



MASTERARBEIT | MASTER'S THESIS

Titel | Title

Behavioral Interventions to Foster Conservation Behavior: An AI-Assisted Meta-Analysis

verfasst von | submitted by

Tim Lukas Bergmann B.Sc.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 913

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Applied Economics

Betreut von | Supervisor:

Assoz. Prof. Dr. sc. ETH Alexandra Brausmann

Abstract

This thesis examines the extent to which behavioral interventions can foster resource conservation behavior. To this end, a meta-analysis is conducted on the basis of 99 randomized controlled field trials that capture objectively observed behavior through measured resource consumption. Study screening and data extraction were carried out using AI-generated suggestions, with the final decisions made manually. In total, 169 effect sizes from 77 publications are synthesized using a three-level meta-regression. The analysis yields a small effect (Hedges' $g = -0.144$), which should be regarded as an upper bound given a potential publication bias. In addition, a moderator analysis is conducted to examine which factors influence the effectiveness of behavioral interventions. Its central hypothesis is that the setting serves as a proxy for the strength of the underlying habit: the more habitual and thus largely automatic a behavior is, the less receptive individuals are to the intervention. The analysis shows that interventions are less effective in habit-dominated household settings than in less routinized (tourism) contexts. In this way, the thesis advances the understanding of how behavioral interventions can contribute to resource conservation.

Zusammenfassung

Diese Masterarbeit untersucht, inwieweit Verhaltensinterventionen ressourcenschonendes Verhalten fördern können. Hierzu wird eine Meta-Analyse auf Basis von 99 randomisierten kontrollierten Feldstudien durchgeführt, die objektiv beobachtetes Verhalten anhand des gemessenen Ressourcenverbrauchs erfassen. Die Studiensichtung und Datenextraktion wurden mithilfe von KI-Vorschlägen und anschließender manueller Entscheidung durchgeführt. Insgesamt fließen 169 gemessene Effektgrößen aus 77 Publikationen ein, die mittels einer Drei-Level-Meta-Regression synthetisiert werden. Die Analyse ergibt einen kleinen Effekt (Hedges' $g = -0,144$), der aufgrund eines potenziellen Publikationsbiases als Obergrenze zu verstehen ist. Darüber hinaus wird eine Moderatorenanalyse durchgeführt, um zu untersuchen, welche Faktoren die Wirkung von Verhaltensinterventionen beeinflussen. Im Zentrum steht die Hypothese, dass der Anwendungskontext als Näherung für die Stärke der zugrunde liegenden Gewohnheit dient: Je stärker ein Verhalten gewohnheitsmäßig und damit weitgehend automatisiert ausgeführt wird, desto geringer ist die Empfänglichkeit der Individuen für die Intervention. Die Analyse zeigt, dass Interventionen in gewohnheitsgeprägten Haushaltssettings schwächer wirken als in weniger routinisierten (Tourismus-)Kontexten. Auf diese Weise erweitert die vorliegende Masterarbeit das Verständnis dafür, wie Verhaltensinterventionen zur Ressourcenschonung beitragen können.

Contents

1	Introduction	4
2	Theoretical Considerations	6
2.1	Theoretical Literature	6
2.2	Experimental Literature	8
3	Data and Methodology	11
3.1	Literature Search and Screening	11
3.2	Inclusion Criteria	12
3.3	Coding of Moderators	13
3.4	Harmonization of Effect Sizes	15
3.5	Meta-Analysis using a Random-Effects Model	16
3.6	Moderator Analytical Methodology	17
3.7	Publication Bias	18
4	Results	19
4.1	Descriptive Results	19
4.2	Meta Analytical Results	24
5	Limitations	29
6	Conclusion	31
	References	33
	Appendix	i

List of Figures

1	Number of effects per intervention type, by year	20
2	Number of effects per intervention type, broken down by conservation behavior, using the full sample of 169 effect	21
3	Share of conservation behavior by setting	21
4	Funnel plot presenting the relationship between precision, measured as standard error, and effect size. Calculated as Hedges' g	22
5	Share of conservation behavior by intervention type	23
6	Intervention-type mix within each conservation behavior	23
A1	PRISMA flow diagram illustrating the study selection process	i
A2	Screenshot showing the built application used for the full-text screening, screening mode	iii
A3	Screenshot showing the built application used for the data extraction, data extraction mode	iv

