Assessing Nudge Impact: A Comprehensive Second-Order Meta-Analysis

Bo Hu [a †], Ziqian Xia [b †], Qingcheng Guo [c], Chang Lu [c d], Sara Constantino [b], Xingda Ju [c d]

† Contributed equally

[a] School of Social and Behavioral Science, Nanjing University, China

[b] Doerr School of Sustainability, Stanford University, United States

[c] School of Psychology, Northeast Normal University, China

[d] Jilin Provincial Key Laboratory of Cognitive Neuroscience and Brain Development, China

This manuscript has been accepted by the *Journal of Behavioral Decision Making* and can be accessed at the following link: https://doi.org/10.1002/bdm.70053.

Correspondence concerning this article should be addressed to:

Xing-Da Ju, School of Psychology, Northeast Normal University, Changchun 130024, China, juxd513@nenu.edu.cn.

Chang Lu, School of Psychology, Northeast Normal University, Changchun 130024, China, luc816@nenu.edu.cn.

Data Availability Statement

This study has been pre-registered (https://osf.io/nmpqv). The analysis script and data during this study were available in the Open Science Framework repository (https://osf.io/ugnf9/?view_only=577ccdb3f1194ff3984e82253947a05d).

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Statement

Not Application.

Author contribution statement

Bo Hu: Conceptualization, Investigation, Software, Formal analysis, Writing – original draft, Writing - Review & Editing.

Ziqian Xia: Conceptualization, Investigation, Formal analysis, Writing – original draft, Writing - Review & Editing.

Qingcheng Guo: Investigation, Writing – original draft, Writing - Review & Editing.

Chang Lu: Conceptualization, Supervision, Investigation, Writing – original draft, Writing - Review & Editing, Project administration.

Sara Constantino: Conceptualization, Investigation, Writing - Review & Editing

Xingda Ju: Conceptualization, Supervision, Investigation, Writing – original draft, Writing - Review & Editing, Project administration.

Abstract

Nudging as a strategy to alter behaviors has garnered increasing attention from both researchers and policymakers. Here, we conduct a second-order meta-analysis, synthesizing 13 articles (14 meta-analyses) that include 1,638 primary studies and approximately 30 million participants. We find a small aggregated effect size across these meta-analyses ($d = 0.27$, 95% CI [0.16, 0.38]), which drops to $d = 0.004$ after adjusting for publication bias. Examining the methodological quality of the meta-analyses, we find that most were rated as low or critically low, suggesting that our findings, which inherit these limitations, should be interpreted with caution. This study provides the most comprehensive synthesis of the effectiveness of nudging to date, while underscoring the urgent need for higher quality, preregistered meta-analyses to clarify the true impact.

Keywords: Nudging, Behavior intervention, Choice architecture, Second-order meta-analysis

Introduction

Thaler and Sunstein (2008) introduced the concept of *Nudging*, which is the practice of altering behavior by designing the physical or psychological environment in which people make decisions while preserving their freedom of choice (i.e., all choice options remain available and there is no coercion). Nudging has gained popularity as a method to influence behavior because it tends to be low-cost and upholds individual autonomy, in contrast to more traditional approaches such as incentives or mandates, and may also be less politically divisive (Chater & Loewenstein, 2023a; Hertwig & Grüne-Yanoff, 2017; Lin et al., 2017). The theory supporting nudges is also built on a rich foundation of studies from psychology and behavioral economics (Johnson & Goldstein, 2003; Kahneman & Tversky, 1988; Tversky & Kahneman, 1981). For example, one widely cited theoretical foundation for nudging is the dual-process theory (Evans, 2008), which distinguishes between automatic, intuitive decision-making (System 1) and deliberate, analytical decision-making (System 2). While traditional interventions often target System 2 by providing information or incentives and assuming rational deliberation, nudges typically aim to influence System 1 processes by shaping the choice environment in ways that guide behavior without requiring conscious effort (Kahneman, 2011).

Practical applications of nudging have demonstrated notable success in various public projects, including tax reforms, smoking cessation campaigns, and climate change mitigation (Halpern, 2015; He et al., 2023). For example, by changing the default option settings to promote pro-environmental behaviors (Berger et al., 2022; He et al., 2023). Governments worldwide have recognized the potential of nudging, leading to the establishment of hundreds of dedicated nudge units around the globe to apply behavioral insights in policy (OECD, 2017).

However, while nudging has been widely adopted as a tool by academics and governments worldwide (Afif et al., 2019; OECD, 2017; WHO & World Bank, 2024; World Bank, 2015), its effectiveness in changing behaviors has come under some scrutiny in recent years (Lin et al., 2017; Schubert, 2017; Sunstein, 2015). And this scrutiny has intensified following the publication of meta-analyses finding small to no effects (DellaVigna & Linos, 2022; Maier et al., 2022; Mertens et al., 2022). Indeed, a recent paper by two scholars who have spent much of their careers advancing these types of solutions to social problems recently wrote a paper arguing that nudges often "yield small or null results" and "their impact may be modest" (Chater & Loewenstein, 2023a, 2023b; Connolly et al., 2025).

Achieving a comprehensive understanding of nudging's effectiveness is essential. We conduct a second-order-meta-analysis that includes 1,638 studies with over 30 million participants to provide (1) a comprehensive assessment of the overall effectiveness of nudge interventions, and (2) identify which conditions or study characteristics, if any, predict greater effectiveness.

Potential moderators of nudge interventions

Because our approach is broad, spanning many studies and disciplines, evaluating the influence of potential moderators is critical to account for heterogeneity across the studies. Numerous factors could moderate the effectiveness of nudge interventions. Our analyses focus on the subset of factors that can be examined with available data from the primary meta-analyses we have included in this study—these are the dimensions that we discuss below. Other contextual factors (e.g., cultural differences, duration of intervention, etc.) or study-level differences in implementation or samples might also influence nudge outcomes, even if they are not captured in our data.

First, past meta-analyses have found that different types of nudges have different effect sizes (Broers et al., 2017; Mertens et al., 2022; Zhang et al., 2023). However, authors classify nudges in different ways (e.g., by their behavioral strategy vs. by their psychological mechanism; Dolan et al., 2012; M. Berger et al., 2022), making comparisons difficult. In this study, we adopt the taxonomy by Münscher et al. (2016), which groups nudges into three broad categories—*decision information* (e.g., providing information or feedback), *decision structure* (e.g., changing default options or choice layouts), and *decision assistance* (e.g., reminders and prompts). Decision information nudges enhance information availability, comprehensibility, and personal relevance. Decision structure nudges alter the choice options within individual decisions, including default options, changing choice efforts, altering choice ranges, and modifying choice consequences. Decision assistance nudges facilitate self-regulation, including by providing reminders and promoting commitments. Previous studies have indicated that the three types of nudges differ in their effectiveness (Mertens et al., 2022; Vlaev et al., 2016).

Second, the effectiveness of nudges may vary by behavioral domain. Some analyses suggest that certain domains, such as energy conservation or privacy, respond differently to nudges than others (Hummel & Maedche, 2019), partly because the instruments differ across domains: privacy nudges are typically warnings or salience cues; health nudges often reduce effort; and finance mainly uses reminders and defaults. In addition, one large meta-analysis found that nudges in the food domain yielded particularly high effects compared to other areas (Mertens et al., 2022), consistent with food choices' low-cost, lowstakes, habit andcue-driven nature, which makes them more sensitive to structural (choice-architecture) changes than higher-stakes domains (e.g., personal finance). In our study, we examine four domains—health, food, environment, and finance—because they represent the most frequently studied and operationally mainstream applications of nudges, account for 92% of the studies we analyze. Where meta-analyses covered more than one domain, we coded the four domain categories and carried forward the reported domain-specific estimates (from subgroups) for the moderation analysis.

