

From the research bench

What is a nudge?

A nudge changes how a choice is presented without removing options or significantly changing incentives. In two US nudge units it averaged 1.4 points.

Author	Published	Reading time	Words
Sinan Isoglu, MBA (Quantic)	18 October 2026	18 min	3,875
Commercial growth leader, lecturer and doctoral researcher			

Management summary	A nudge, in the definition Thaler and Sunstein gave and DellaVigna and Linos quote, is choice architecture that alters behaviour in a predictable way without forbidding any options or significantly changing economic incentives. This concept guide sets out the kinds of nudge the trials distinguish; reports DellaVigna and Linos's comparison of 126 trials by two US nudge units, which averaged 1.4 percentage points, with published trials that averaged 8.7; explains their model, which attributes about 70 per cent of the gap to selective publication made worse by low statistical power; summarises what a meta-analysis by Jachimowicz et al. found about the size and moderators of default effects; and gives a record for testing a change to a renewal process. The sources are one census of government trials with a comparison sample of published trials, and one meta-analysis of 58 default datasets.
--------------------	---

Contents	§ 1	What is a nudge?	2
	§ 2	Which kinds of nudge do the trials distinguish?	3
	§ 3	How large were nudge effects in published trials and in nudge units?	4
	Figure 1	Average effect by mechanism, in nudge units and in published trials	5
	§ 4	Why did the published trials report larger effects?	5
	§ 5	How large are default effects, and what changes them?	7
	§ 6	Which changes to a renewal count as nudges?	9
	Figure 2	A nudge test record	11
	§ 7	How far does this evidence travel to B2B contracts?	12
		References	13

What is a nudge?

§ 1 of 7

DellaVigna and Linos (2022)¹ open their study with the definition from Thaler and Sunstein's 2008 book **Nudge**, which defined nudges as “choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives.” The book is not held for this guide, so the definition is quoted as DellaVigna and Linos print it (p. 81). They call such interventions light-touch, “including simplification, personalization, and social-norm comparison”.

¹ DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica*, 90(1), 81–116.

Besides the predictable effect, the definition sets two conditions: no option is forbidden, and the economic incentives do not change significantly. DellaVigna and Linos¹ apply the second when they build their data, keeping only trials that “do not use financial incentives” (p. 82). They count a change of default as a nudge too, but leave it out: they “exclude interventions with default changes (excluding 2 nudges in 1 trial)”, because these were “the rare exception among Nudge Unit interventions in our sample (only two treatment arms of one trial), and substantively different”, and they drop the two default trials from their published sample as well (pp. 82, 86 and 89).

In this guide, a nudge is a change to the way a decision is put to the people who make it, such as the wording of a notice, the order of the options, the timing of a reminder or the option selected in advance, made to move their choices in a predictable direction while every option stays available and the economic incentives stay essentially as they were. This guide draws the line where DellaVigna and Linos¹ draw it for their samples, which keep no trial with a financial incentive: a discount, a fee or a bonus for one option is not a nudge here, however small, because it changes the incentive, and neither is withdrawing an option. A nudge unit is a team that designs nudges and tests them in randomized trials; in these sources, a federal government team and an office that worked mainly with local governments. In a supplier's business, the comparable decisions are the ones it puts to its own customers: the renewal notice, the order form and the terms it selects in advance, the step that enrolls an account in a new billing method. That application is this guide's reading; neither source studied business buyers.

Which kinds of nudge do the trials distinguish?

§ 2 of 7

DellaVigna and Linos¹ code every nudge “by its policy area, communication channel, and behavioral mechanisms” (p. 89). For the nudge units, the most frequent mechanisms include “simplification of a letter or notice”; “drawing on personal motivation such as personalizing the communication or using loss aversion to motivate action”; “using implementation intentions or planning prompts”; “exploiting social cues or building social norms into the communication”; “adjusting framing or formatting of existing communication”; and nudging “people towards an active choice or making some choices more salient” (pp. 89 and 90). Their Table I files these as simplification and information, personal motivation, reminders and planning prompts, social cues, framing and formatting, and choice design. Its note says that the frequencies for mechanisms “are not mutually exclusive”, since one nudge can use several, and the published trials used fewer per treatment arm: “an average of 1.5, compared to an average of 2.2 in Nudge Unit trials” (pp. 90 and 92).