List of Tables

1	Meta-analytic estimates by moderator subgroup.	25
2	Three-level meta-analytic estimates by moderator subgroup	26
3	Marginal omnibus Wald tests by moderator	28
4	Marginal omnibus Wald tests by moderator (social comparison excluded).	29
A1	Search strategy used for the literature search in Scopus	ii
A2	Coded studies included in the meta-analysis	xiv

List of Acronyms

API	Application Programming Interface
CR2	Cluster-Robust Variance Estimation (Type 2)
HTZ	Hotelling–Zhang Approximation
IPCC	Intergovernmental Panel on Climate Change
LLM	Large Language Model
OLS	Ordinary Least Squares
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT	Randomized Controlled Trial
REML	Restricted Maximum Likelihood
SMD	Standardized Mean Difference
SD	Standard Deviation
SE	Standard Error
TPB	Theory of Planned Behavior

1 Introduction

Human activities are a major driver of environmental degradation and climate change, contributing substantially to the depletion of natural resources and increased greenhouse gas emissions (IPCC, 2022). Furthermore, the United Nations Environment Programme reports that plastic waste equivalent to 2,000 garbage trucks is discharged into global waterways daily (United Nations Environment Programme, 2022). While both of these issues require systemic transformations on a global scale, they also entail that everyday decisions about our available resources need to adapt. These include, but are not limited to, decisions about heating or cooling, watering techniques, and waste management. At the same time, the Intergovernmental Panel on Climate Change (IPCC, 2022) highlights that demand-side solutions, including changes in individual behavior and consumption patterns, have considerable potential to contribute to climate change mitigation. In this context, interventions aimed at promoting pro-environmental behavior have received growing scholarly and political attention. Among these, non-price interventions stand out for their relatively low implementation costs and potential for large-scale deployment. Nonetheless, evidence regarding the effectiveness of such interventions is scattered across behavioral domains, intervention strategies, and contextual settings.

Therefore, this thesis reviews the theoretical foundations of pro-environmental behavior and the behavioral barriers that may hinder sustainable action. It then examines the main categories of non-price interventions used to address these barriers and synthesizes evidence from field experiments to assess their effectiveness in promoting pro-environmental behavior. To ensure a reliable synthesis of the evidence, the meta-analysis is restricted to randomized controlled field trials investigating conservation behavior. Following Nisa et al. (2019), this restriction aims to reduce potential biases arising from less rigorous study designs, such as laboratory experiments or studies relying on self-reported behavioral outcomes, thereby providing more credible estimates of intervention effectiveness. Accordingly, this thesis addresses the following research question: What is the overall effectiveness of behavioral interventions in promoting resource conservation behavior (electricity, water, and waste reduction)?

Beyond estimating the overall effect of non-price interventions, a moderator analysis is conducted to examine the influence of intervention strategy, behavioral domain, setting, and sample size on the efficacy of behavioral interventions. The investigation of sources of heterogeneity is important because, as the literature readily argues, there is no silver bullet to foster pro-environmental behavior; expanding the knowledge on why interventions yield heterogeneous effects can help align intervention design more closely with target populations. Moreover, most interventions aim at deliberative decision-making and neglect

the effects of habits on behavior. Examining the setting in which studies take place may help draw conclusions on the relevance of habits. Which is why, unlike previous meta-analyses (e.g., Nisa et al., 2019), this thesis focuses exclusively on conservation behaviors involving energy, water, and waste usage. These behaviors share the characteristics of being embedded in everyday routines and generally do not require capital investments, making them particularly suitable for examining habitual decision-making. This thesis tests the hypothesis that, in a household setting, habit strength negatively affects behavioral response. Together, the question: Does the effectiveness of behavioral interventions depend on how habitual the targeted behavior is, such that interventions are less effective in stable, routine-dominated settings (e.g., households) than in disrupted ones (e.g., tourism), is answered by this thesis.