Third, the setting of the experiment (lab vs. field) might influence nudge effectiveness. Laboratory studies typically minimize extraneous variation and enhance adherence, reduce measurement noise, and increase effect sizes. By contrast, field experiments have greater external validity but can also limit full control over the environment and introduce implementation challenges, noncompliance, spillovers, and contextual shocks that can dilute effects. Consistent with these mechanisms, evidence syntheses report larger impacts in published academic field Randomized controlled trials (RCTs) than in government-run, at-scale field trials (DellaVigna & Linos, 2022), echoing long-standing discussions about the translation of research from the lab to the field. Thus, while one might expect attenuation of effects outside the lab, the magnitude, variability, and contribution to heterogeneity of nudge effectiveness remain an open question.

Finally, the study design could play a role. RCTs with clear control groups are considered a gold standard for causal inference (Oxford Centre for Evidence-Based Medicine, 2011). In nudging research, however, many studies use pre-post designs or other less rigorous designs. Without a control, apparent improvements can simply reflect background change, measurement drift, or statistical artefacts rather than true intervention effects (Schweizer et al., 2016; Shadish et al., 2002). Thus, one might expect that nudging effects are more precisely estimated and also smaller in studies with stronger designs than in those without proper control groups. We consider whether the study's design (presence or absence of control groups) moderates the estimated effect sizes.

Note that while the first two sources of heterogeneity – nudge type and behavioral domain – are about differences in the effect of the intervention on behavior change, the last two sources of heterogeneity are more about differences in effects due to study design or measurement and analysis choices.

The present study

Prior evidence syntheses on nudges span both domain-specific and cross-domain meta-analyses. Food-focused reviews (Arno & Thomas, 2016; Broers et al., 2017; Cadario & Chandon, 2017; Zhang et al., 2023) typically report small-to-moderatepositive effects—especially for decision-structure interventions such as defaults, portioning, placement, and other effort-reducing changes. Health-focused syntheses often find smaller or null averages when interventions rely primarily on decision-information tools (e.g., labels, messages) rather than structural changes (Cadario & Chandon, 2017; Mertens et al., 2022). Finance-related reviews more often examine reminders and defaults, with effects that are generally smaller than in food (DellaVigna & Linos, 2022; Jachimowicz et al., 2019). Cross-domain meta-analyses (e.g., Hummel & Maedche, 2019; Mertens et al., 2022) reinforce these patterns: decision-structure nudges tend to outperform decision-information and decision-assistance categories on average, and context (lab vs. field; academic vs. at-scale implementation) matters—field trials run by government "nudge units" typically report attenuated impacts relative to published academic field RCTs (DellaVigna & Linos, 2022).

These reviews motivate a second-order synthesis that can (a) aggregate across meta-analyses, (b) probe heterogeneity along theoretically salient dimensions (nudge type, domain, setting), and (c) address bias more systematically.

Second-order meta-analyses are comprehensive assessments that provide an overview of existing systematic reviews and meta-analyses (Ioannidis, 2009; Papatheodorou, 2019), allowing for a more extensive evaluation of the overall effectiveness of nudging in the extant literature. By synthesizing published meta-analyses, we maximize the use of available evidence and can assess patterns at the meta-analytic level, offering a complementary perspective to first-order meta-analyses. Rather than re-estimating a single pooled effect from primary studies, second-order meta-analyses allow us to address review-level questions such as: *Do recent meta-analyses converge in their conclusions? How large is the between-review heterogeneity? What is the methodological quality of these reviews? How much study overlap do they share?* Compared to previous meta-analyses, our second-order meta-analysis synthesizes 13 published meta-analyses, encompassing 1,638 primary studies and roughly 30 million participants across multiple domains. This substantially expands the time horizon, sample size, and domain coverage beyond any single prior synthesis and allows us to examine whether the latest literature, particularly pandemic-era and climate-relevant nudges, has shifted average effects or credibility. Taken together, this approach provides an updated and broader assessment of the effectiveness and limitations of nudging, complementing first-order meta-analyses by evaluating the robustness of conclusions across multiple meta-analyses and offering decision-relevant prediction intervals for future applications.

Method

This study has been preregistered, and the analysis script and data were available in the Open Science Framework (https://osf.io/ugnf9/?view_only=577ccdb3f1194ff3984e82253947a05d).

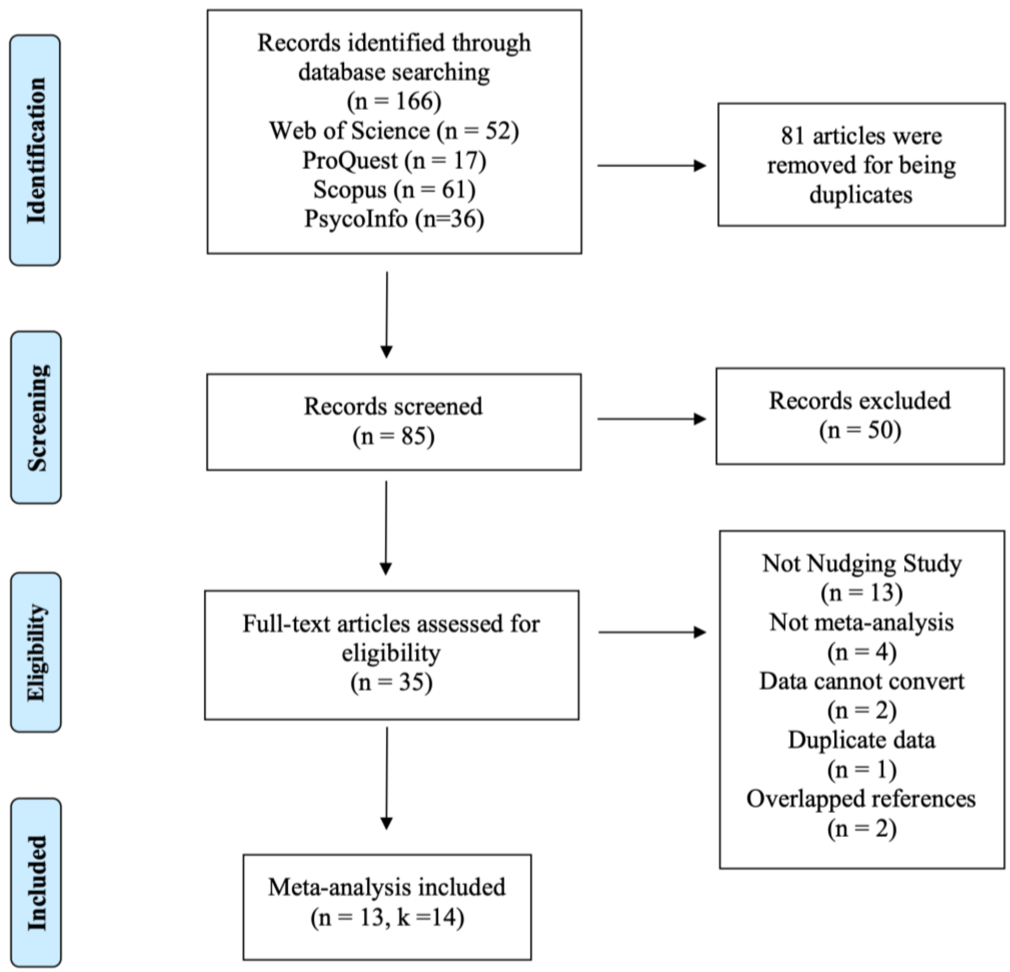
2.1 Search strategy

Following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines (Page et al., 2021), a systematic search was conducted on September 27, 2023, across four databases: Web of Science, Proquest, Scopus, and PsycInfo.

