

Aus der Forschung

Was ist ein Nudge?

Ein Nudge ändert die Darstellung einer Wahl, streicht keine Option und ändert Anreize kaum. In zwei US-Nudge-Einheiten brachte er im Schnitt 1,4 Punkte.

Autor	Veröffentlicht	Lesezeit	Wörter
Sinan Isoglu, MBA (Quantic) Führungskraft für Umsatzwachstum, Dozent und Doktorand	18. Oktober 2026	21 Min.	4.462

Management Summary Ein Nudge ist nach der Definition von Thaler und Sunstein, die DellaVigna und Linos zitieren, Entscheidungsarchitektur, die Verhalten vorhersehbar ändert, ohne Optionen zu verbieten oder wirtschaftliche Anreize wesentlich zu ändern. Der Leitfaden nennt die Arten von Nudges in den Studien; berichtet, wie DellaVigna und Linos 126 Studien zweier US-Nudge-Einheiten, im Schnitt 1,4 Prozentpunkte, mit publizierten Studien, im Schnitt 8,7, vergleichen; erklärt ihr Modell, das rund 70 Prozent der Lücke selektiver Publikation zuschreibt, verstärkt durch geringe Teststärke; fasst zusammen, was die Meta-Analyse von Jachimowicz et al. zu Größe und Moderatoren des Default-Effekts fand; und bietet ein Blatt zum Testen einer Änderung bei Verlängerungen. Die Quellen sind eine Vollerhebung staatlicher Studien mit Vergleichsstichprobe und eine Meta-Analyse von 58 Datensätzen zu Voreinstellungen.

Inhalt	§ 1	Was ist ein Nudge?	2
	§ 2	Welche Arten von Nudges unterscheiden die Studien?	3
	§ 3	Wie groß waren Nudge-Effekte in publizierten Studien und in Nudge-Einheiten?	5
	Abbildung 1	Durchschnittlicher Nudge-Effekt nach Mechanismus, in den Nudge-Einheiten und in publizierten Studien	6
	§ 4	Warum berichteten die publizierten Studien größere Effekte?	7
	§ 5	Wie groß sind Default-Effekte, und was verändert sie?	9
	§ 6	Welche Änderungen an einer Vertragsverlängerung sind Nudges?	12
	Abbildung 2	Ein Testblatt für einen Nudge	13
	§ 7	Wie weit trägt diese Evidenz bei B2B-Verträgen?	14
		Referenzen	16

Was ist ein Nudge?

§ 1 von 7

DellaVigna und Linos (2022)¹ beginnen ihre Studie mit der Definition aus dem Buch **Nudge**, das Thaler und Sunstein im Jahr 2008 veröffentlichten: Nudges sind demnach Entscheidungsarchitektur, die das Verhalten von Menschen vorhersehbar verändert, ohne Optionen zu verbieten oder ihre wirtschaftlichen Anreize wesentlich zu ändern („choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives“). Das Buch liegt diesem Leitfaden nicht vor; die Definition wird daher so zitiert, wie DellaVigna und Linos sie abdrucken (S. 81). Sie nennen solche Eingriffe sanft, darunter Vereinfachung, Personalisierung und Vergleiche mit sozialen Normen („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.

Neben der vorhersehbaren Wirkung stellt die Definition zwei Bedingungen: Keine Option wird verboten, und die wirtschaftlichen Anreize ändern sich nicht wesentlich. DellaVigna und Linos¹ wenden die zweite an, wenn sie ihre Daten zusammenstellen: Sie behalten nur Studien, die keine finanziellen Anreize einsetzen („do not use financial incentives“, S. 82). Auch eine geänderte Voreinstellung zählen sie als Nudge, lassen sie aber weg. Sie schließen Eingriffe mit geänderten Voreinstellungen aus, zwei Nudges in einer Studie („exclude interventions with default changes (excluding 2 nudges in 1 trial)“), weil diese unter den Eingriffen der Nudge-Einheiten die seltene Ausnahme und inhaltlich andersartig waren („the rare exception among Nudge Unit interventions in our sample (only two treatment arms of one trial), and substantively different“), und sie streichen die beiden Studien mit Voreinstellungen auch aus ihrer Stichprobe publizierter Studien (S. 82, 86 und 89).

In diesem Leitfaden ist ein Nudge eine Änderung daran, wie eine Entscheidung den Menschen vorgelegt wird, die sie treffen, etwa am Wortlaut eines Schreibens, an der Reihenfolge der Optionen, am Zeitpunkt einer Erinnerung oder an der vorab ausgewählten Option. Sie soll die Wahl in eine vorhersehbare Richtung lenken, während jede Option verfügbar bleibt und die wirtschaftlichen Anreize im Wesentlichen bleiben, wie sie waren. Dieser Leitfaden zieht die Grenze dort, wo DellaVigna und Linos¹ sie für ihre Stichproben ziehen, in denen keine Studie mit finanziellem Anreiz bleibt: Ein Rabatt, eine Gebühr oder ein Bonus für eine Option ist hier kein Nudge, wie klein er auch

sein mag, weil er den Anreiz ändert, und das Streichen einer Option ebenso wenig. Eine Nudge-Einheit (Nudge Unit) ist ein Team, das Nudges entwirft und in randomisierten Studien testet; in diesen Quellen ein Team der US-Bundesregierung und ein Büro, das vor allem mit Kommunalverwaltungen arbeitete. Im Geschäft eines Anbieters sind die vergleichbaren Entscheidungen diejenigen, die er seinen eigenen Kunden vorlegt: das Verlängerungsschreiben, das Bestellformular mit den Bedingungen, die es vorab auswählt, und der Schritt, mit dem ein Kunde für ein neues Abrechnungsverfahren angemeldet wird. Diese Übertragung ist die Lesart dieses Leitfadens; keine der beiden Quellen hat Geschäftskunden untersucht.

