeability in each specific choice situation.

3 Inattentive choice and nudges

Assume that inequality (4) holds, and the choice is made inattentively. To understand how this effortless choice is made, we join the strand of literature that makes a distinction between actual experienced utility and what determines the decision (commonly referred to as decision or choice utility) (see, e.g., Chetty, 2015; Gabaix, 2014; Handel and Schwartzstein, 2018). The inattentive choice between x and y is thus based on choice utility V rather than on actual utility U. The inattentive decision maker will choose the alternative that yields the highest choice utility, V, which consists of two parts: expected utility and preference-irrelevant attributes.[5] Corresponding with Step 1 of the decision-making process, we weigh the two parts of the choice utility with a rescaling of the confidence parameter, $\vartheta(\theta) : [0.5, 1] \rightarrow [0, 1]$:

$$V(i) = \vartheta E_2[U(i)] + (1 - \vartheta)\mu_i, \quad i = x, y \tag{5}$$

[5]Preference-irrelevant attributes correspond to what others such as Bernheim and Rangel (2009) and Chetty (2015) denote as ancillary conditions.

where $E_2[U]$ is the expected utility in the inattentive choice in Step 2, which may be different from the *ex ante* expected utility, $E[U]$. μ contains attributes that are irrelevant for the choice in a neoclassical way–that is, they are irrelevant for the experienced utility (such as placement on a shelf). Hence, the inattentive choice is determined by the sign of the difference

$$V(x) - V(y) = \vartheta\big(E_2[U(x)] - E_2[U(y)]\big) + (1-\vartheta)\big(\mu_x - \mu_y\big), \tag{6}$$

or, in short,

$$\Delta V = \vartheta \Delta E_2[U] + (1-\vartheta)\Delta\mu. \tag{7}$$

If $V(x) - V(y) > 0$, then x is chosen, and y is chosen if the inequality is reversed. If $V(x) = V(y)$, the choice is made by a random draw.

Although the choice between x and y is made inattentively, the choice is not independent of preference-related attributes, $\Delta E_2[U]$, and the degree of confidence, $\vartheta(\theta)$. The individual may have some prior belief that one of the alternatives is the best (a belief which may or may not be correct), and if $\vartheta = 1$, this prior belief completely determines the outcome of the inattentive choice. The other extreme is that $\vartheta = 0$, in which $\Delta E_2[U]$ has no effect at all, but the choice is completely based on the preference-irrelevant attributes μ's. Whenever $\vartheta \in (0,1)$, both mechanisms are at play to various degrees. We define a *pure nudge* as an alteration of μ_x or μ_y, or both, and nothing else. For a pure nudge to have an effect, these attributes therefore have to play a significant role in the inattentive decision. A *preference nudge* is instead an alteration of the expected utility in the second step $E_2[U]$, but not of U itself.

Definition 1. *A pure nudge is an alteration of a preference-irrelevant attribute in an inattentive choice situation.*

Definition 2. *A preference nudge is an alteration of expected utility in an inattentive choice situation, without altering actual utility.*

A nudge thus means altering one of the two components in equation (7), without altering actual utility or required effort—that is, without altering the attentive choice. For the nudge to be effective, it must change the sign of ΔV in (7). We can therefore identify two general claims:

Claim 1. *A nudge is more likely to be effective the less important the choice is considered.*

The less important the choice, the more likely it is to be made inattentively in the first place. If the inattentive choice is unimportant ($\Delta E_2[U]$ of a small magnitude), the preference-irrelevant attributes have a stronger weight in the choice utility. Hence, a pure nudge is more likely to be effective the less important the choice is. However, also a preference nudge is more likely to work when the value of $\Delta E_2[U]$ is small, because a smaller intervention is needed to reverse the sign. Moreover:

Claim 2. *The lower the confidence, the more likely it is that a pure nudge is more effective than a preference nudge.*

In the inattentive choice, the relative effectiveness of a pure nudge versus a preference nudge depends on the confidence, ϑ, of the decision maker. For a given $\Delta E_2[U]$, the less confident the decision maker is, the more likely it is that a pure nudge is more effective than a preference nudge: At the limit, a preference nudge cannot have any effect if $\vartheta = 0$, and a pure nudge can have no effect if $\vartheta = 1$. Moreover, for a confident decision maker to be affected by a preference nudge, the choice must be sufficiently unimportant, so that $\Delta E_2[U]$ is easily reversed.