Building upon the retrieval methods of existing studies (Bergquist et al., 2023; Mertens et al., 2022), this study employed the following search strategy: "Nudge" OR "Nudging" OR "Choice architecture" AND "Meta-analysis" OR "Research synthesis" OR "Quantitative review" OR "Meta-analytic structural equation modeling" OR "Meta-analytic SEM" OR "MASEM" OR "Meta-analytic path analysis" OR "Meta-regression" OR "Cumulative meta-analysis" OR "Mega-analysis" OR "Bayesian meta-analysis" OR "Second-order meta-analyses" OR "Secondary use of meta-analytic data."

The titles and abstracts were reviewed independently by two evaluators to select potentially relevant meta-analyses. In cases of disagreement, consensus was reached through discussion with a third author. After the title and abstract screening, two evaluators conducted further screening by reading the full texts to identify eligible meta-analyses for inclusion in the Second-order meta-analyses. The research flowchart is shown in Figure 1. At the full-text screening stage, we excluded two meta-analyses owing to very high overlap with other included reviews (Jachimowicz et al., 2019; Ratnayake et al., 2023) and one meta-analysis whose set of primary studies was entirely redundant with another review (Cadario & Chandon, 2017). We also excluded two meta-analyses that reported effects as weighted mean differences or compliance effects but did not provide sufficient statistics for conversion to a common effect-size metric (Antinyan & Asatryan, 2020; R. Li et al., 2021).

Figure 1. Literature search flowchart



Note: n denotes the number of included articles; k denotes the total number of meta-analyses across those articles.

2.2 Inclusion criteria

The inclusion criteria were established to ensure the inclusion of relevant meta-analyses in the analysis. The inclusion criteria were as follows:

(a) The meta-analysis focused on the effectiveness of ‘nudging’ interventions, as defined by Thaler and Sunstein (2008). Following the modal formulations across those reviews, such interventions steer behavior by changing the choice architecture while preserving decision makers’ freedom of choice. Because terminology varies across disciplines, we adopt a descriptive harmonization approach: we do not overwrite first-order authors’ labels ex post, but instead catalogue their definitions and report the range of included formulations (see Table S1). This approach allows our second-order synthesis to reflect how “nudging/choice architecture” is actually operationalized in the published meta-analytic literature.

(b) The meta-analysis provided quantitative data, enabling the calculation of effect sizes;

(c) Meta-analyses covering studies published before 2008 were excluded, since the concept of ‘nudging’ was formally introduced in 2008;

(d) Meta-analyses that did not reference Thaler and Sunstein (2008) or otherwise clearly indicate they were examining *nudging* (as opposed to generic behavioral interventions) were excluded, to ensure relevance to the nudging literature;

(e) When two meta-analyses covered overlapping primary studies and the pairwise overlap exceeded 15%—the very high overlap—we retained the review that was more recent, included more primary studies, or had the higher study quality rating.

2.3 Quality assessment

The quality of the included meta-analyses was independently assessed by two evaluators using A Measurement Tool to Assess Systematic Reviews (AMSTAR-2, Shea et al., 2017). AMSTAR-2 consists of 16 items, with questions 2, 4, 7, 9, 11, 13, and 15 considered as critical domains, covering protocol preregistration, a comprehensive literature search, a justified list of exclusions, risk-of-bias appraisal for included studies, appropriate meta-analytic methods, consideration of risk-of-bias in interpreting findings, and assessment of publication bias. This assessment of quality ensures that the selected meta-analyses meet established standards and have been rigorously evaluated for their methodological rigor and reliability.

The quality of the meta-analyses was assessed based on the following criteria:

(a) High quality: No critical flaws or only one non-critical weakness;

(b) Moderate quality: More than one non-critical weakness;

(c) Low quality: One critical flaw with or without non-critical weaknesses;

(d) Critically low quality: More than one critical flaw with or without non-critical weaknesses.

2.4 Data synthesis

In this meta-analysis, we aimed to assess the effectiveness of nudging interventions in influencing behavioral outcomes. In line with recent recommendations (Funder & Ozer, 2019), an effect size r of .05 ($d \approx 0.10$) can be very small for single events but consequential in the not-very-long run[1]; $r = .10$ ($d \approx 0.20$) remains small at the single-event level but more consequential over time; $r = .20$ ($d \approx 0.41$) represents a medium effect with explanatory and practical value even in the short run; and $r = .30$ ($d \approx 0.63$) reflects a large effect potentially powerful both short- and long-term. We coded the direction of behavior change

[1] For example, a person who routinely comes home from a disliked job in a state of fatigue may be slightly more likely ($r = .05$) to snap during stressful conversations with a spouse; even if self-control replenishes overnight, that small daily risk can accumulate across many occasions into meaningful marital strain.

in Cohen's d to reflect the expected intervention effect. In our analysis, a positive Cohen's d value indicates a positive effect, while a negative value represents a negative effect. Existing studies have found large heterogeneity in nudge study effects (Mertens et al., 2022; Szaszi et al., 2022). We thus used random-effects models to re-estimate pooled standardized mean differences with 95% CI.

Heterogeneity among the included meta-analyses was assessed using Cochran's Q statistic and the I^2 statistic (Higgins et al., 2003). Cochran's Q assesses whether the observed differences in effect sizes are greater than expected by chance, while the I^2 statistic provides a percentage estimate of the total variation across meta-analyses, helping quantify the degree of inconsistency among the results. Because individual study-level data were unavailable for most of the included meta-analyses, we conducted heterogeneity analyses at the meta-analysis level. This means that rather than pooling raw study-level effect sizes into a unified first-order meta-analysis, we extract moderator results as reported in each meta-analysis and synthesize these aggregate estimates across the set of meta-analyses.

We implemented a robust Bayesian model-averaged meta-analysis (RoBMA), which effectively adjusts for publication bias (Bartoš et al., 2022). This approach combines selection models and PET-PEESE via Bayesian model averaging, yielding bias-adjusted, model-averaged effect sizes and Bayes-factor evidence. By integrating different presumptions about effect size, heterogeneity, and publication bias from prior distributions, we were able to reliably determine the effect size (Berkhout et al., 2023; Gronau et al., 2021). In essence, the RoBMA approach does not assume a single fixed level of publication bias; instead, it averages over multiple bias scenarios with associated prior probabilities (with a higher prior weight on the possibility of no bias). The Appendix describes the model settings of RoBMA, which include a discrete mixture of publication bias models ranging from no bias to high bias.

We also applied frequentist bias correction techniques (Egger's regression test and PET-PEESE) independently of the RoBMA to cross-verify the evidence of publication bias (Duval & Tweedie, 2000; Egger et al., 1997; Stanley & Doucouliagos, 2014). The PET-PEESE was done as a separate analysis on the aggregated data, not within the Bayesian model. Small-study effects bias was considered present when the regression asymmetry test produced a p-value of 0.05 or less.

2.6 Study overlap

Second-order meta-analysis incorporates a large number of systematic reviews on the same topic, which is likely to lead to overlap in primary studies across the reviews (Lunny et al., 2021). We can calculate the corrected covered area (CCA) for the entire matrix of included reviews (Hennessy & Johnson, 2020). The CCA represents the degree of study overlap between included meta-analyses and can range from 0 to 100%; higher values indicate substantial re-use of the same primary studies, whereas lower values indicate more independent evidence coverage. Consistent with Pieper et al. (2014), which categorize CCA > 15% overlap as very high, we applied >15% as the exclusion threshold for meta-analyses. The formula used to calculate the CCA is

$$\frac{N - r}{(r \times c) - r}$$

where N is the total number of included publications (including double counting) in evidence synthesis (this is the sum of the ticked boxes in the citation matrix); r is the number of rows (number of index publications); and c is the number of columns (number of reviews).