The units reached people mostly “through email, letter, or postcard”, while in-person contact was common in the published trials (p. 89). One federal trial shows the type. Service members had to re-enrol in a retirement savings plan by a deadline; the control group received the usual email, while “the treatment group received a simplified, personalized email with loss framing and clear action steps” (p. 86). Other unit trials encouraged utility customers to sign up for automatic payment and electronic bills, or reminded residents to pay a bill (pp. 86, 89 and 97).

Defaults are a further kind, close to what DellaVigna and Linos¹ call choice design (p. 102), though their samples leave defaults out. Jachimowicz et al. (2019)² start from the observation that when people decide “with a pre-selected choice option”, “they are more likely to select that option”. They count defaults among “the most widely employed tools in the choice architecture toolbox” and stress that “they are only a single tool in the choice architect’s toolbox”: a smaller, curated choice set, a limited window in which to sign up, clearer information and an active choice, in which people “are required to make a choice”, are others (pp. 159 to 161).

² Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, 3(2), 159–186.

How large were nudge effects in published trials and in nudge units? § 3 of 7

The census covers every trial run by two US units: the Office of Evaluation Sciences, “which collaborates with federal government agencies”, and the North America office of the Behavioural Insights Team, which “worked primarily with local government agencies” (p. 85). “Both units kept a comprehensive record of all trials from inception in 2015”, and by July 2019 they had run 165 (p. 82). Keeping randomized trials in the field with a clear control group, no financial incentives, a binary outcome and no default change, the authors hold 126 trials with 241 nudges and more than 23 million participants. Their comparison sample comes from two recent meta-analyses of published nudge trials, restricted in the same way, and holds 26 trials with 74 nudges and 505,337 participants (pp. 81 and 82).

In the published trials, “the average impact of a nudge is very large”: take-up rose by 8.7 percentage points, “a 33.4 percent increase over the average control group take-up of 26.0 percent”. In the units’ trials the average was 1.4 points, “an 8.0 percent increase over an average control group take-up of 17.3 percent”, which the authors call “still sizable and highly statistically significant” (pp. 81 to 83). Take-up is the share of people who did what a trial asked of them, such as getting vaccinated or enrolling in a savings plan. The unit average is “about one sixth the size of the estimate in academic papers”, and the authors “can reject the hypothesis of equal effect sizes in the two samples” (p. 83). They add that the 1.4 points come “with a marginal cost that is typically zero or close to zero” (p. 84).

Figure 1 splits both averages by mechanism. In the units, only choice design averaged more than 3 points; in the published trials, every mechanism averaged more than 5.

Figure 1 Average effect by mechanism, in nudge units and in published trials

In the nudge units, only choice design averaged more than 3 points. In the published trials, every mechanism averaged more than 5.