Hence, we can now claim the following:

Claim 3. *Making an inattentive choice is a necessary, but not sufficient, condition for being nudgeable.*

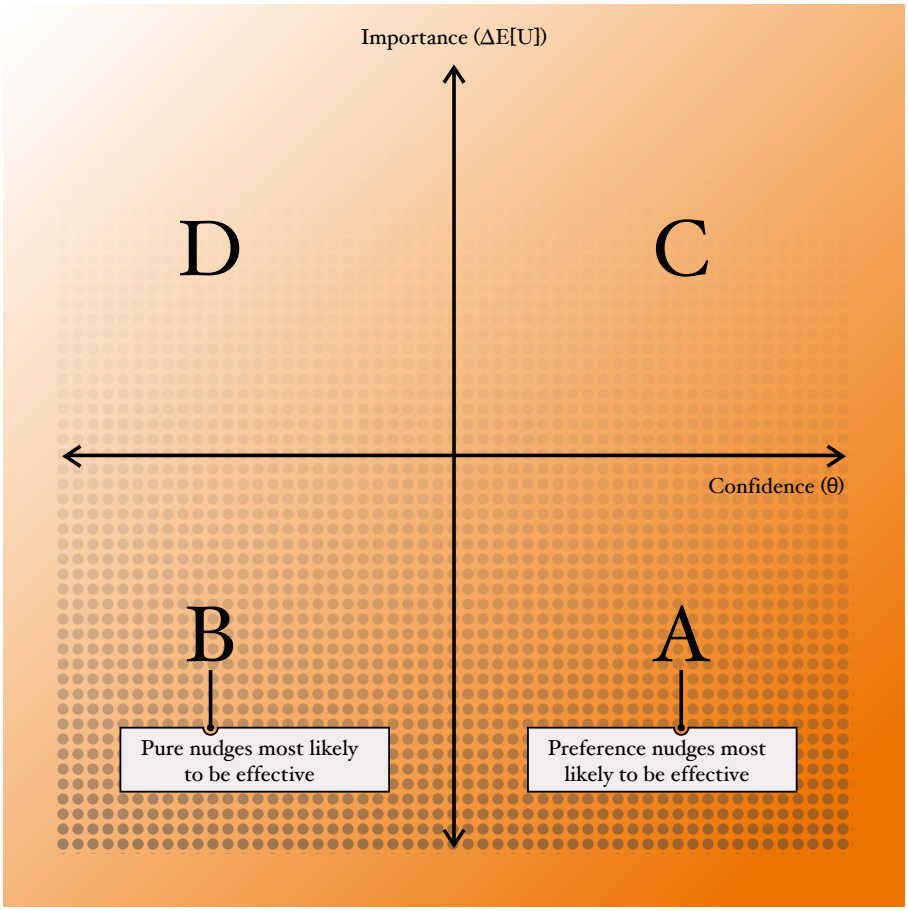

Figure 3: Nudgeability and attentiveness in the importance and confidence space

In Figure 3, the intensity of the color still indicates the probability of an inattentive choice rather than an attentive one, but here we have added a new dimension: opacity, which illustrates the nudgeability. As can be noted, nudgeability coincides with the degree of unimportance of the choice. Hence, the likelihood of a nudge being effective is higher in areas A and B than in C and D. As noted above, a pure nudge is more likely to be effective in B, while a preference nudge is more likely to be effective in A. However, this does not imply that a nudge cannot also be effective for choices in the upper part of Figure 3. Consider a choice situation in D for which the required effort to make a rational choice is very high. Because of the high cost, the choice is made inattentively and a nudge could actually be effective. The most unlikely choice situations to be nudgeable are found in the upper-right corner (C). A confident decision maker who is strongly convinced that one option is better than the other will hardly alter behavior just due to a nudge.