When two meta-analyses overlapped by more than 15%, we retained the one that was (a) more recent, (b) included more primary studies, or (c) had a higher quality (AMSTAR-2) score. Using this approach, we identified two overlapping cases and excluded two meta-analyses: specifically, Jachimowicz et al. (2019) overlapped substantially with the more recent default-nudge meta-analysis by Zhao et al. (2022) and was therefore excluded, and Ratnayake et al. (2023) overlapped with other meta-analyses on similar topics (Slapo et al., 2021) and was excluded to avoid redundancy. After these exclusions, the final Corrected Covered Area (CCA) was 1.17% (Figure S1), indicating minimal overlap among the included reviews (Lunny et al., 2021).

2.5 Data extraction

For each meta-analysis included in this study, two evaluators independently coded the data. The data coding process involved extracting the following information: authors, publication year, intervention effect (including the effect size and 95% CI), number of included articles, number of included studies, sample size, and moderator variables.

Result

We included 13 articles (published between 2016 and 2023) in our review (Arno & Thomas, 2016; Broers et al., 2017; Cadario & Chandon, 2017; DellaVigna & Linos, 2022; Hummel & Maedche, 2019; Ioannou et al., 2021; Y. Li et al., 2023; Mertens et al., 2022; Slapo et al., 2021; Waheed, 2023; Wyse et al., 2021; Zhang et al., 2023; Zhao et al., 2022), with one article (DellaVigna & Linos, 2022) containing two distinct meta-analyses, yielding 14 meta-analyses estimates in total (see Table 1 for an overview of included meta-analyses). Across these meta-analyses, a total of 655 primary articles (encompassing 1,638 studies) were analyzed, with an aggregate sample size of approximately 29.9 million participants. Among these meta-analyses, five focused on the food domain, three were related to the health domain, one focused on the privacy domain (disclosure of personal information), and six were not restricted to a single domain.

Table 1. Characteristics of the included meta-analyses

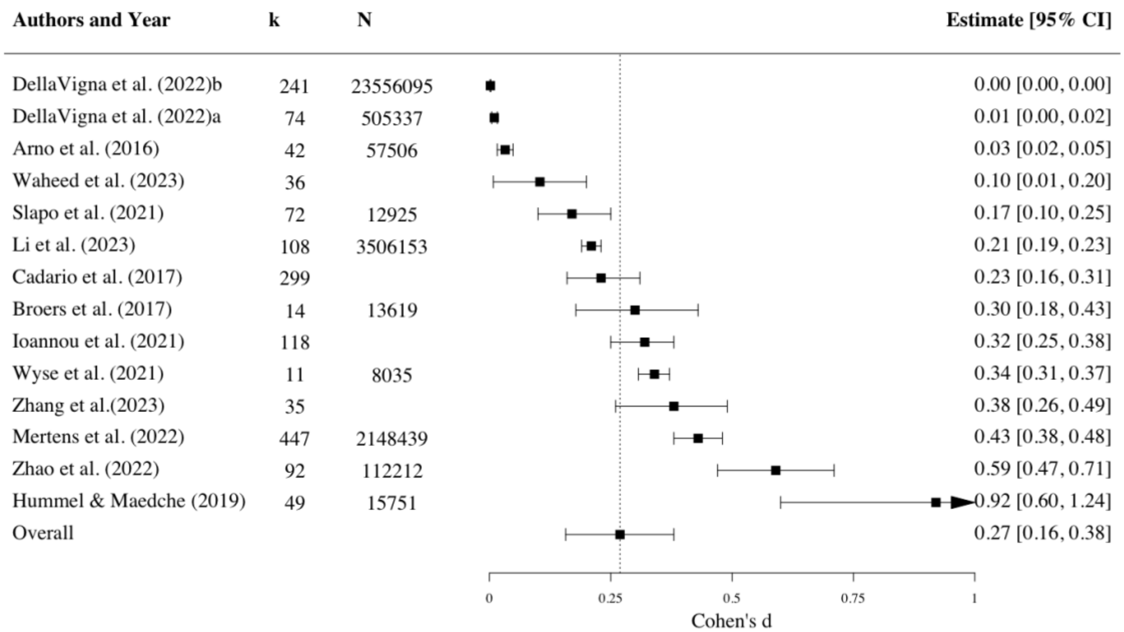
Study name	Article (study)	Sample	Percent of studies with control groups	Nudge type	Experiment	Domain
Arno et al. (2016)	36(42)	57506	74%	Multi	Multi Lab and Field	Health
Broers et al. (2017)	11(14)	13619	39%	Multi	Lab and Field	Food
Cadario et al. (2017)	92(299)	NA	41%	Multi	Field	Food
DellaVigna et al. (2022)a	26(74)	505337	100%	Multi	Lab and Field	Multi
DellaVigna et al. (2022)b	121(241)	23556095	100%	Multi	Field	Multi
Hummel & Maedche (2019)	11 (49)	15751	100%	Multi	Lab and Field	Multi
Ioannou et al. (2021)	54(118)	NA	NA	Multi	Lab and Field	Privacy
Li et al. (2023)	32(108)	3506153	100%	Multi	Lab and Field	Multi
Mertens et al. (2022)	214(447)	2148439	100%	Multi	Lab and Field	Multi
Slapo et al. (2021)	30(72)	12925	78%	Multi	Lab and Field	Food
Waheed et al. (2023)	14(36)	NA	100%	Multi	Lab and Field	Health
Wyse et al. (2021)	11(11)	6360	100%	Multi	Lab and Field	Food
Zhang et al.(2023)	22(35)	NA	23%	Multi	Lab and Field	Food
Zhao et al. (2022)	58(92)	112212	100%	Default	Lab and Field	Multi

Note: NA indicates that the literature did not report the relevant information. Multi indicates that the meta-analysis examined multiple moderator categories (i.e., included different types of moderators

Using a conventional random-effects model (Figure 2), we find an overall mean effect size of d = 0.27 (95% CI [0.16, 0.38], $p < .001$), indicating a small but statistically significant positive effect of nudge interventions. There was substantial heterogeneity among these meta-analyses ($Q = 1476.10$, $p < .001$, $I^2 = 99.89\%$). The high heterogeneity likely reflects variation across behavioral domains, nudge categories, study design, and experimental settings included in the underlying reviews. It is important to note that the random error in these analyses is likely to be smaller than usual because the effects of aggregating many studies tend to stabilize the variability. The quality assessment using AMSTAR-2 revealed that one meta-analysis was of moderate quality, three were of low quality, and nine were critically low quality (Table S2). Very few meta-analyses were pre-registered (only 3 of 13), and few conducted a risk of bias assessment of primary studies (only 4 of 13). When critically low-quality studies were excluded and the pooled effect was re-evaluated, d = 0.29, 95% CI [0.21, 0.36].

A Robust Bayesian Meta-Analysis (RoBMA) (Figure S2), which accounts for potential publication bias by averaging across multiple bias models, yielded a very similar model-averaged effect size (d = 0.29, 95% CI [0.15, 0.43]). This convergence suggests that our overall estimate is robust to the different analytic approaches (at least before additional bias correction mechanisms are applied). The estimate for heterogeneity was τ=0.26, 95% CI [0.18; 0.39].