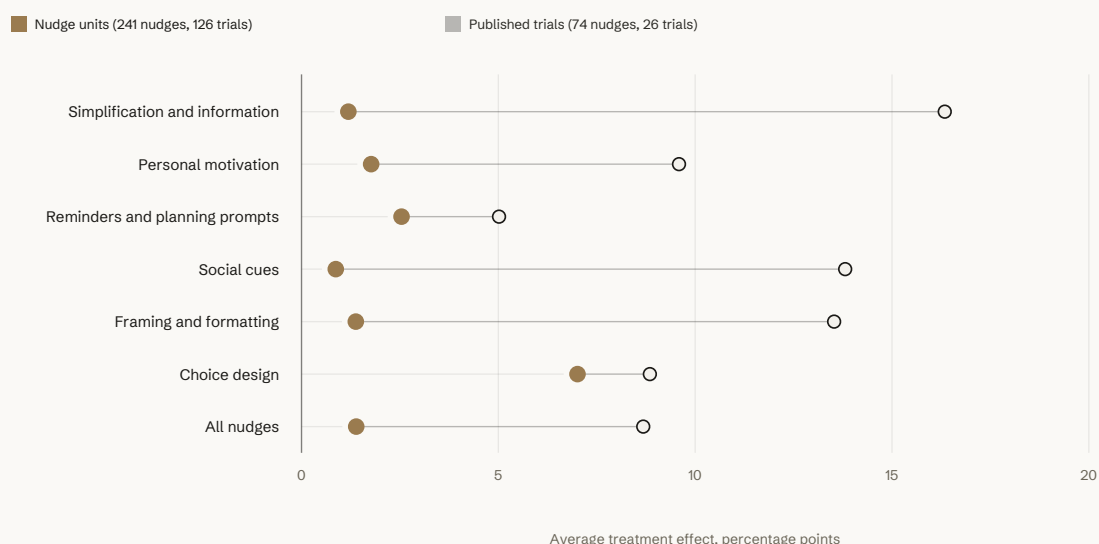


Table I prints averages without standard errors. In the published trials, simplification and information rests on 4 nudges in 2 trials and choice design on 15 nudges in 9; in the units, choice design rests on 15 nudges in 12 trials. The authors caution against reading differences between categories causally.

Source DellaVigna and Linos (2022), p. 92, Table I (continued), Mechanism rows and Total: average treatment effect in percentage points of take-up; Nudge Units 241 nudges in 126 trials, Academic Journals 74 nudges in 26 trials. The mechanism categories are not mutually exclusive, so one nudge can count in several rows; default changes are excluded from both samples.

Some of these averages rest on few nudges, and the authors caution “against a causal interpretation of these heterogeneity results”, because the differences between categories “may reflect feasibility constraints” (p. 102).

Why did the published trials report larger effects?

§ 4 of 7

The authors examine three dimensions, while allowing that others may also play a role: the statistical power of the trials, the features of the nudges, and selective publication, which for them covers “not only whether a journal would publish a paper, but also whether a researcher would write up a study” (pp. 83 and 84).

The median published nudge was tested on a treatment arm of 484 people and could detect, with 80 per cent power, an effect of 6.3 percentage points; that smallest detectable effect is called the minimum detectable effect. The median unit arm had 10,006 people and a minimum detectable effect of 0.8 points, so “the statistical power in the Academic Journals sample is an order of magnitude smaller” (p. 83). In a forecast survey of a convenience sample, the median nudge practitioner expected 1.95 points for the unit trials, and the median academic expected 4.0 points for the units and 7.0 for the published trials. The authors read this as suggestive: academics’ expectations “may shape their trial design”, though the difference in power may also reflect different views of what effect size is relevant for policy or publishable (pp. 83 and 100).

Selective publication shows in the published trials’ test statistics. Nine papers had a most significant arm with a t-statistic between 1.96, the usual threshold for significance, and 2.96, and only two had one between 0.96 and 1.96, which suggests that the probability of publication for papers without a significant result “is only a fraction” of that for papers with one (p. 106). The unit trials show no such pattern, which the authors find “consistent with the fact that we observe the universe of completed trials” (p. 106). Their model, which corrects for selective publication, estimates that papers with no significant result “only have one-tenth the probability of being published”, and that without this selection the published average would be 3.89 points, “quite a bit lower than the unweighted estimate of 8.7 pp., but still higher (though not significantly so) than the estimated Nudge Unit ATE”, the average treatment effect in the units (p. 111).

Their summary is that “selective publication in the Academic Journals sample, exacerbated by low statistical power, explains about 70 percent of the difference in effect sizes between the two samples”, and that “Different nudge characteristics account for most of the residual difference” (p. 81). The share moves with the model: 66 per cent in the version without nudge characteristics, and 77 per cent in the two versions with them, which assign about a fifth of the gap to the characteristics, all with wide bootstrap standard errors (Table V, p. 109).