Before going into the policy implications of our model, we also want to say a

few words about what is *not* a nudge. We do not regard an intervention that affects actual *ex post* utility or the cost of making an attentive choice as a nudge. Making one option more easily accessible or providing information would thus not qualify as a nudge. We would still call these behavioral interventions, though. Information also could alter people's choices when they are made attentively. It could either reduce the effort cost, so that the choice is made attentively in the first place, or make one change one's opinion about the alternatives, so that the outcome of the choice would be different. Information could thus alter behavior in both the attentive and inattentive situations, while a preference nudge would have an effect only on the inattentive choice. Of course, a specific intervention could be a nudge to some individuals but not to others.

Further, we have not yet explicitly discussed the possibility of not making a choice at all, which can be seen as the most effortless choice rule possible. Providing a default option may therefore be a powerful tool. In our model, a default option could work entirely through the μ's and would then count as a pure nudge. If x is the default option, then this would imply that $\mu_x > 0$ or $\mu_y = 0$. In fact, a default option is nothing but the extreme of salience. However, in some circumstances a default can be interpreted by the individual as guidance, suggesting that one option is preferable. In such cases the default is a preference nudge, since it will have an effect on $\Delta E_2[U]$ rather than on the μ's. We discuss this in more detail in Section 4.

4 Policy implications for behavioral intervention

Going back to our schematic categorization and our discussions on nudgeability, we now draw some policy implications for behavioral interventions in general. We start by adding a number of choice situations to Figure 3. Although there are individual differences in terms of both confidence and importance, the suggested positioning should be viewed as illustrative examples.

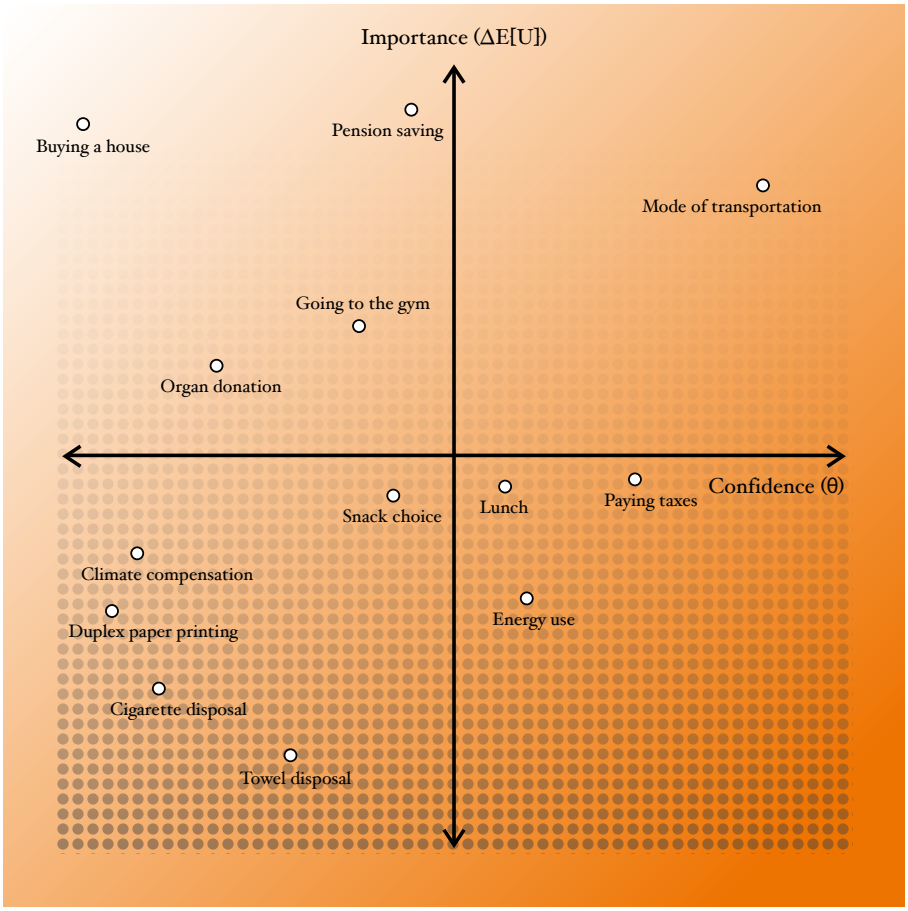

Figure 4: Examples of choices in the importance and confidence space

Choices in the lower part of Figure 4, which are choices that the decision maker considers less important, are those where nudges are most likely to be effective. Examples include towel reuse, duplex printing, snack choice, and what to have for lunch. Commonly used pure nudges for those choices are default options and framing (e.g., placing apples at eye level and candy bars closer to the floor or a vegetarian option on top of the menu). Preference nudges, such as social-norm messages of others' behavior, are also common (e.g., information about how much energy your neighbors use or a note in the shower indicating the share of hotel guests who reuse their towels). However, as (7) shows, in a specific choice situation (categorized as either A or B), pure and preference nudges both could affect the choice of an individual.[6] Returning to Figure 4, the choice of what to have for lunch could possibly be affected by both types of nudges. Changing the order of dishes