Figure 2. Forest plot of the effects of the nudging intervention



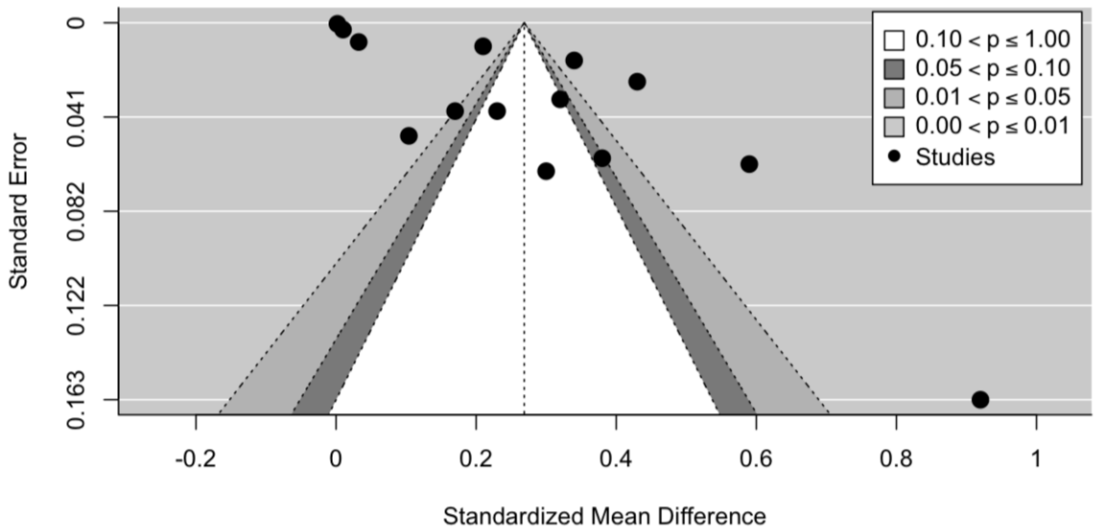
Moderator analysis was conducted for 12 meta-analyses, including nudging categories, behavioral domains, experimental types, study design, and location. As shown in Table 2, the results indicated that nudging categories (Q=2.84, p=.242), behavioral domains (Q=4.42, p=.220), experimental types (Q=0.67, p=.413), and study design (β=0.004, p=.986) did not significantly impact the effectiveness of nudging.

Table 2. Moderator effects test of the nudging intervention

Moderator		Include meta-analysis	*d*	Lower limit	Upper limit	*k*
Main effect		13	0.27	0.16	0.38	1638
Category	Decision information	11	0.23	0.16	0.30	360
	Decision structure	11	0.40	0.23	0.57	365
	Decision assistance	6	0.30	0.09	0.52	123
Domain	Environment	2	0.45	0.35	0.54	98
	Health	4	0.17	-0.01	0.36	166
	Finance	2	0.51	-0.43	1.45	53
	Food	5	0.33	0.18	0.47	510
Experiment	Field	6	0.63	-0.08	1.34	755
	Laboratory	2	0.52	0.36	0.67	157

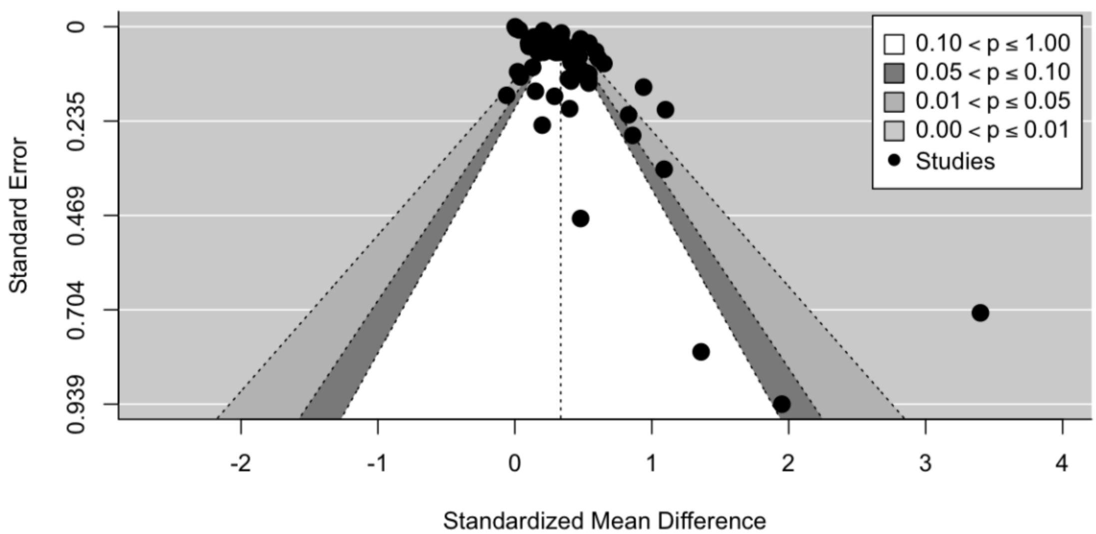
We next assessed potential publication bias in our meta-analysis. Funnel plot asymmetry suggested that smaller-sample meta-analyses tended to report higher effect sizes (Figure 3), consistent with publication bias. Correspondingly, Egger's regression test was significant ($z = 4.25$, $p < .001$), indicating bias. Applying the PET-PEESE correction (which adjusts for small-study effects via a regression of effect size on standard error), the estimated true effect of nudging was essentially null (adjusted $d = 0.003$), reinforcing the evidence that the positive overall effect may be driven by publication bias.

Figure 3. Funnel plot of nudging intervention



Second, we used all 59 subgroup effects (consisting of 57 subgroup effects and two mean effect sizes from meta-analyses that did not report subgroup effects) to further assess publication bias (Bergquist et al., 2023; Busch & Friede, 2018). As shown in Figure 4, funnel plots revealed a potential risk of publication bias. The trim-and-fill method added 12 studies; overall effect decreased from d = 0.34, 95% CI [0.27, 0.39] to d = 0.27, 95% CI [0.20, 0.34] when adjusting for publication bias. The Egger test indicated significant publication bias (z=5.92, p<.001). In the PET-PEESE test, when correcting for small-sample effects, the nudging interventions under study have no effect (d= 0.004).

Figure 4. Funnel plot of nudging intervention by subgroup effect size



Finally, we assessed the publication bias of the included meta-analyses. A total of 11 meta-analyses reported their publication bias (Table S3); six meta-analyses reported potential or significant publication bias, and five meta-analyses reported minimal publication bias. We evaluate a new second-order meta-analysis using adjusted effect sizes from three meta-analyses (Slapo et al., 2021; Wyse et al., 2021; Zhang et al., 2023). The results showed that the overall effect increased from d = 0.27, 95% CI [0.16, 0.38] to d = 0.28, 95% CI [0.16, 0.41].

The publication bias analyses above focus mainly on intervention publication bias. We also attempted to mitigate the impact of pilot publication bias, which occurs when only successful pilot studies are selected for further research and publication, leading to a biased view of the overall effectiveness of the intervention(von Klinggraeff et al., 2023). During the literature search, we attempted to include grey literature, such as dissertations, conference papers, and preprints. In addition, we included government work reports to reduce potential bias (DellaVigna & Linos, 2022).

Discussion

This study synthesizes 13 articles (14 meta-analyses) covering 1,638 primary studies and nearly 30 million participants to reassess the effectiveness of nudging. The unadjusted pooled estimate indicates a small positive effect. However, multiple diagnostics for small-study and selection bias consistently attenuate this estimate toward zero. There are no significant differences were found across moderators.