Since the publication of previous meta-analyses, the evidence base on non-price interventions has grown considerably, warranting an updated synthesis (see Figure 1). Furthermore, existing meta-analyses have primarily differentiated interventions by behavioral domain or intervention type, without distinguishing the setting in which the intervention is employed, a distinction that may act as a proxy for habit strength. By coding the settings in which conservation behaviors take place, this thesis investigates whether intervention effectiveness varies systematically with the habitual strength of the targeted behavior. In addition, recent advances in meta-analytic methodology enable a more rigorous treatment of statistically dependent effect sizes. Because many primary studies report multiple treatment arms, conventional meta-analytic approaches either discard information or assume independence between effect sizes. This thesis instead employs a three-level meta-regression to account for the nested structure of effect sizes within studies (Van den Noortgate et al., 2013; Cheung, 2014), and applies cluster-robust variance estimation (Tipton, 2015).

The results of the conducted meta-analysis indicate that non-price interventions produce small but statistically significant effects on conservation behavior. The moderator analyses further suggests that intervention effectiveness varies systematically across behavioral settings, with household contexts exhibiting weaker effects than less routine-dominated environments.

The remainder of this thesis is structured as follows. Section 2 reviews the relevant theoretical and empirical literature on non-price interventions. Section 3 presents the meta-analytical methodology used to estimate their effectiveness and identify moderating factors. Section 4 reports the results of the meta-analysis, including the moderator analyses, and situates these findings in relation to the estimates obtained by previous meta-analytical work. Section 5 discusses the limitations of these results. Section 6 concludes, summarizing the key findings and outlining directions for future research.

2 Theoretical Considerations

The following chapter begins by outlining the theoretical foundations underlying the development and maintenance of behavior, which are essential to understanding the mechanisms by which behavior can be changed. Subsequently, it provides an overview of behavioral interventions employed in the field and presents the first evidence on their effectiveness in promoting behavior change in resource conservation.

2.1 Theoretical Literature

Standard economic theory predicts that individuals engage in environmentally friendly behavior when the expected private benefit exceeds the associated private costs. Environmental problems, however, are often characterized as a common goods dilemma (Ostrom, 1990) in which individual decision-making may generate negative externalities. Consequently, individuals may have insufficient incentives to adopt pro-environmental behaviors, even when they are socially desirable. An early discussion on how a society may handle negative externalities is provided by Pigou (1932). From a Pigouvian standpoint, corrective taxes on socially (environmentally) harmful activities help internalize external costs by aligning private incentives with societal welfare (Pigou, 1932). One prominent example of an environmentally grounded Pigouvian tax is Sweden’s national carbon tax, implemented in 1991, which increases the price of carbon-intensive fuels to better reflect their social costs (Hammar and Åkerfeldt, 2015).

Although the classical economic approach offers mechanisms to address externalities, its practical application is often constrained by political resistance, administrative costs, and challenges in rigorously monitoring and enforcing environmentally relevant behavior. Moreover, evidence from behavioral economics suggests that individuals do not always respond according to economic incentives (Tversky and Kahneman, 1981; Tversky and Kahneman, 1986). Furthermore, pro-environmental behavior is influenced by a range of psychological and contextual factors, including social norms, habits, attitudes, and situational constraints (Ajzen, 1991; Verplanken and Aarts, 1999). These considerations have motivated growing interest in non-price interventions to foster environmentally friendly behavior.

Several theoretical frameworks have been developed to explain how behavior cultivates, and understanding the factors that drive behavior is essential for designing effective behavioral interventions. These frameworks can be broadly categorized as either considering behavior as the outcome of a deliberate decision-making process or the product of more automatic mechanisms. Deliberative theories, such as the Theory of Planned Behavior (TPB; Ajzen, 1991), propose that behavior is guided by conscious evaluations, social

norms, and perceived behavioral control. In contrast, other scholars argue that behavior, especially frequent behavior, is often shaped by heuristics, cognitive biases, and habitual responses that operate with limited deliberation (Kahneman, 2003; Thaler and Sunstein, 2008; Verplanken et al., 1998; Verplanken and Aarts, 1999).