Welche Arten von Nudges unterscheiden die Studien?

§ 2 von 7

DellaVigna und Linos¹ ordnen jeden Nudge nach Politikfeld, Kommunikationsweg und Verhaltensmechanismus ein („by its policy area, communication channel, and behavioral mechanisms“, S. 89). Für die Nudge-Einheiten zählen sie zu den häufigsten Mechanismen die Vereinfachung eines Schreibens oder Bescheids („simplification of a letter or notice“), die persönliche Motivation, etwa durch Personalisierung oder durch Verlustaversion („drawing on personal motivation such as personalizing the communication or using loss aversion to motivate action“), Wenn-dann-Vorsätze oder Planungshilfen („using implementation intentions or planning prompts“), soziale Hinweise und Normen („exploiting social cues or building social norms into the communication“), ein anderes Framing oder eine andere Formatierung bestehender Schreiben („adjusting framing or formatting of existing communication“) sowie Anstöße zu einer aktiven Wahl oder zur Hervorhebung einzelner Optionen („people towards an active choice or making some choices more salient“, S. 89 und 90). In ihrer Tabelle I heißen diese Vereinfachung und Information, persönliche Motivation, Erinnerungen und Planungshilfen, soziale Hinweise, Framing und Formatierung sowie Gestaltung der Wahl. Laut ihrer Anmerkung schließen sich die Häufigkeiten der Mechanismen nicht gegenseitig aus („are not mutually exclusive“), weil ein Nudge mehrere nutzen kann, und die publizierten Studien setzten je Versuchsgruppe weniger davon ein („an average of 1.5, compared to an average of 2.2 in Nudge Unit trials“, S. 90 und 92).

Die Einheiten erreichten die Menschen meist per E-Mail, Brief oder Postkarte („through email, letter, or postcard“), während in den publizierten Studien persönlicher Kontakt verbreitet war (S. 89). Eine Studie des Office of Evaluation Sciences zeigt den Typ.

Militärangehörige mussten sich bis zu einer Frist erneut für einen Altersvorsorgeplan anmelden; die Kontrollgruppe erhielt die übliche E-Mail, die Versuchsgruppe eine vereinfachte, persönliche E-Mail, die mit drohenden Verlusten argumentierte und klare Handlungsschritte nannte („the treatment group received a simplified, personalized email with loss framing and clear action steps“, S. 86). Andere Studien der Einheiten warben bei Kunden von Versorgungsbetrieben für automatische Zahlung und elektronische Rechnungen oder erinnerten Bürger daran, eine Rechnung zu begleichen (S. 86, 89 und 97).

Voreinstellungen sind eine weitere Art, verwandt mit dem, was DellaVigna und Linos Gestaltung¹ der Wahl nennen (S. 102), auch wenn ihre Stichproben Voreinstellungen ausschließen. Jachimowicz et al. (2019)² gehen davon aus, dass Menschen, die mit einer vorab ausgewählten Option entscheiden („with a pre-selected choice option“), diese Option eher wählen („they are more likely to select that option“). Sie zählen Voreinstellungen zu den am häufigsten eingesetzten Werkzeugen der Entscheidungsarchitektur („the most widely employed tools in the choice architecture toolbox“) und betonen, dass sie nur eines von mehreren sind („they are only a single tool in the choice architect’s toolbox“): Eine kleinere, kuratierte Auswahl, ein begrenztes Zeitfenster für die Anmeldung, klarere Informationen und eine aktive Wahl, bei der Menschen sich entscheiden müssen („are required to make a choice“), gehören ebenfalls dazu (S. 159 bis 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.

Wie groß waren Nudge-Effekte in publizierten Studien und in Nudge-Einheiten?

§ 3 von 7

Die Vollerhebung umfasst jede Studie zweier US-Einheiten: des Office of Evaluation Sciences, das mit Bundesbehörden zusammenarbeitet („which collaborates with federal government agencies“), und des nordamerikanischen Büros des Behavioural Insights Team, das vor allem mit Kommunalverwaltungen arbeitete („worked primarily with local government agencies“, S. 85). Beide Einheiten führten von ihrer Gründung 2015 an ein vollständiges Verzeichnis aller Studien („Both units kept a comprehensive record of all trials from inception in 2015“), bis Juli 2019 waren es 165 Studien (S. 82). Beschränkt auf randomisierte Feldstudien mit klarer Kontrollgruppe, ohne finanzielle Anreize, mit binärer Zielgröße und ohne geänderte Voreinstellung, umfasst ihre Stichprobe 126 Studien mit 241 Nudges und mehr als 23 Millionen Teilnehmenden. Die Vergleichsstichprobe stammt aus zwei neueren Meta-Analysen publizierter Nudge-Studien, ist ebenso eingeschränkt und umfasst 26 Studien mit 74 Nudges und 505.337 Teilnehmenden (S. 81 und 82).