Our unadjusted pooled estimate showed a significant effect of nudging interventions (d = 0.27, 95% CI [0.16, 0.38]), an effect in the lower-small range for single instances but one that may still accumulate into meaningful consequences when applied widely or repeatedly. However, publication bias analysis indicated that the true effect may be null. It is worth noting that this effect size is smaller than in previous meta-analyses across various behavioral domains (d = 0.43, Mertens et al., 2022). There are several possible reasons for this difference: on one hand, recent meta-analyses of nudging have reported lower intervention effects (Ratnayake et al., 2023; Slapo et al., 2021; Waheed, 2023), suggesting that the effectiveness of nudging may have been overestimated previously. On the other hand, this study includes more research results from Nudging Units, which often provide more field experiments and contain lower effect sizes compared to studies published in academic journals (DellaVigna & Linos, 2022). Our definitional pluralism in the underlying literature: the included first-order meta-analyses use overlapping but not identical formulations of nudging/choice architecture, and our second-order conclusions therefore speak to this broader, as-used construct rather than to a single prescriptive definition.

After correcting for publication bias, the aggregated effect of nudging was rendered virtually zero. This finding aligns with prior warnings that substantial publication bias in the nudging literature may have inflated effect size estimates and impeded a clear understanding of nudging's true impact (Maier et al., 2022). While methods like PET-PEESE have their limitations and can potentially overcorrect (Sala et al., 2019). These results should be interpreted with caution, considering that there may be other unmeasured factors (e.g., varying contexts, intervention fidelity, or publication biases in the original studies) that influence nudge effectiveness.

Because most of the meta-analyses we synthesized were rated as low or critically low quality, our second-order meta-analysis inherits those weaknesses. This means our aggregate findings should be interpreted with caution. Even though our second-order meta-analysis is constrained by the generally low methodological quality of the input studies, it remains informative in several important respects. First, it reveals a consistent pattern across the included meta-analyses: most reported positive effects of nudging, but these effects were likely inflated by publication bias, with bias-corrected estimates converging near zero. Second, it highlights systemic issues—such as pervasive bias, extreme heterogeneity, and uneven methodological rigor—that cut across behavioral domains and study designs. Importantly, the alignment of our conclusions with those reached by independent re-analyses suggests that these patterns are robust and reflect genuine features of the nudging evidence base rather than artefacts of our synthesis (Maier et al., 2022).

Our moderator analyses did not detect any statistically significant differences in effect size by nudge type, behavioral domain, experimental setting, or study design. A key limitation of our moderator analysis is that it was performed at the meta-analysis level rather than at the individual-study level. Due to the lack of publicly available study-level datasets in most included meta-analyses, we relied on moderator results reported by the original authors. However, given the extremely high heterogeneity in our data, it is likely that nudging's effectiveness does vary considerably across different contexts and intervention types, even if our broad subgroup tests lacked power to confirm it. Prior work and patterns in our data hint at meaningful variations: for example, meta-analyses focusing on default-type nudges often found larger effects than those on informational nudges (Jachimowicz et al., 2019; Mertens et al., 2022), but that advantage can disappear in certain domains. Likewise, we observed that nudging interventions in health and finance domains, on average, yielded negligible effects ($d \approx 0$ in each), whereas effects in domains like environment and food were positive. This could imply that nudging is not a universally effective strategy across all areas of behavior – for instance, entrenched health-related behaviors might require stronger or different interventions beyond nudges – although these domain differences were not statistically significant in our meta-regression.

Similarly, we found no significant difference between lab-based and field experiments in our moderator test; nonetheless, the point estimates aligned with expectations from the literature: nudges tested in field settings tended to show smaller effects (and in our aggregate analysis of field studies, the effect was non-significant) compared to those demonstrated in more controlled or academic settings. This pattern suggests that many nudges may not translate as powerfully into real-world contexts, likely due to the influence of external factors and reduced adherence when scaling up. In other words, caution is warranted when generalizing nudge findings from optimal conditions to everyday implementation.

Our second-order meta-analysis did not detect statistically significant moderation by nudge type, behavioural domain, experimental setting or study design, likely because the analyses rely on aggregate data and have limited power. The broadened theoretical landscape indicates that many factors not captured in the available data may shape nudging effects. Baseline motivation and the possibility of substituting other behaviours can diminish observed effects (Saccardo et al., 2025), and misalignment with cultural norms or group values can undermine intervention success (Acierno et al., 2025). Further, evidence from nudgeability research indicates that personal preferences moderate nudge effects: people cannot be nudged into choices they do not want, and effects do not hinge on either nudge transparency or whether decisions are made under System 1 or System 2 (de Ridder et al., 2022). Socioeconomic status may modulate responsiveness, but the evidence is limited and often confined to financial defaults (Allcott, 2011; Beshears et al., 2010; Ghesla et al., 2020; Madrian & Shea, 2001). These insights suggest that the heterogeneity observed in the nudging literature likely reflects a complex interplay of individual, contextual and design factors. Future experimental work should report richer contextual variables and participant characteristics, test for these moderators directly, and explore personalised or context-adaptive nudging strategies.

Recommendations

For future nudging research, we offer the following recommendations. First, develop a unified taxonomy for nudges. The nudging literature currently lacks a consistent classification scheme—some frameworks categorize nudges by the behavioral strategy employed (Dolan et al., 2012; Münscher et al., 2016) and others by underlying psychological mechanism or transparency (Cadario & Chandon, 2017; Y. Li et al., 2023). This inconsistency hinders cross-study comparisons. We recommend that future work converge on a clearer, multi-dimensional taxonomy of nudges (for instance, classifying each nudge by both its functional technique and cognitive mechanism). Establishing such a framework would facilitate more precise analyses of which types of nudges work best and why.

Second, the included meta-analyses of nudging currently suffer from low quality. The most common critical flaws were lack of preregistration and absence of risk-of-bias assessments in the meta-analyses, which greatly lowered their AMSTAR-2 scores. Only three meta-analyses have been preregistered (Mertens et al., 2022; Slapo et al., 2021; Wyse et al., 2021). Similarly, only four meta-analyses have conducted risk assessments for the included meta-analyses (Broers et al., 2017; Ioannou et al., 2021; Slapo et al., 2021; Wyse et al., 2021). These diminish the reliability of meta-analytical results. Future meta-analyses focusing on nudging should follow a more comprehensive workflow and provide open data for evaluation and re-analysis by others.

Third, examine nudging through multiple lenses to fully assess its importance. Our second-order meta-analysis provides a broad overview, but future research should probe deeper into specific aspects of nudging's effectiveness. For instance, studies could investigate the long-term impacts of nudges (do effects persist or wane over time?), their cost-effectiveness relative to other interventions, and potential boundary conditions (e.g., cultural factors or ethical considerations that might moderate effectiveness). Such targeted research efforts, approaching the question from different angles, will complement aggregate evidence and offer a more nuanced understanding of when and how nudging works.

Fourth, improve the reporting and documentation of nudge interventions. We found that many primary studies and meta-analyses did not include detailed information about intervention implementation, context, or sample characteristics. Going forward, providing comprehensive descriptions of interventions (e.g., exact choice architecture changes, population demographics, and setting) and outcomes is essential. This would enable more granular evidence syntheses—such as meta-analyses focusing on specific subtypes of nudges or specific contexts (school settings, health clinics, etc.)—which in turn can identify what works best in which circumstances.

Finally, we note an intriguing pattern in our findings: nudges in certain domains (notably the environmental domain, such as energy-saving nudges) showed among the largest effects, whereas others (like health-related nudges) showed minimal impact. Although our moderator analysis could not confirm domain differences with confidence, this suggests a fruitful direction for future research. We encourage further investigations into why nudging appears more effective in some domains than others. It may be that environmental behaviors are more susceptible to subtle changes in nudging, or simply that interventions in that domain have been better designed and implemented. Unpacking these domain-specific dynamics will help tailor nudging strategies to where they can be most beneficial.