[6]It is only in the two extreme cases where $\vartheta \in \{0, 1\}$ that only one of the nudges could be effective.

on the menu (a pure nudge) could affect the choice (Kurz, 2017), but we could also think of a preference nudge that potentially could have an effect such as a sign saying "Chef's recommendation" or "Vegetarian food is good for the climate". As a comparison, we have placed climate compensation in the lower left corner, where a pure nudge is more likely to have an effect than a preference nudge. A commonly used nudge is a default to climate compensate when booking flight tickets (Araña and León, 2013). However, when trying to nudge individuals with a very high degree of confidence, Löfgren et al. (2012) found only little effect of a default option. Then a preference nudge could have been effective instead. Both of these results are in line with the claims of our model.

What about important choices? In the top-left corner of Figure 4, we could place several choices that most often are made attentively, such as buying a house. However, important choices made by unconfident decision makers are made inattentively if the required effort to make an attentive choice is very high. Then even important choices could be nudged. Think about organ donation, which is important, but to the individual perhaps not one of the most important choices. As we have discussed previously, the required effort of making an attentive choice may be sufficiently high to prevent the individual from making the choice attentively. A pure nudge, such as changing the default option, may then be effective if people tend to adhere to a no-choice. However, providing completely new information could of course affect the choice of the attentive decision maker. This would reduce the required effort or even alter the actual utility of the options.[7]

Choices affected by habits can be placed in the top-right corner, where a confident decision-maker makes an important choice. The individual is convinced that one option is better than the other and is also very confident that this is the preferred option. We claim that this situation is the hardest to affect by any behavioral intervention. If one has a strong habit of driving to work it will be very difficult to nudge such a person to commute by bus instead. For information to be effective, it would have to be very alarming to actually cause a change. Hence, if one wants to affect choices influenced by habit formation, rather than using behavioral interventions, a tax or even a ban could be necessary.

[7]Note that information also could have an effect in the inattentive choice via the same mechanisms as a preference nudge.

5 Interpreting the empirical nudge literature in light of our model

In this section, we take a first look at the existing literature from the perspective of our model to explore whether our claims and predictions are consistent with what has been found empirically and experimentally. While it is beyond the scope of this paper to make an extensive literature review, we have made an effort to include at least a large share of the empirical nudge literature. The studies that we reviewed are listed in Table 1. In the empirical literature, the term "nudge" is used quite broadly and includes everything from framing and changes in default to commitment devices and information. Clearly, our definitions of nudges are more narrow. It is clear that many cognitive nudges, such as default and framing (which includes increasing salience and visibility), would be pure nudges according to our definition, and reminding people of social norms could be a preference nudge. It is important to note that receiving information about what others do would count as a nudge only if it would not affect utility in the attentive choice. If being informed about a social or moral norm would affect an individual's rational choice, then we would consider it to be information instead, and not a nudge. In line with this, our review includes nudges of social norm type but excludes studies that are clearly informational or based on commitments, since such interventions are likely to alter the attentive choice. We do include reminders, acknowledging that some reminders may be information rather than nudges. For example, a general reminder from the gym would count as a pure nudge, whereas a reminder that you should not forget your booked spinning class would not. In the former case, the reminder may induce an inattentive choice to actually go to the gym, whereas the choice of exercising has already been made in the latter (although you may have forgotten it).

In Section 3, we identified what is decisive for whether a nudge is effective. Table 1 presents for each study whether the nudge had an effect; which category, A–D, we think best describes the choice; and a judgement of required effort to make an attentive choice. It should be noted that the categorization of the studies is clearly subjective, and this should be seen only as a first (weak) test of our theory. If the findings of the empirical and experimental literature show a pattern consistent with our theory, this would arguably strengthen the model and encourage further studies testing different aspects of it.