Low power did its damage through selection. In the authors’ simulations, giving the published trials the units’ sample sizes while keeping the selection lowers their simulated average only from 7.33 to 6.26 points, while removing the selection lowers it to 3.81 (Table VI, pp. 112 and 113). Run the other way, the units’ own nudges, with the

published trials' power and selection, would have averaged 3.35 points, "more than twice the observed effect size", and with the published trials' power alone 1.39 against 1.40 as observed: "in the absence of publication bias (Row 7), having lower statistical power would not bias the average point estimate" (pp. 112 and 113).

Nudge features matter too. In a regression that does not model selective publication, controls for policy area, medium, mechanism, the control group's take-up and the time to the outcome "reduce the difference in point estimate between the samples by two thirds, from 7.3 pp. (Column 1) to 2.4 pp. (Column 2)"; larger estimates went with environmental topics, with no previous communication, with in-person contact, with framing and formatting and especially with choice design (pp. 101 and 103). The authors add that "some of these nudge categories may be capturing differences in effect sizes due to selective publication", since in-person nudges had smaller samples, and that differences in academic involvement "do not appear to explain our findings" (pp. 101 and 103). The two thirds and the 70 per cent therefore cannot be added together; in the model that has both, the characteristics take about a fifth of the gap (Table V, p. 109).

The evidence over anecdote essay uses this comparison as a lesson in what a number has to survive before it reaches a decision, and the uplift specification essay uses it to show why an uplift claim must name its evidence object. This guide asks what the comparison says about nudges. The ratio of one sixth is a fact about this published sample and this census, not a discount factor for other settings: the unit trials "are not exact replications of the Academic Journal nudges", so the authors' results "cannot measure the magnitude" of two forces in publication, luck or p-hacking that inflates the effects that get published, and selection of the interventions that really work (p. 85).

How large are default effects, and what changes them? § 5 of 7

Jachimowicz et al. (2019)² define the default effect as "the difference in choice between the opt-out condition versus that in the opt-in condition" (p. 162): in an opt-out condition the option is selected in advance and people act to leave it, in an opt-in condition they act to choose it. Their meta-analysis pooled 58 datasets from 55 studies in 35 articles, with 73,675 participants in all, from a literature search that ended in May 2017 (pp. 159 and 163).

On average, a pre-selected option was chosen more often by $d = 0.68$ standard deviations (95 per cent confidence interval 0.53 to 0.83), or

0.63 in a model that allows for several observations from the same study or article; the authors call it “a medium-sized effect given conventional criteria” (p. 164). Cohen’s d states a difference in standard deviations, so it cannot be compared directly with DellaVigna and Linos’s¹ percentage points. In the 38 datasets with a yes-or-no outcome, opt-out defaults raised uptake “by an absolute average of 27.24%” (p. 165). That figure cannot be set beside the 1.4 points either: it sets an option selected in advance against one people must choose for themselves, across laboratory and field studies, while DellaVigna and Linos set a nudge against the usual communication, or none, in government services, with defaults left out.

The average hides wide variation. In 46 observations the effect was significant and positive, in ten it was not significant and in two it was significant and negative, and the heterogeneity statistic I^2 was 98.21 per cent, and 98.01 in the three-level model, which is “considered to be very high heterogeneity” (pp. 166 and 168). For a given use, the authors warn, the effect “may be substantially lower or higher than the meta-analytic average” (p. 175). A trim-and-fill analysis estimated that eight studies were missing and that including them would raise the average to $d = 0.80$, so “if anything, default studies finding larger effects are missing from the literature”; a regression test for funnel asymmetry was not significant, and the authors ask that the results “be interpreted with caution” (p. 166).

Two sets of factors went with the size of the effect. Defaults in consumer choices, “defined as decisions related to buying a product or service”, had larger effects ($b = 0.73$, $p = 0.003$), and those in environmental choices smaller ones, significant only at the 10 per cent level ($p = 0.08$); “No other study characteristic explained any further systematic variance”, including whether a study was a field or a laboratory experiment (pp. 168 to 171). The authors “can only speculate” why, and suggest that preferences about consumption may be held less strongly than preferences about the environment (pp. 176 and 178).