Conclusion

Synthesizing 13 articles (14 meta-analyses) across multiple behavioral domains, our second-order meta-analysis indicates that nudging yields, on average, small and context-dependent effects alongside very high heterogeneity. Signals of small-study and publication bias are nontrivial, and bias-adjusted estimates approach zero; moreover, the moderator patterns are inconsistent and do not reliably account for the dispersion of effects.

While some domains appear more promising in raw estimates, these advantages are not robust once uncertainty and bias are considered. The overarching takeaway is therefore cautious: nudges can work, but their typical impact is modest and uneven.

For policy and practice, this points to a complementary role for nudges rather than a substitute for structural or incentive-based instruments. We recommend routine field testing with transparent preregistration, careful monitoring of implementation fidelity, and reporting on persistence, spillovers, and basic costs to enable meaningful comparisons. For research, clearer and more consistent intervention taxonomies, stronger risk-of-bias safeguards, and living evidence syntheses that incorporate new field studies will help identify when, for whom, and through which pathways nudges are most dependable, thereby improving external validity and decision relevance without overclaiming average effects.

Reference

Acierno, J., Tedaldi, E., Ginn, J., Goldwert, D., Vlasceanu, M., Geiger, S. J., Sparkman, G., & Constantino, S. M. (2025). *Assessing the role of culture in moderating social norm interventions: A global experiment* (No. xzj7r_v1). PsyArXiv. https://doi.org/10.31234/osf.io/xzj7r_v1

Afif, Z., Islan, W. W., Calvo-Gonzalez, O., & Dalton, A. G. (2019). *Behavioral science around the world: Profiles of 10 countries*. World Bank Group (eMBeD, Poverty & Equity GP). https://documents.worldbank.org/en/publication/documents-reports/documentdetail/710771543609067500/behavioral-science-around-the-world-profiles-of-10-countries

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9), 1082–1095. https://doi.org/10.1016/j.jpubeco.2011.03.003

Antinyan, A., & Asatryan, Z. (2020). Nudging for Tax Compliance: A Meta-Analysis. *Marketing Science*, *39*(3), 459–665. https://doi.org/10.2139/ssrn.3680357

Arno, A., & Thomas, S. (2016). The efficacy of nudge theory strategies in influencing adult dietary behaviour: A systematic review and meta-analysis. *BMC Public Health*, *16*, 1. https://doi.org/10.1186/s12889-016-3272-x

Bartoš, F., Maier, M., Quintana, D. S., & Wagenmakers, E.-J. (2022). Adjusting for Publication Bias in JASP and R: Selection Models, PET-PEESE, and Robust Bayesian Meta-Analysis. *Advances in Methods and Practices in Psychological Science*, *5*(3), 25152459221109259. https://doi.org/10.1177/25152459221109259

Berger, S., Kilchenmann, A., Lenz, O., Ockenfels, A., Schlöder, F., & Wyss, A. M. (2022). Large but diminishing effects of climate action nudges under rising costs. *Nature*

Human Behaviour, *6*(10), Article 10. https://doi.org/10.1038/s41562-022-01379-7

Bergquist, M., Thiel, M., Goldberg, M. H., & van der Linden, S. (2023). Field interventions for climate change mitigation behaviors: A second-order meta-analysis. *Proceedings of the National Academy of Sciences*, *120*(13), e2214851120. https://doi.org/10.1073/pnas.2214851120

Berkhout, S. W., Haaf, J. M., Gronau, Q. F., Heck, D. W., & Wagenmakers, E.-J. (2023). A tutorial on Bayesian model-averaged meta-analysis in JASP. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02093-6

Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2010). The limitations of defaults. *Retirement Research Center Working Paper*, *RRC NB10-2*. https://www.nber.org/sites/default/files/2020-08/orrc10-02.pdf

Broers, V. J. V., De Breucker, C., Van den Broucke, S., & Luminet, O. (2017). A systematic review and meta-analysis of the effectiveness of nudging to increase fruit and vegetable choice. *European Journal of Public Health*, *27*(5), 912–920. https://doi.org/10.1093/eurpub/ckx085

Busch, T., & Friede, G. (2018). The Robustness of the Corporate Social and Financial Performance Relation: A Second-Order Meta-Analysis. *Corporate Social Responsibility and Environmental Management*, *25*(4), 583–608. https://doi.org/10.1002/csr.1480

Cadario, R., & Chandon, P. (2017). Which Healthy Eating Nudges Work Best? A Meta-Analysis of Field Experiments. *Marketing Science*, *77*, 1–53. https://doi.org/10.1287/mksc.2018.1128

Chater, N., & Loewenstein, G. (2023a). The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray. *Behavioral and Brain Sciences*, *46*, e147. https://doi.org/10.1017/S0140525X22002023

Chater, N., & Loewenstein, G. (2023b). Where next for behavioral public policy? *Behavioral and Brain Sciences*, *46*, e181. https://doi.org/10.1017/S0140525X23002091

Connolly, D. J., Loewenstein, G., & Chater, N. (2025). An s-frame agenda for behavioral public policy research. *Behavioural Public Policy*, *9*(3), 593–613. https://doi.org/10.1017/bpp.2024.58

de Ridder, D., Kroese, F., & van Gestel, L. (2022). Nudgeability: Mapping Conditions of Susceptibility to Nudge Influence. *Perspectives on Psychological Science*, *17*(2), 346–359. https://doi.org/10.1177/1745691621995183

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

Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, *33*(1), 264–277. https://doi.org/10.1016/j.joep.2011.10.009

Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot–Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629–634. http://www.jstor.org/stable/25175671

Evans, J. St. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social

Cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Ghesla, C., Grieder, M., & Schubert, R. (2020). Nudging the poor and the rich – a field study on the distributional effects of green electricity defaults. *Energy Economics*, *86*, 104616. https://doi.org/10.1016/j.eneco.2019.104616

Gronau, Q. F., Heck, D. W., Berkhout, S. W., Haaf, J. M., & Wagenmakers, E.-J. (2021). A Primer on Bayesian Model-Averaged Meta-Analysis. *Advances in Methods and Practices in Psychological Science*, *4*(3), 25152459211031256. https://doi.org/10.1177/25152459211031256

Halpern, D. (2015). *Inside the nudge unit: How small changes can make a big difference*. Random House.