In den publizierten Studien ist die durchschnittliche Wirkung eines Nudges sehr groß („the average impact of a nudge is very large“): Die Inanspruchnahme stieg um 8,7 Prozentpunkte, ein Plus von 33,4 Prozent gegenüber einer durchschnittlichen Inanspruchnahme von 26,0 Prozent in den Kontrollgruppen („a 33.4 percent increase over the average control group take-up of 26.0 percent“). In den Studien der Einheiten lag der Durchschnitt bei 1,4 Punkten, ein Plus von 8,0 Prozent gegenüber 17,3 Prozent in den Kontrollgruppen („an 8.0 percent increase over an average control group take-up of 17.3 percent“), was die Autoren noch immer beachtlich und hoch signifikant nennen („still sizable and highly statistically significant“, S. 81 bis 83). Die Inanspruchnahme ist der Anteil der Menschen, die taten, worum eine Studie sie bat, etwa sich impfen zu lassen oder sich für einen Sparplan anzumelden. Der Durchschnitt der Einheiten beträgt etwa ein Sechstel des Werts in den akademischen Aufsätzen („about one sixth the size of the estimate in academic papers“), und die Autoren können die Hypothese gleicher Effektstärken in beiden Stichproben verwerfen („can reject the hypothesis of equal effect sizes in the two samples“, S. 83). Sie fügen hinzu, dass die 1,4 Punkte mit Grenzkosten verbunden sind, die meist bei null oder nahe null liegen („with a marginal cost that is typically zero or close to zero“, S. 84).

Abbildung 1 schlüsselt beide Durchschnitte nach Mechanismus auf. In den Einheiten lag nur die Gestaltung der Wahl im Schnitt über 3

Punkten; in den publizierten Studien lag jeder Mechanismus im Schnitt über 5.

Abbildung 1 Durchschnittlicher Nudge-Effekt nach Mechanismus, in den Nudge-Einheiten und in publizierten Studien

In den Nudge-Einheiten lag nur die Gestaltung der Wahl im Schnitt über 3 Punkten. In den publizierten Studien lag jeder Mechanismus im Schnitt über 5.

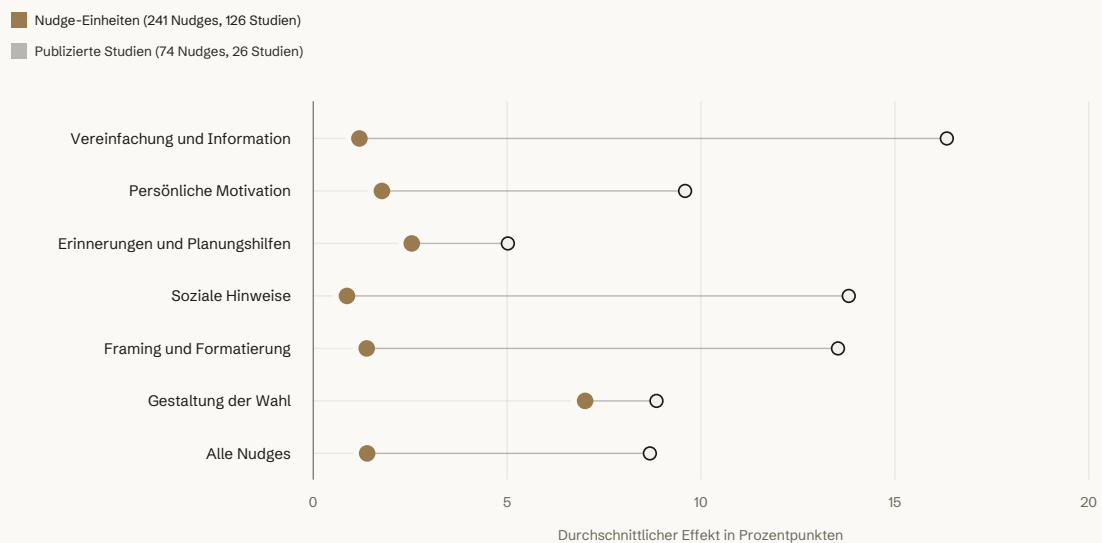


Tabelle I gibt Durchschnitte ohne Standardfehler an. In den publizierten Studien beruht Vereinfachung und Information auf 4 Nudges in 2 Studien und die Gestaltung der Wahl auf 15 Nudges in 9; in den Einheiten beruht die Gestaltung der Wahl auf 15 Nudges in 12 Studien. Die Autoren warnen davor, Unterschiede zwischen den Kategorien kausal zu lesen.