The TPB (Ajzen, 1991) posits that behavior results from a conscious decision-making process in which attitudes toward the behavior, subjective norms, and perceived behavioral control collectively influence the intention to act. Attitudes toward the behavior refer to an individual’s overall evaluation of performing it, whereas subjective norms capture the perceived social pressure to engage in or refrain from it. Perceived behavioral control reflects an individual’s perception of the ease of performing the behavior and, according to Ajzen (2002), comprises both perceived self-efficacy (i.e., confidence in one’s ability to perform the behavior) and perceived controllability (i.e., the extent to which performing the behavior is perceived to be under one’s control).

Whether the formed intention leads to actual behavior further depends on actual behavioral control and external obstacles. For instance, according to the TPB, the intention to conserve water forms when an individual views water conservation as a noble goal, believes others share this view, and feels capable of conserving water in their daily life. However, situational factors may hinder the execution of intentions. For example, when using an automatic faucet, the running duration may be longer than needed, but the faucet’s default setting cannot be altered. From an intervention perspective, the TPB implies that pro-environmental behavior can be encouraged by targeting attitudes, subjective normative beliefs, or by enhancing perceived behavioral control. Notably, TPB argues that the relative importance of these determinants varies across behaviors and contexts (Ajzen, 1991). Consequently, an intervention targeting the most influential determinant in a given setting may be sufficient to strengthen pro-environmental intentions without addressing all three simultaneously.

Nonetheless, most behaviors targeted by behavioral interventions are daily actions that may result from habitual performance rather than deliberate assessment. Hence, the relevance of habitual actions warrants discussion. In contrast to the deliberate action theories, scholars argue that in many situations, behavior is automatic (Verplanken and Aarts, 1999).

A theoretical approach that posits that an individual first considers different aspects of norms and attitudes toward that behavior may underestimates the automaticity with which a behavior is performed. Verplanken and Aarts (1999) connect the Theory of Planned Behavior to habits and argue that both intention and behavior are influenced by the strength of prior habit. An established habit reduces the likelihood of deviant behavior. Nonetheless, an intervention may prompt the individual to reconsider a habit,

thereby creating conditions for a new habit to form. Verplanken and Aarts (1999, p. 105) argue that habits are goal-directed in that their initiation is often intentional, whereas their subsequent execution occurs without conscious intention. Interventions that rely on changing attitudes or norms may leave the behavior untouched if the behavior happens fully automatically. While interventions could initiate a deliberate change of behavior. This, however, poses the question of whether interventions built on these assumptions alter behaviors in settings where they are most habitual.

Taken together, the discussed theories suggest that intervention strategies aimed at promoting pro-environmental behavior need to understand the sort of behavior they address and take into account the respective challenges to change it.

2.2 Experimental Literature

The use of behavioral interventions to improve household resource conservation dates back to research on the late 1970s energy crisis (Delmas et al., 2013). Since then, these strategies have become widespread, especially with the rise of behavioral economics and technological advancements that facilitate scalability. Having discussed the theoretical foundation used to motivate behavioral interventions, the experimental work conducted in the field is now discussed.

Appeals, Prompts, and Reminders:

Appeals, prompts, and nudges, as defined herein, are intended to motivate individuals to engage in more environmentally sustainable behaviors. These are generally communicated through messages, reminders, or cues positioned in proximity to the targeted action (for example, appeals to conserve water are situated in the bathroom). Appeals are commonly used in the tourism sector as low-barrier interventions that address the specific needs of a given area or hotel (Greene et al., 2025), and have also been applied in drought scenarios to inform water conservation efforts (Tonke, 2024). A meta-analysis conducted by Nisa et al. (2019) revealed that appeals exerted a small effect ($d = -0.27$) on environmentally relevant behaviors. Nonetheless, the authors also acknowledge that, due to significant heterogeneity across studies, the interpretation of these findings should be approached with caution. Bergquist et al. (2023) concluded with an effect size of $d = 0.28$.¹