He, G., Pan, Y., Park, A., Sawada, Y., & Tan, E. S. (2023). Reducing single-use cutlery with green nudges: Evidence from China's food-delivery industry. *Science*, *381*(6662), eadd9884. https://doi.org/10.1126/science.add9884

Hennessy, E. A., & Johnson, B. T. (2020). Examining overlap of included studies in meta-reviews: Guidance for using the corrected covered area index. *Research Synthesis Methods*, *11*(1), 134–145. https://doi.org/10.1002/jrsm.1390

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and Boosting: Steering or Empowering Good Decisions. *Perspectives on Psychological Science*, *12*(6), 973–986. https://doi.org/10.1177/1745691617702496

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics*, *80*, 47–58. https://doi.org/10.1016/j.socec.2019.03.005

Ioannidis, J. P. A. (2009). Integration of evidence from multiple meta-analyses: A primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *CMAJ*, *181*(8), 488–493. https://doi.org/10.1503/cmaj.081086

Ioannou, A., Tussyadiah, I., Miller, G., Li, S., & Weick, M. (2021). Privacy nudges for disclosure of personal information: A systematic literature review and meta-analysis. *Plos One*, *16*(8). https://doi.org/10.1371/journal.pone.0256822

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

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux. https://book.douban.com/subject/6754574/

Li, R., Zhang, Y., Cai, X., Luo, D., Zhou, W., Long, T., Zhang, H., Jiang, H., & Li, M. (2021). The nudge strategies for weight loss in adults with obesity and overweight: A systematic review and meta-analysis. *Health Policy*, *125*(12), 1527–1535. https://doi.org/10.1016/j.healthpol.2021.10.010

Li, Y., Chen, W., & Zhang, S. (2023). The behavioral effects of nudge: A meta-analysis based on a dual perspective of “Cognitive Pathway” and “Transparency.” *Advances* *in Psychological Science*, *31*(12), 2275. https://doi.org/10.3724/SP.J.1042.2023.02275

Lin, Y., Osman, M., & Ashcroft, R. (2017). Nudge: Concept, Effectiveness, and Ethics. *Basic and Applied Social Psychology*, *39*(6), 293–306. https://doi.org/10.1080/01973533.2017.1356304

Lunny, C., Pieper, D., Thabet, P., & Kanji, S. (2021). Managing overlap of primary study results across systematic reviews: Practical considerations for authors of overviews of reviews. *BMC Medical Research Methodology*, *21*(1), 140. https://doi.org/10.1186/s12874-021-01269-y

Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *The Quarterly Journal of Economics*, *116*(4), 1149–1187. https://doi.org/10.1162/003355301753265543

Maier, M., Bartoš, F., Stanley, T. D., Shanks, D. R., Harris, A. J. L., & Wagenmakers, E.-J. (2022). No evidence for nudging after adjusting for publication bias. *Proceedings of the National Academy of Sciences*, *119*(31), e2200300119. https://doi.org/10.1073/pnas.2200300119

Mertens, S., Herberz, M., Hahnel, U. J. J., & Brosch, T. (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(1). https://doi.org/10.1073/pnas.2107346118

Münscher, R., Vetter, M., & Scheuerle, T. (2016). A Review and Taxonomy of Choice Architecture Techniques. *Journal of Behavioral Decision Making*, *29*(5), 511–524. https://doi.org/10.1002/bdm.1897

OECD. (2017). *Behavioural insights and public policy: Lessons from around the world*. OECD Publishing. https://doi.org/10.1787/9789264270480-en

Oxford Centre for Evidence-Based Medicine. (2011). *Levels of evidence*. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 1–11. https://doi.org/10.1186/s13643-021-01626-4

Papatheodorou, S. (2019). Umbrella reviews: What they are and why we need them. *European Journal of Epidemiology*, *34*(6), 543–546. https://doi.org/10.1007/s10654-019-00505-6

Pieper, D., Antoine, S.-L., Mathes, T., Neugebauer, E. A. M., & Eikermann, M. (2014). Systematic review finds overlapping reviews were not mentioned in every other overview. *Journal of Clinical Epidemiology*, *67*(4), 368–375. https://doi.org/10.1016/j.jclinepi.2013.11.007

Ratnayake, G., Peters, R., Lee, J., Chen, Z., & Elmi, N. (2023). Nudging Interventions to Promote Vaccination: A Systematic Review and Meta-Analysis. *Canadian Geriatrics Journal*, *26*(2), 291. https://doi.org/10.5770/cgj.26.686

Saccardo, S., Dai, H., Han, M. A., Raja, N., Vangala, S., & Croymans, D. M. (2025). *Scaling nudges: The roles of baseline motivation and substitution effects*. https://bfi.uchicago.edu/wp-content/uploads/2025/02/Scaling-Nudges-The-Roles-of-Baseline-Motivation-and-Substitution-Effects.pdf

Sala, G., Aksayli, N. D., Tatlidil, K. S., Tatsumi, T., Gondo, Y., & Gobet, F. (2019). Near and Far Transfer in Cognitive Training: A Second-Order Meta-Analysis. *Collabra: Psychology*, *5*(1), 18. https://doi.org/10.1525/collabra.203

Schubert, C. (2017). Green nudges: Do they work? Are they ethical? *Ecological Economics*, *132*, 329–342. https://doi.org/10.1016/j.ecolecon.2016.11.009

Schweizer, M. L., Braun, B. I., & Milstone, A. M. (2016). Research methods in healthcare epidemiology and antimicrobial stewardship – quasi-experimental designs. *Infection Control and Hospital Epidemiology*, *37*(10), 1135–1140. https://doi.org/10.1017/ice.2016.117

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, *358*, j4008. https://doi.org/10.1136/bmj.j4008

Slapo, H., Schjoll, A., Stromgren, B., Sandaker, I., & Lekhal, S. (2021). Efficiency of In-Store Interventions to Impact Customers to Purchase Healthier Food and Beverage Products in Real-Life Grocery Stores: A Systematic Review and Meta-Analysis. *Foods*, *10*(5). https://doi.org/10.3390/foods10050922

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

Sunstein, C. R. (2015). Nudges, Agency, and Abstraction: A Reply to Critics. *Review of Philosophy and Psychology*, *6*(3), 511–529. https://doi.org/10.1007/s13164-015-

0266-z

Szaszi, B., Higney, A., Charlton, A., Gelman, A., Ziano, I., Aczel, B., Goldstein, D. G., Yeager, D. S., & Tipton, E. (2022). No reason to expect large and consistent effects of nudge interventions. *Proceedings of the National Academy of Sciences*, *119*(31), e2200732119. https://doi.org/10.1073/pnas.2200732119

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

Vlaev, I., King, D., Dolan, P., & Darzi, A. (2016). The theory and practice of “nudging”: Changing health behaviors. *Public Administration Review*, *76*(4), 550–561. https://doi.org/10.1111/puar.12564

von Klinggraeff, L., Ramey, K., Pfledderer, C. D., Burkart, S., Armstrong, B., Weaver, R. G., & Beets, M. W. (2023). The mysterious case of the disappearing pilot study: A review of publication bias in preliminary behavioral interventions presented at health behavior conferences. *Pilot and Feasibility Studies*, *9*(1), Article 1. https://doi.org/10.1186/s40814-023-01345-8

Waheed, H. (2023). Nudging smokers away from lighting up: A meta-analysis of framing effect in current smokers. *Journal of Behavioral and Experimental Economics*, *104*. https://doi.org/10.1016/j.socec.2023.101998

WHO & World Bank. (2024). *Behavioral science around the world: Volume III: public health*. World Health Organization; International Bank for Reconstruction and Development / The World Bank. https://www.who.int/publications/i/item/B09046

World Bank. (2015). *World development report 2015: Mind, society, and behavior*. World Bank. https://doi.org/10.1596/978-1-4648-0342-0

Wyse, R., Jackson, J. K., Delaney, T., Grady, A., Stacey, F., Wolfenden, L., Barnes, C., McLaughlin, M., & Yoong, S. L. (2021). The Effectiveness of Interventions Delivered Using Digital Food Environments to Encourage Healthy Food Choices: A Systematic Review and Meta-Analysis. *Nutrients*, *13*(7). https://doi.org/10.3390/nu13072255

Zhang, J., Huang, Y., Zhu, J., & Zhao, L. (2023). A meta-analysis on the effectiveness of food-waste reducing nudges. *Food Policy*, *120*. https://doi.org/10.1016/j.foodpol.2023.102480

Zhao, N., Liu, X., Li, S., & Zheng, R. (2022). Nudging effect of default options: A meta-analysis. *Advances in Psychological Science*, *30*(6), 1230–1241. https://doi.org/10.3724/SP.J.1042.2022.01230