Quelle DellaVigna und Linos (2022), S. 92, Tabelle I (Fortsetzung), Zeilen zu den Mechanismen und Gesamtwert: durchschnittlicher Effekt in Prozentpunkten der Inanspruchnahme; Nudge-Einheiten 241 Nudges in 126 Studien, akademische Zeitschriften 74 Nudges in 26 Studien. Die Mechanismen schließen sich nicht gegenseitig aus, sodass ein Nudge in mehreren Zeilen zählen kann; geänderte Voreinstellungen sind in beiden Stichproben ausgeschlossen.

Manche dieser Durchschnitte beruhen auf wenigen Nudges, und die Autoren warnen vor einer kausalen Deutung dieser Unterschiede („against a causal interpretation of these heterogeneity results“), weil sie Grenzen der Umsetzbarkeit widerspiegeln können („may reflect feasibility constraints“, S. 102).

Warum berichteten die publizierten Studien größere Effekte?

§ 4 von 7

Die Autoren prüfen drei Dimensionen und schließen weitere nicht aus: die Teststärke der Studien, die Merkmale der Nudges und die selektive Publikation, zu der sie nicht nur zählen, ob eine Zeitschrift einen Aufsatz veröffentlicht, sondern auch, ob Forschende eine Studie überhaupt aufschreiben („not only whether a journal would publish a paper, but also whether a researcher would write up a study“, S. 83 und 84).

Im Median wurde ein publizierter Nudge an einer Versuchsgruppe von 484 Personen getestet und konnte mit 80 Prozent Teststärke einen Effekt ab 6,3 Prozentpunkten nachweisen; dieser kleinste nachweisbare Effekt heißt im Englischen minimum detectable effect. Im Median umfasste eine Versuchsgruppe der Einheiten 10.006 Personen, mit einem kleinsten nachweisbaren Effekt von 0,8 Punkten; die Teststärke der akademischen Stichprobe ist damit eine Größenordnung kleiner („the statistical power in the Academic Journals sample is an order of magnitude smaller“, S. 83). In einer Prognoseumfrage unter einer Gelegenheitsstichprobe erwarteten die befragten Nudge-Praktiker im Median 1,95 Punkte für die Studien der Einheiten, die befragten Akademiker im Median 4,0 Punkte für die Einheiten und 7,0 für die publizierten Studien. Die Autoren lesen das als Hinweis: Die Erwartungen der Akademiker könnten ihr Studiendesign prägen („may shape their trial design“), doch der Unterschied in der Teststärke kann auch daher rühren, dass beide Seiten unterschiedlich beurteilen, welche Effektstärke für die Politik bedeutsam oder publizierbar ist (S. 83 und 100).

Die selektive Publikation zeigt sich in den Teststatistiken der publizierten Studien. Neun Aufsätze hatten als signifikantestes Ergebnis eine t-Statistik zwischen 1,96, der üblichen Signifikanzschwelle, und 2,96, und nur zwei eine zwischen 0,96 und 1,96. Das spricht dafür, dass Aufsätze ohne signifikantes Ergebnis nur mit einem Bruchteil der Wahrscheinlichkeit erscheinen, mit der Aufsätze mit einem solchen Ergebnis veröffentlicht werden („is only a fraction“, S. 106). Die Studien der Einheiten zeigen kein solches Muster, was nach den Autoren dazu passt, dass sie die Gesamtheit der abgeschlossenen Studien beobachten („consistent with the fact that we observe the universe of completed trials“, S. 106). Ihr Modell, das die selektive Publikation herausrechnet, schätzt, dass Aufsätze ohne signifikantes Ergebnis nur ein Zehntel der Publikationswahrscheinlichkeit haben („only have one-tenth the

probability of being published“), und dass der publizierte Durchschnitt ohne diese Selektion bei 3,89 Punkten läge: deutlich unter dem ungewichteten Wert von 8,7, aber noch immer, wenn auch nicht signifikant, über dem durchschnittlichen Effekt in den Einheiten („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“, S. 111).

Ihr Fazit lautet, dass selektive Publikation in der akademischen Stichprobe, verstärkt durch geringe Teststärke, rund 70 Prozent des Unterschieds in den Effektstärken der beiden Stichproben erklärt („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“) und dass unterschiedliche Merkmale der Nudges den größten Teil des Rests erklären („Different nudge characteristics account for most of the residual difference“, S. 81). Der Anteil hängt vom Modell ab: 66 Prozent in der Fassung ohne Merkmale der Nudges, 77 Prozent in den beiden Fassungen mit ihnen, die den Merkmalen rund ein Fünftel der Lücke zuschreiben, jeweils mit großen Bootstrap-Standardfehlern (Tabelle V, S. 109).

Die geringe Teststärke wirkte über die Selektion. In den Simulationen der Autoren sinkt der simulierte Durchschnitt der publizierten Studien nur von 7,33 auf 6,26 Punkte, wenn sie die Stichprobengrößen der Einheiten erhalten, die Selektion aber bleibt; ohne Selektion sinkt er auf 3,81 (Tabelle VI, S. 112 und 113). Umgekehrt hätten die eigenen Nudges der Einheiten mit der Teststärke und der Selektion der publizierten Studien im Schnitt 3,35 Punkte gezeigt, mehr als das Doppelte des beobachteten Effekts („more than twice the observed effect size“), mit der Teststärke der publizierten Studien allein dagegen 1,39 gegenüber beobachteten 1,40: Ohne Publikationsbias würde eine geringere Teststärke den durchschnittlichen Schätzwert nicht verzerren („in the absence of publication bias (Row 7), having lower statistical power would not bias the average point estimate“, S. 112 und 113).