Table 1: Summarizing nudge studies

Author	Effect	Type of case (A-D)	Effort
Pure nudge: Default			
Abadie and Gay (2006)	YES	D	High
Araña and León (2013)	YES	B	High
Bronchetti et al. (2013)	NO	D	High
Egebark and Ekström (2016)	YES	B	Low
Johnson and Goldstein (2003)	YES	D	High
Li et al. (2013)	YES	D	High
Löfgren et al. (2012)	NO	A/C	Low
Pure nudge: Visibility/Framing			
Altmann and Traxler (2014)	YES	B	Medium
Blumenthal et al. (2001)	NO	B/D	Medium
Fryer Jr et al. (2012)	YES	C	Low
Gerber and Rogers (2009)	YES	A/B	Medium
Goldzahl et al. (2018)	NO	D	Medium
Guo et al. (2017)	YES	B	Medium
Haynes et al. (2013)	YES	B/D	Low
Hossain and List (2012)	YES	C	Low
Karlan et al. (2012)	NO	B/D	Low
Karlan et al. (2016)	NO	B/D	Low
Kessler and Roth (2014)	YES	D	High
Kurz (2017)	YES	A/B	Medium
Sanders (2017)	YES	B	Low
Schulz et al. (2018)	YES	B	Low
Preference nudge: Social norm			
Allcott (2011)	YES	A/B	Medium
Alm et al. (2017)	YES	B/D	Medium
Beshears et al. (2015)	YES	D	High
Castro and Scartascini (2015)	NO	B/D	Medium
Chen et al. (2010)	YES	B	Low
Ferraro and Price (2013)	YES	A/B	Medium
Goldstein et al. (2008)	YES	B	Low
Goldzahl et al. (2018)	NO	D	Medium
Hallsworth et al. (2017)	YES	B/D	Medium
Holladay et al. (2016)	YES	B/D	Medium
Sanders and Smith (2016)	YES	B/D	Medium
Shang and Croson (2009)	YES	B	Low
Sudarshan (2017)	YES	B	Medium
Pure nudge / Preference nudge / Information: Reminders			
Altmann and Traxler (2014)	YES	B	Medium
Apesteguia et al. (2013)	YES	A/B	Low
Ariel (2012)	NO	B/D	Medium
Bott et al. (2017)	YES	B/D	Medium
Cadena and Schoar (2011)	YES	D	Medium
Calzolari and Nardotto (2016)	YES	A	Medium
Castleman and Page (2015)	YES	D	High
Chande et al. (2015)	YES	C	Low
Dale and Strauss (2009)	YES	A/B	Medium
Damgaard and Gravert (2018)	YES	B	Low
Gerber and Green (2000)	YES/NO	A/B	Medium
Guyton et al. (2017)	YES	B/D	High
Habla and Muller (2018)	YES	A/B	Medium
Haynes et al. (2013)	YES	B/D	Low
Karlan et al. (2012)	NO	B/D	Low
Karlan et al. (2016)	NO	B/D	Low
Nickerson and Rogers (2010)	YES/NO	A/B	Medium
Tørnblad et al. (2014)	NO	C	High

A few things can be noted. First, very few studies concern choices categorized as C (important and high confidence). The most commonly studied choice situation seems to be B (unimportant and unconfident). This is what we would expect from our model, given that researchers tend to study nudges in situations where they *ex ante* are most likely to be effective. Perhaps surprising, we find many studies characterized as D (important and unconfident), although in most of these studied cases, we also judge the required effort to be sufficiently high for the decision to be made inattentively, and therefore nudgeable.

Our theory suggests that nudges should be more effective in A/B situations than in C/D. In line with this, we actually see that nudges in the studied A/B situations seem to be effective. Moreover, most studies that do not find an effect are categorized as either D or B/D situations. In most situations that we have labeled as B/D, we judge that the importance of the choice varies significantly among individuals, and it is therefore difficult to specifically label the choices. One such example is tax compliance (e.g., Hallsworth et al., 2017), which for some individuals may be very important while for others less so.