The second set concerns how a default works. The authors take three channels from earlier research: endorsement, when the default is read as what the choice architect recommends; ease, because staying with the pre-selected option takes less effort than choosing another; and endowment, when the default is felt to be the status quo. Two trained coders, not part of the author team, rated each study’s design on each channel. Endorsement ($b = 0.32$, $p = 0.038$) and endowment ($b = 0.31$, $p = 0.044$) were significant moderators, while “Contrary to our

prediction, ease is not a significant moderator” (pp. 171 to 174). The authors call their ratings “only noisy approximations”, note that the pooled studies rarely varied how hard it was to leave a default, which may partly explain the null result for ease, and report that endowment was significant only at the 10 per cent level with a more robust estimator (pp. 174 and 176). They cite one study in which defaults were less effective when people did not trust the choice architect, and they note that defaults rest on the assumption that they “allow those whose preferences are different from the default option to easily select an alternative” (pp. 172 and 179).

DellaVigna and Linos¹ left defaults out of both samples. Adding the few default arms makes both estimates “slightly larger”, “with the caveat that these are just 3 arms in the Academic Journals sample and 2 arms in the Nudge Units sample” (p. 97), and they name the units’ institutional constraints, which make features “such as default changes” less likely, as the first of three reasons why the 1.4-point estimate “likely is a lower bound of the impact of behavioral science”; the other two are that an organization can adopt the most successful arm of a trial and that researchers can build later interventions on the most successful results (p. 84). Jachimowicz et al.² draw a practical conclusion from the variation: “one must test applications rigorously, rather than just assuming that they will always work as expected” (p. 177). Separately, to learn how the preferences of the people facing a default are distributed, they suggest a forced-choice condition, which shows what people choose without one (p. 179).

Which changes to a renewal count as nudges?

§ 6 of 7

The guide’s definition gives a commercial team a test to run before any trial: does the change keep every option open, add no discount, fee or bonus, and leave price and other incentives essentially as they were? An illustration, with no company data: a supplier of maintenance contracts for industrial equipment renews about 400 contracts a quarter and is considering five changes. A one-page renewal notice with the deadline and three steps, in place of the full terms letter, is simplification and information. A reminder two weeks before the deadline falls under reminders and planning prompts. A renewal form that opens with the customer’s current service level selected, every other level as easy to pick as before, is a default that reflects the status quo. Three per cent off for renewing early changes the economic incentive, the kind of financial incentive DellaVigna and Linos¹ keep out of their samples, and withdrawing monthly billing forbids an option, so neither of these two is a nudge on this guide’s definition.

The meta-analysis says most about the default, but less about this default than it may seem. A default felt as the status quo is the endowment case, which went with larger effects, but the articles the meta-analysis pooled, 34 of them starred in its reference list and one, Dinner et al. (2011), named in its forest plot, concern choices such as organ donation, end-of-life care, vaccination, green energy and charitable giving; none, judging by their titles, concerns business buyers or a contract renewal. Keeping every other service level as easy to select as before is what the assumption behind defaults requires, and the example classifies this default without recommending it.

A record for testing such a change has six fields: the decision and its options, the change and its mechanism, whether it is a nudge at all, the comparison, the smallest effect the test can detect, and what is reported. The fields follow DellaVigna and Linos's¹ definition and coding, their restriction to randomized trials with a control group, their power calculations and their finding on selective publication, and Jachimowicz et al.²'s call to test defaults rather than assume that they work.

Figure 2 A nudge test record

Write down the decision, the change, whether it is a nudge, the comparison, the detectable effect and the report before the test starts.

Decision	Change	Nudge?	Comparison	Detectable	Report
Whose choice, which options?	Which mechanism or default?	Options kept, incentives unchanged?	Random assignment, arm sizes?	Smallest effect at 80% power?	Outcome, window, every arm?