Auch die Merkmale der Nudges spielen eine Rolle. In einer Regression, die selektive Publikation nicht modelliert, verringern Kontrollvariablen für Politikfeld, Kommunikationsweg, Mechanismus, die Inanspruchnahme der Kontrollgruppe und die Zeit bis zur Zielgröße den Unterschied der Schätzwerte um zwei Drittel, von 7,3 auf 2,4 Prozentpunkte („reduce the difference in point estimate between the samples by two thirds, from 7.3 pp. (Column 1) to 2.4 pp. (Column 2)“); größere Schätzwerte gingen mit Umweltthemen einher, mit fehlender vorheriger Kommunikation, mit persönlichem Kontakt, mit Framing und Formatierung und vor allem mit der Gestaltung der Wahl (S. 101 und

103). Die Autoren fügen hinzu, dass manche Kategorien zum Teil Effekte der selektiven Publikation abbilden könnten („some of these nudge categories may be capturing differences in effect sizes due to selective publication“), weil Nudges mit persönlichem Kontakt kleinere Stichproben hatten, und dass Unterschiede in der akademischen Beteiligung den Befund nicht zu erklären scheinen („do not appear to explain our findings“, S. 101 und 103). Die zwei Drittel und die 70 Prozent lassen sich daher nicht addieren; im Modell, das beides enthält, entfällt auf die Merkmale rund ein Fünftel der Lücke (Tabelle V, S. 109).

Der Beitrag Evidenz statt Anekdote nutzt diesen Vergleich als Lehre darüber, was eine Zahl überstanden haben muss, bevor sie in eine Entscheidung eingeht, und der Beitrag zur Uplift-Spezifikation nutzt ihn, um zu zeigen, warum eine Uplift-Aussage ihr Evidenzobjekt benennen muss. Dieser Leitfaden fragt, was der Vergleich über Nudges aussagt. Das Verhältnis von einem Sechstel ist ein Befund über diese publizierte Stichprobe und diese Vollerhebung, kein Abschlag für andere Kontexte: Die Studien der Einheiten sind keine exakten Replikationen der akademischen Nudges („are not exact replications of the Academic Journal nudges“), daher können die Ergebnisse der Autoren die Stärke zweier Kräfte bei der Publikation nicht messen („cannot measure the magnitude“): Glück oder P-Hacking, die publizierte Effekte aufblähen, und die Auswahl der Eingriffe, die tatsächlich wirken (S. 85).

Wie groß sind Default-Effekte, und was verändert sie?

§ 5 von 7

Jachimowicz et al. (2019)² definieren den Default-Effekt als Unterschied in der Wahl zwischen der Opt-out- und der Opt-in-Bedingung („the difference in choice between the opt-out condition versus that in the opt-in condition“, S. 162): In der Opt-out-Bedingung ist die Option vorab ausgewählt, und Menschen müssen handeln, um sie abzuwählen; in der Opt-in-Bedingung müssen sie handeln, um sie zu wählen. Ihre Meta-Analyse fasst 58 Datensätze aus 55 Studien in 35 Aufsätzen zusammen, mit insgesamt 73.675 Teilnehmenden, aus einer Literatursuche, die im Mai 2017 endete (S. 159 und 163).

Im Schnitt wurde eine vorab ausgewählte Option um $d = 0,68$ Standardabweichungen häufiger gewählt (95-Prozent-Konfidenzintervall 0,53 bis 0,83), in einem Modell, das mehrere Beobachtungen aus derselben Studie oder demselben Aufsatz berücksichtigt, um 0,63; die Autoren sprechen nach üblichen Kriterien von einem mittleren Effekt („a medium-sized effect given conventional criteria“, S. 164). Cohens d gibt einen Unterschied in

Standardabweichungen an und lässt sich daher nicht direkt mit den Prozentpunkten von DellaVigna und Linos¹ vergleichen. In den 38 Datensätzen mit einer Ja-oder-nein-Zielgröße erhöhten Opt-out-Voreinstellungen die Inanspruchnahme im absoluten Durchschnitt um 27,24 Prozent („by an absolute average of 27.24%“, S. 165). Auch diese Zahl lässt sich nicht mit den 1,4 Punkten vergleichen: Sie stellt eine vorab ausgewählte Option einer gegenüber, die Menschen selbst wählen müssen, in Labor- und Feldstudien, während DellaVigna und Linos einen Nudge mit der üblichen Kommunikation oder mit gar keiner vergleichen, bei staatlichen Leistungen und ohne Voreinstellungen.