All in all, although our interpretation of the empirical literature to fit our model is at best a crude illustration, it corroborates the mechanisms identified in our model.

6 Conclusion

We have offered a framework that allows us to understand important mechanisms of decision making. We think of decision making as a two-step process. In the first step, one decides whether the decision should be attentively or inattentively made in the second step. In the terminology of Kahneman (2003), the attentive choice is made using System 2, where the reasoning and rational individual maximizes utility, potentially due to some effort in terms of, e.g., a cognitive cost. An inattentive choice is instead made using the intuitive System 1, whereby one uses some simplified choice rule to avoid costly optimization. Making the attentive choice is thus costly, but the outcome is the optimal one. The required cost is likely to vary considerably across choices and across individuals. How important it is to reach the optimal outcome may also vary. If the cost is high or the choice is unimportant,

it is very likely that it will not pay off to exert the effort, and one will make an inattentive choice instead. The confidence of the individual may also be decisive. The kind of confidence we regard is the *ex ante* confidence that the outcome of an inattentive choice would be the preferred one. Hence, a choice rule built on habit or mimicking the choice of a trustworthy person could make the individual confident that attentive optimization is not needed for the optimal option to be chosen.

If one decides to make the attentive choice, one is by definition not susceptible to nudging. One could, however, be affected by other behavioral interventions, such as information. This could reduce the required effort to make the attentive choice, or new information could affect the choice itself in the second step.

If the choice is instead inattentively made, one may be susceptible to the two distinct kinds of nudges that we define in the paper. We think of an inattentive choice as based on *choice utility*, which is likely to differ from *actual* utility. Choice utility is based on two parts, both expected utility of the choice alternatives and attributes that are totally irrelevant in terms of preferences. Altering any of these preference-irrelevant attributes is what we define as a *pure nudge*. Examples of pure nudges are changing default options, altering menu order, and reducing the plate size. *Preference nudges* are defined as altering the expected utility in the inattentive choice without altering actual utility. This may require some explanation. Actual utility is based on certain attributes, and those are the ones that determine the outcome in the attentive choice. A preference nudge makes one of the options appear better or worse for the inattentive decision maker without affecting any of these attributes. Examples of preference nudges are alerting taxpayers that a majority pay their taxes on time and reminding hotel guests that reusing towels is good for the environment.

The line between preference nudges and information may be fine, though. Both knowledge and what attributes are decisive for attentive optimization vary across individuals. Hence, something that is a nudge for one person may well count as information for another.

In addition to offering clear definitions of nudges, we have discussed in what choice situations different behavioral interventions are more or less likely to be effective. In a choice situation that is unimportant and where the individual is unconfident, a pure nudge is most likely to have an effect. If the individual is confident when making the unimportant choice, a preference nudge would presumably have

a stronger effect. Important choices made by unconfident individuals are the most likely to be made attentively. Then information would be more likely to affect the decision. Only if a high degree of effort is needed to make an attentive choice do unconfident individuals make important choices inattentively, and thereby become nudgeable. For important decisions made by confident individuals, such as those due to habit formation, neither information, nor nudges are likely to be effective. This may explain why travel habits are so hard to affect by behavioral interventions.

When categorizing a large number of empirical nudge studies, we see that they corroborate our theory to a large extent. Very few studies deal with the kinds of choice situations that we claim to be the least nudgeable. Nudges are, however, found to alter behavior in situations that we claim to be nudgeable. Hence, the model we propose and its predictions may guide those who want to alter people's behavior using behavioral interventions.

Our positive analysis can also be used as an input to the normative discussion on nudging. For example, assume that we know that a choice could be effectively nudged based on our model. This does not necessarily mean that a nudge *should* be implemented. A nudge only works when the choice is made inattentively and for some decisions, perhaps the mere fact that the choice is made attentively is more valuable than the outcome of the choice. For instance, in cases where decisions involve ethical and moral considerations. The welfare implications in such choice situations should be thoroughly analysed before implementing a nudge.

References

Abadie, A. and Gay, S. (2006). The impact of presumed consent legislation on cadaveric organ donation: A cross-country study. *Journal of Health Economics*, 25(4):599 – 620.

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

Alm, J., Bloomquist, K. M., and McKee, M. (2017). When you know your neighbour pays taxes: Information, peer effects and tax compliance. *Fiscal Studies*, 38(4):587–613.