Report every arm, significant or not: low power biased the published average only together with selection.

One row per proposed change. A change that removes an option or alters the price is recorded and set aside, because it is not a nudge. State the smallest detectable effect before the test starts.

Source Author's decision worksheet. The decision, change and nudge fields follow DellaVigna and Linos (2022), pp. 81, 82 and 89 to 92; the comparison field follows DellaVigna and Linos (2022), pp. 82 and 85 to 86, and Jachimowicz et al. (2019), pp. 177 and 179; the detectable field follows DellaVigna and Linos (2022), pp. 83 and 98; the report field follows DellaVigna and Linos (2022), pp. 84, 106 and 111 to 113. The fields are reader inputs; no company data is supplied.

For the renewal example, suppose 200 accounts are assigned at random to each process and 80 per cent of contracts are renewed by the deadline under the current one. At the 80 per cent power that the authors' calculations use, and with a two-sided test at the 5 per cent level, the smallest detectable difference is about 11 percentage points ($2.8 \times 2 \times 0.8 \times 0.2/200$ approximately 0.112), and a difference would have to reach about 7.8 points to be significant at all. Both are many times the units' average of 1.4 points. If renewal nudges moved business buyers by amounts like that, which the sources do not test, a quarter's trial could not detect one, and a result that did reach significance would have the profile the authors warn about in under-powered published trials: large, and "unlikely to replicate at the same effect size" (p. 85). Reporting every arm, significant or not, keeps a team's own record free of the selection that inflated the published average. These are the author's examples and have not been tested.

How far does this evidence travel to B2B contracts?

§ 7 of 7

Not far without a test of its own. DellaVigna and Linos (2022)¹ studied US government services, with outcomes such as vaccination, benefit enrolment, bill payment and registration, reached mostly by letter, email or postcard; they call 1.4 points “a reasonable estimate for the average impact of a nudge at scale on government services” (p. 84). They “do not observe the micro-data for each trial”, their published sample is drawn from two meta-analyses under the same restrictions, and they caution against reading the differences between categories causally (pp. 82, 102 and 114). Jachimowicz et al. (2019)² pooled 58 opt-in versus opt-out comparisons from laboratory and field studies; their channel ratings are coders’ judgements, and their study characteristics “reflect the decisions of researchers about what studies to conduct and how to implement the default intervention” (p. 176). Neither source studied business buyers, buying committees or contract renewals. The [external validity guide](#) sets out how to name the population, treatment, outcome and context that a result is meant to transfer to.

On this guide’s reading, three things carry over with some confidence. The guide’s own definition sorts proposed changes before anyone tests them: a discount or a withdrawn option is not a nudge. A large effect from a small test calls for doubt when it is the result that was kept (the significant arm or the study that was written up), because in the authors’ simulations low power raised the average only together with that selection; the units’ 1.4 points show how far an at-scale average in government services fell below a published one, not what a renewal nudge will do. And a default’s effect varies too widely to be assumed; it has to be measured where it is used. Whether renewal and billing nudges move business buyers by amounts like those in government services, and whether the consumer result for defaults holds for firms, the sources do not say.

 End of essay

References

- 1 DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica*, 90(1), 81–116.
<https://doi.org/10.3982/ECTA18709>
- 2 Jachimowicz, J. M., Duncan, S., Weber, E. U., & Johnson, E. J. (2019). When and why defaults influence decisions: A meta-analysis of default effects. *Behavioural Public Policy*, 3(2), 159–186.
<https://doi.org/10.1017/bpp.2018.43>

How to cite this essay

Isoglu, S. (2026, October 18). What is a nudge? Sinan Isoglu.
<https://isoglu.com/insights/journal/what-is-a-nudge/>
Online version: <https://isoglu.com/insights/journal/what-is-a-nudge/>

Sinan Isoglu, MBA (Quantic)

Commercial growth leader, lecturer and doctoral researcher
© 2026 Sinan Isoglu · isoglu.com · All rights reserved