Hinter dem Durchschnitt steht eine große Streuung. In 46 Beobachtungen war der Effekt signifikant positiv, in zehn nicht signifikant und in zwei signifikant negativ, und das Heterogenitätsmaß I^2 lag bei 98,21 Prozent, im dreistufigen Modell bei 98,01 Prozent, was als sehr hohe Heterogenität gilt („considered to be very high heterogeneity“, S. 166 und 168). Für eine konkrete Anwendung, warnen die Autoren, kann der Effekt deutlich unter oder über dem Durchschnitt der Meta-Analyse liegen („may be substantially lower or higher than the meta-analytic average“, S. 175). Eine Trim-and-Fill-Analyse schätzte, dass acht Studien fehlen und dass der Durchschnitt mit ihnen auf $d = 0,80$ stiege; wenn überhaupt, fehlen also eher Studien mit größeren Effekten („if anything, default studies finding larger effects are missing from the literature“). Ein Regressionstest auf Asymmetrie im Trichterdiagramm war nicht signifikant, und die Autoren bitten, die Ergebnisse mit Vorsicht zu lesen („be interpreted with caution“, S. 166).

Zwei Gruppen von Faktoren hingen mit der Größe des Effekts zusammen. Voreinstellungen bei Konsumententscheidungen, definiert als Entscheidungen über den Kauf eines Produkts oder einer Dienstleistung („defined as decisions related to buying a product or service“), hatten größere Effekte ($b = 0,73$, $p = 0,003$), solche bei Umweltentscheidungen kleinere, nur auf dem 10-Prozent-Niveau signifikant ($p = 0,08$); kein anderes Studienmerkmal erklärte weitere systematische Streuung („No other study characteristic explained any further systematic variance“), auch nicht, ob eine Studie ein Feld- oder ein Laborexperiment war (S. 168 bis 171). Die Autoren können über die Gründe nur spekulieren („can only speculate“) und vermuten, dass Präferenzen beim Konsum weniger fest sind als Präferenzen in Umweltfragen (S. 176 und 178).

Die zweite Gruppe betrifft, wie eine Voreinstellung wirkt. Die Autoren übernehmen aus früherer Forschung drei Kanäle: Empfehlung

(endorsement), wenn die Voreinstellung als Empfehlung des Entscheidungsarchitekten gelesen wird; Bequemlichkeit (ease), weil es weniger Mühe kostet, bei der vorab ausgewählten Option zu bleiben, als eine andere zu wählen; und Besitzstand (endowment), wenn die Voreinstellung als Status quo empfunden wird. Zwei geschulte Kodierer, die nicht zum Autorenteam gehörten, bewerteten das Design jeder Studie für jeden Kanal. Empfehlung ($b = 0,32$, $p = 0,038$) und Besitzstand ($b = 0,31$, $p = 0,044$) waren signifikante Moderatoren, Bequemlichkeit entgegen der Vorhersage der Autoren nicht („Contrary to our prediction, ease is not a significant moderator“, S. 171 bis 174). Die Autoren nennen ihre Bewertungen nur fehlerbehaftete Näherungen („only noisy approximations“), merken an, dass die zusammengefassten Studien kaum variierten, wie schwer es war, eine Voreinstellung zu verlassen, was den Nullbefund für die Bequemlichkeit zum Teil erklären könnte, und berichten, dass der Besitzstand mit einem robusteren Schätzer nur noch auf dem 10-Prozent-Niveau signifikant war (S. 174 und 176). Sie zitieren eine Studie, in der Voreinstellungen weniger wirkten, wenn die Menschen dem Entscheidungsarchitekten nicht vertrauten, und halten fest, dass Voreinstellungen auf der Annahme beruhen, dass Menschen mit anderen Präferenzen leicht eine Alternative wählen können („allow those whose preferences are different from the default option to easily select an alternative“, S. 172 und 179).

DellaVigna und Linos¹ ließen Voreinstellungen in beiden Stichproben weg. Nimmt man die wenigen Versuchsgruppen mit Voreinstellungen hinzu, werden beide Schätzwerte etwas größer, mit dem Vorbehalt, dass es nur 3 Gruppen in der akademischen und 2 in der Stichprobe der Einheiten sind („with the caveat that these are just 3 arms in the Academic Journals sample and 2 arms in the Nudge Units sample“, S. 97). Sie nennen die institutionellen Grenzen der Einheiten, die Merkmale wie geänderte Voreinstellungen unwahrscheinlicher machen („such as default changes“), als ersten von drei Gründen dafür, dass die 1,4 Punkte wahrscheinlich eine Untergrenze der Wirkung der Verhaltenswissenschaft sind („likely is a lower bound of the impact of behavioral science“); die beiden anderen sind, dass eine Organisation die erfolgreichste Versuchsgruppe einer Studie übernehmen kann und dass Forschende spätere Eingriffe auf den erfolgreichsten Ergebnissen aufbauen können (S. 84). Jachimowicz et al.² ziehen aus der Streuung eine praktische Folgerung: Anwendungen müssen gründlich getestet werden, statt anzunehmen, dass sie immer wie erwartet wirken („one must test applications rigorously, rather than just assuming that they will always work as expected“, S. 177). Um zu erfahren, wie die Präferenzen der Menschen verteilt sind, denen eine Voreinstellung

vorgelegt wird, schlagen sie außerdem eine Bedingung erzwungener Wahl vor, die zeigt, was Menschen ohne Voreinstellung wählen (S. 179).