Altmann, S. and Traxler, C. (2014). Nudges at the dentist. *European Economic Review*, 72:19–38.

Apesteguia, J., Funk, P., and Iriberri, N. (2013). Promoting rule compliance in daily-life: Evidence from a randomized field experiment in the public libraries of Barcelona. *European Economic Review*, 64:266–284.

Araña, J. E. and León, C. J. (2013). Can defaults save the climate? Evidence from a field experiment on carbon offsetting programs. *Environmental and Resource Economics*, 54(4):613–626.

Ariel, B. (2012). Deterrence and moral persuasion effects on corporate tax compliance: Findings from a randomized controlled trial. *Criminology*, 50(1):27–69.

Bernheim, B. D. and Rangel, A. (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics*, 124(1):51–104.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., and Milkman, K. L. (2015). The effect of providing peer information on retirement savings decisions. *The Journal of Finance*, 70(3):1161–1201.

Blumenthal, M., Christian, C., Slemrod, J., and Smith, M. G. (2001). Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota. *National Tax Journal*, 54(1):125–138.

Bott, K., Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2017). You've got mail: A randomised field experiment on tax evasion. Discussion Paper 10/2017, Norwegian School of Economics.

Bronchetti, E. T., Dee, T. S., Huffman, D. B., and Magenheim, E. (2013). When a nudge isn't enough: Defaults and saving among low-income tax filers. *National Tax Journal*, 66(3):609–634.

Cadena, X. and Schoar, A. (2011). Remembering to pay? Reminders vs. financial incentives for loan payments. Working Paper 17020, National Bureau of Economic Research.

Calzolari, G. and Nardotto, M. (2016). Effective reminders. *Management Science*, 63(9):2915–2932.

Castleman, B. L. and Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization*, 115:144–160.

Castro, L. and Scartascini, C. (2015). Tax compliance and enforcement in the pampas evidence from a field experiment. *Journal of Economic Behavior & Organization*, 116:65–82.

Chande, R., Luca, M., Sanders, M., Soon, X.-Z., Borcan, O., Barak Corren, N., Linos, E., Kirkman, E., and Robinson, S. (2015). Curbing adult student attrition: Evidence from a field experiment. Working Paper 15-065, Harvard Business School NOM Unit.

Chen, Y., Harper, F. M., Konstan, J., and Li, S. X. (2010). Social comparisons and contributions to online communities: A field experiment on MovieLens. *American Economic Review*, 100(4):1358–98.

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, 105(5):1–33.

Dale, A. and Strauss, A. (2009). Don't forget to vote: Text message reminders as a mobilization tool. *American Journal of Political Science*, 53(4):787–804.

Damgaard, M. T. and Gravert, C. (2018). The hidden costs of nudging: Experimental evidence from reminders in fundraising. *Journal of Public Economics*, 157:15–26.

Egebark, J. and Ekström, M. (2016). Can indifference make the world greener? *Journal of Environmental Economics and Management*, 76:1–13.

Farhi, E. and Gabaix, X. (2018). Optimal taxation with behavioral agents. Working Paper 21524, National Bureau of Economic Research.

Ferraro, P. J. and Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1):64–73.

Fryer Jr, R. G., Levitt, S. D., List, J., and Sadoff, S. (2012). Enhancing the efficacy of teacher incentives through loss aversion: A field experiment. Working Paper 18237, National Bureau of Economic Research.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710.

Gerber, A. S. and Green, D. P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review*, 94(3):653–663.

Gerber, A. S. and Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, 71(1):178–191.

Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3):472–482.

Goldzahl, L., Hollard, G., and Jusot, F. (2018). Increasing breast-cancer screening uptake: A randomized controlled experiment. *Journal of Health Economics*, 58:228 – 252.

Guo, Z., Zhao, J., Whong, C., Mishra, P., and Wyman, L. (2017). Redesigning subway map to mitigate bottleneck congestion: An experiment in Washington DC using Mechanical Turk. *Transportation Research Part A: Policy and Practice*, 106:158–169.

Guyton, J., Langetieg, P., Manoli, D., Payne, M., Schafer, B., and Sebastiani, M. (2017). Reminders and recidivism: Using administrative data to characterize nonfilers and conduct EITC outreach. *American Economic Review*, 107(5):471–75.

Habla, W. and Muller, P. (2018). Can email reminders overcome limited attention? An experiment on short- and long-run effects of reminders in the gym. Mimeo, University of Gothenburg.

Hallsworth, M., List, J. A., Metcalfe, R. D., and Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148:14–31.

Handel, B. and Schwartzstein, J. (2018). Frictions or mental gaps: What's behind the information we (don't) use and when do we care? *The Journal of Economic Perspectives*, 32(1):155–178.

Hansen, P. G. (2016). The definition of nudge and libertarian paternalism: Does the hand fit the glove? *European Journal of Risk Regulation*, 7(1):155–174.

Hausman, D. M. and Welch, B. (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy*, 18(1):123–136.

Haynes, L. C., Green, D. P., Gallagher, R., John, P., and Torgerson, D. J. (2013). Collection of delinquent fines: An adaptive randomized trial to assess the effectiveness of alternative text messages. *Journal of Policy Analysis and Management*, 32(4):718–730.

Holladay, J. S., LaRiviere, J., Novgorodsky, D. M., and Price, M. (2016). Asymmetric effects of non-pecuniary signals on search and purchase behavior for energy-efficient durable goods. Working Paper 22939, National Bureau of Economic Research.

Hossain, T. and List, J. A. (2012). The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science*, 58(12):2151–2167.

Johnson, E. J. and Goldstein, D. (2003). Do defaults save lives? *Science*, 302(5649):1338–1339.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9):697–720.

Karlan, D., McConnell, M., Mullainathan, S., and Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science*, 62(12):3393–3411.

Karlan, D., Morten, M., and Zinman, J. (2012). A personal touch: Text messaging for loan repayment. Working Paper 17952, National Bureau of Economic Research.

Kessler, J. B. and Roth, A. E. (2014). Don't take 'no' for an answer: An experiment with actual organ donor registrations. Working Paper 20378, National Bureau of Economic Research.

Kurz, V. (2017). Nudging to reduce meat consumption: Immediate and persistent effects of an intervention at a university restaurant. Working Paper in Economics 712, University of Gothenburg.

Li, D., Hawley, Z., and Schnier, K. (2013). Increasing organ donation via changes in the default choice or allocation rule. *Journal of Health Economics*, 32(6):1117–1129.

Löfgren, Å., Martinsson, P., Hennlock, M., and Sterner, T. (2012). Are experienced people affected by a pre-set default option—Results from a field experiment. *Journal of Environmental Economics and Management*, 63(1):66–72.

Murphy, J., Hofacker, C., and Mizerski, R. (2006). Primacy and recency effects on clicking behavior. *Journal of Computer-Mediated Communication*, 11(2):522–535.

Nickerson, D. W. and Rogers, T. (2010). Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. *Psychological Science*, 21(2):194–199.

Reis, R. (2006). Inattentive consumers. *Journal of Monetary Economics*, 53(8):1761 – 1800.

Salant, Y. (2011). Procedural analysis of choice rules with applications to bounded rationality. *American Economic Review*, 101(2):724–748.

Sanders, M. (2017). Social influences on charitable giving in the workplace. *Journal of Behavioral and Experimental Economics*, 66:129–136.

Sanders, M. and Smith, S. (2016). Can simple prompts increase bequest giving? Field evidence from a legal call centre. *Journal of Economic Behavior & Organization*, 125:179–191.

Schulz, J., Thiemann, P., and Thöni, C. (2018). Nudging generosity: Choice architecture and cognitive factors in charitable giving. *Journal of Behavioral and Experimental Economics*, 74:139 – 145.

Shang, J. and Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *The Economic Journal*, 119(540):1422–1439.

Sudarshan, A. (2017). Nudges in the marketplace: The response of household electricity consumption to information and monetary incentives. *Journal of Economic Behavior & Organization*, 134:320 – 335.

Thaler, R. and Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth and happiness*. Penguin Books.

Tørnblad, S. H., Kallbekken, S., Korneliussen, K., and Mideksa, T. K. (2014). Using mobility management to reduce private car use: Results from a natural field experiment in Norway. *Transport Policy*, 32:9–15.