Welche Änderungen an einer Vertragsverlängerung sind Nudges?

§ 6 von 7

Die Definition dieses Leitfadens gibt einem kaufmännischen Team einen Test, den es vor jeder Studie anwenden kann: Bleibt jede Option offen, kommt kein Rabatt, keine Gebühr und kein Bonus hinzu, und bleiben Preis und andere Anreize im Wesentlichen, wie sie waren? Ein Beispiel ohne Unternehmensdaten: Ein Anbieter von Wartungsverträgen für Industrieanlagen verlängert rund 400 Verträge im Quartal und erwägt fünf Änderungen. Ein einseitiges Verlängerungsschreiben mit Frist und drei Schritten statt des vollständigen Vertragsschreibens ist Vereinfachung und Information. Eine Erinnerung zwei Wochen vor der Frist gehört zu den Erinnerungen und Planungshilfen. Ein Verlängerungsformular, in dem die aktuelle Leistungsstufe des Kunden vorausgewählt ist und jede andere Stufe so leicht wählbar bleibt wie zuvor, ist eine Voreinstellung, die den Status quo abbildet. Drei Prozent Nachlass für eine frühe Verlängerung ändern den wirtschaftlichen Anreiz, die Art finanzieller Anreize, die DellaVigna und Linos¹ aus ihren Stichproben heraushalten, und die Abschaffung der monatlichen Abrechnung streicht eine Option; keine dieser beiden Änderungen ist nach der Definition dieses Leitfadens ein Nudge.

Zur Voreinstellung sagt die Meta-Analyse am meisten, über diese Voreinstellung aber weniger, als es scheinen mag. Eine Voreinstellung, die als Status quo empfunden wird, ist der Fall des Besitzstands, der mit größeren Effekten einherging; doch die zusammengefassten Aufsätze, 34 davon im Literaturverzeichnis mit einem Stern markiert und einer, Dinner et al. (2011), im Forest-Plot genannt, betreffen Entscheidungen wie Organspende, Behandlung am Lebensende, Impfung, Ökostrom und Spenden, und keiner betrifft, nach den Titeln zu urteilen, Geschäftskunden oder eine Vertragsverlängerung. Jede andere Leistungsstufe so leicht wählbar zu lassen wie zuvor, ist, was die Annahme hinter Voreinstellungen verlangt, und das Beispiel ordnet diese Voreinstellung ein, ohne sie zu empfehlen.

Ein Eintrag für den Test einer solchen Änderung hat sechs Felder: die Entscheidung und ihre Optionen, die Änderung und ihren Mechanismus, ob sie überhaupt ein Nudge ist, den Vergleich, den kleinsten Effekt, den der Test nachweisen kann, und was berichtet wird. Die Felder folgen der Definition und Kodierung von DellaVigna und Linos¹, ihrer Beschränkung auf randomisierte Studien mit

Kontrollgruppe, ihren Berechnungen der Teststärke und ihrem Befund zur selektiven Publikation sowie der Forderung von Jachimowicz et al.², Voreinstellungen zu testen, statt ihre Wirkung anzunehmen.

Abbildung 2 Ein Testblatt für einen Nudge

Entscheidung, Änderung, Nudge-Prüfung, Vergleich, nachweisbaren Effekt und Bericht festhalten, bevor der Test beginnt.

Entscheidung	Änderung	Nudge?	Vergleich	Nachweisbar	Bericht
Wessen Wahl, welche Optionen?	Mechanismus oder Voreinstellung?	Optionen erhalten, Anreize gleich?	Zufallszuteilung, Gruppengrößen?	Kleinsten Effekt bei 80 % Teststärke?	Zielgröße, Zeitraum, alle Gruppen?

Jede Gruppe berichten, signifikant oder nicht: Geringe Teststärke verzerrte den publizierten Durchschnitt nur zusammen mit Selektion.

Eine Zeile je vorgeschlagener Änderung. Eine Änderung, die eine Option streicht oder den Preis ändert, wird festgehalten und beiseitegelegt, weil sie kein Nudge ist. Den kleinsten nachweisbaren Effekt angeben, bevor der Test beginnt.

Quelle Entscheidungsblatt des Autors. Die Felder für Entscheidung, Änderung und Nudge folgen DellaVigna und Linos (2022), S. 81, 82 und 89 bis 92; das Feld für den Vergleich folgt DellaVigna und Linos (2022), S. 82 und 85 bis 86, und Jachimowicz et al. (2019), S. 177 und 179; das Feld für den nachweisbaren Effekt folgt DellaVigna und Linos (2022), S. 83 und 98; das Feld für den Bericht folgt DellaVigna und Linos (2022), S. 84, 106 und 111 bis 113. Die Felder sind Eingaben der Lesenden; es werden keine Unternehmensdaten geliefert.

Angenommen, im Beispiel der Verlängerung werden je 200 Kundenkonten zufällig einem der beiden Verfahren zugeteilt, und unter dem bisherigen Verfahren werden 80 Prozent der Verträge fristgerecht verlängert. Bei den 80 Prozent Teststärke, mit denen die Autoren rechnen, und einem zweiseitigen Test auf dem 5-Prozent-Niveau liegt der kleinste nachweisbare Unterschied bei rund 11 Prozentpunkten ($2 \cdot 8 \times 2 \times 0,8 \times 0,2/200$ approximately $0,112$), und ein Unterschied müsste rund 7,8 Punkte erreichen, um überhaupt signifikant zu sein. Beides ist ein Vielfaches des Durchschnitts der Einheiten von 1,4 Punkten. Würden Nudges bei Verlängerungen Geschäftskunden in

ähnlichem Maß bewegen, was die Quellen nicht prüfen, könnte ein Test über ein Quartal keinen davon nachweisen, und ein Ergebnis, das doch signifikant würde, hätte das Profil, vor dem die Autoren bei publizierten Studien mit geringer Teststärke warnen: groß und wahrscheinlich nicht in derselben Größe replizierbar („unlikely to replicate at the same effect size“, S. 85). Jede Versuchsgruppe zu berichten, signifikant oder nicht, hält die eigenen Aufzeichnungen eines Teams frei von der Selektion, die den publizierten Durchschnitt aufgebläht hat. Das sind Beispiele des Autors, sie sind nicht geprüft.

Wie weit trägt diese Evidenz bei B2B-Verträgen?

§ 7 von 7

Ohne eigenen Test nicht weit. DellaVigna und Linos (2022)¹ untersuchten staatliche Leistungen in den USA, mit Zielgrößen wie Impfung, Anmeldung zu Leistungen, Zahlung von Rechnungen und Registrierung, meist per Brief, E-Mail oder Postkarte; sie halten 1,4 Punkte für eine vernünftige Schätzung der durchschnittlichen Wirkung eines Nudges im breiten Einsatz bei staatlichen Leistungen („a reasonable estimate for the average impact of a nudge at scale on government services“, S. 84). Sie beobachteten die Einzeldaten der Studien nicht („do not observe the micro-data for each trial“), ihre publizierte Stichprobe stammt aus zwei Meta-Analysen mit denselben Einschränkungen, und sie warnen davor, die Unterschiede zwischen Kategorien kausal zu lesen (S. 82, 102 und 114). Jachimowicz et al. (2019)² fassten 58 Vergleiche von Opt-in und Opt-out aus Labor- und Feldstudien zusammen; ihre Bewertungen der Kanäle sind Urteile von Kodierern, und ihre Studienmerkmale spiegeln die Entscheidungen der Forschenden darüber, welche Studien sie durchführen und wie sie die Voreinstellung umsetzen („reflect the decisions of researchers about what studies to conduct and how to implement the default intervention“, S. 176). Keine der beiden Quellen hat Geschäftskunden, Buying Center oder Vertragsverlängerungen untersucht. Der Leitfaden zur externen Validität zeigt, wie man Population, Intervention, Ergebnis und Kontext benennt, auf die ein Ergebnis übertragen werden soll.

Nach der Lesart dieses Leitfadens lassen sich drei Dinge mit einiger Zuversicht übertragen. Die eigene Definition des Leitfadens sortiert vorgeschlagene Änderungen, bevor jemand sie testet: Ein Rabatt oder eine gestrichene Option ist kein Nudge. Ein großer Effekt aus einem kleinen Test verdient Zweifel, wenn er das Ergebnis ist, das übrig blieb (die signifikante Versuchsgruppe oder die Studie, die aufgeschrieben wurde), denn in den Simulationen der Autoren hob geringe Teststärke den Durchschnitt nur zusammen mit dieser Selektion an; die 1,4 Punkte

der Einheiten zeigen, wie weit ein Durchschnitt im breiten Einsatz bei staatlichen Leistungen unter einem publizierten lag, nicht, was ein Nudge bei Verlängerungen bewirkt. Und der Effekt einer Voreinstellung schwankt zu stark, um ihn vorauszusetzen; er muss dort gemessen werden, wo sie eingesetzt wird. Ob Nudges bei Verlängerung und Abrechnung Geschäftskunden in ähnlichem Maß bewegen wie bei staatlichen Leistungen und ob der Befund zu Konsumententscheidungen für Voreinstellungen auch für Unternehmen gilt, sagen die Quellen nicht.

 Ende des Beitrags

Referenzen

Zitate stehen in der Sprache des Originals; die deutschen Wiedergaben im Fließtext sind sinngemäß und stammen vom Autor.

- 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>

Zitierweise

Isoglu, S. (2026, 18. Oktober). Was ist ein Nudge? Sinan Isoglu.
<https://isoglu.com/de/insights/journal/was-ist-ein-nudge/>
Online-Fassung: <https://isoglu.com/de/insights/journal/was-ist-ein-nudge/>

Sinan Isoglu, MBA (Quantic)

Führungskraft für Umsatzwachstum, Dozent und Doktorand
© 2026 Sinan Isoglu · isoglu.com · Alle Rechte vorbehalten