Information & Education: Information and education have been deployed as behavioral interventions for resource conservation since the earliest field experiments of the 1970s and 1980s, as documented in meta-analytic reviews covering this period (Delmas

¹Despite the signs hinting in opposite directions, the effects in both studies represent an increase in pro-environmental behavior. Throughout this section, negative and positive effect sizes may both indicate increases in pro-environmental behavior, as sign conventions differ across the cited meta-analyses. In this thesis, negative values denote increased conservation.

et al., 2013), and remain the most common single ingredient in contemporary intervention strategies (Nisa et al., 2019). Informational and educational interventions can be grouped into those that provide conservation-relevant knowledge in the form of appliance-use tips, guidance on efficient usage behaviors, thereby lowering the search costs of acting on conservation preferences, and awareness interventions that aim to make the negative consequences of performed behavior salient. Typical delivery is through low-cost channels, such as leaflets, door hangers, emails, SMS, or in-home displays. Meta-analyses find small average effects with substantial heterogeneity across specific implementation strategies (Delmas et al., 2013; Nisa et al., 2019). Recent experimental efforts leverage the growing availability of data and digital connectivity to include personalized information and increase communication. These approaches are grounded in research showing that generic content does not effectively engage households with diverse consumption habits and that interventions that use more sustained engagement are more effective (Nisa et al., 2019). Additionally, in the post-2015 literature, strategies often combine informational approaches with other methods, such as social norm messages or social comparison, rather than relying on any strategy alone.

Feedback: The use of feedback interventions to promote resource conservation has been increasingly adopted, enabled through the rollout of smart metering and utility billing. They are characterized by providing individuals or households with feedback on their own consumption patterns over time or in real time. Comparison is only made to one’s own consumption. This bridges the gap between perceived and actual consumption to end-users by making consumption salient. Today, in-home displays or shower equipment can provide real-time consumption feedback (Tiefenbeck et al., 2018). Meta-analyses examining the effectiveness of feedback interventions using randomized controlled trial (RCT) samples provide small average effects of -7.6% (Nemati and Penn, 2020), or a Cohen’s d from -0.048 (Nisa et al., 2019) to 0.16 in Bergquist et al. (2023).

Social norm messages: Normative messaging and social comparison are closely related behavioral approaches grounded in theories of social influence and normative conduct. Social norm messages are based on the premise that individuals’ behavior is shaped by their perceptions of what others do (descriptive norms) or what they perceive others would approve of (injunctive norms). Consequently, informing individuals about prevailing social norms can influence their behavior when choosing beverage containers (Loschelder et al., 2019) and towel re-use in a hotel setting in Thailand (Morgan and Chompreeda, 2015).

Social comparison: Social comparison interventions provide individuals with feedback on their behavior relative to a reference group. These interventions are mainly used as household consumption reports for energy (Allcott, 2011), water (Bhanot, 2017), or

waste (Bonan et al., 2025). Comparison groups may be defined at scales ranging from immediate neighbors to the state or national level. Recent studies are exploring ways to enhance the effects of these reports by examining the physical and psychological distance between the user and different reference groups (Brülisauer et al., 2020; Wang et al., 2025). Furthermore, most often, the comparison information is paired with conservation knowledge (Bhanot, 2017), information about environmental consequences (Asensio and Delmas, 2015), or norm messages (Wang et al., 2025).

Meta-analytical evidence suggests that the effectiveness of normative intervention strategies is modest in absolute terms ($d = -0.077$, Nisa et al., 2019; $d = 0.37$, Bergquist et al., 2023, or in the longer run, $d = -0.07$, Lemmel et al., 2025). Supporting its relative effectiveness, Pellerano et al. (2017) provide evidence that social comparison interventions outperform monetary interventions. Bergquist et al. (2023) further argue that the effect size may be more reliable in social comparison interventions as they are usually applied on a greater scale and a more heterogeneous sample.

Goal setting: Goal-setting interventions typically require participants to formulate specific goals and involve monitoring progress over time, making them more resource-intensive to implement than the intervention strategies mentioned above. Consequently, they appear to be relatively uncommon in the pro-environmental behavior literature. Despite these challenges, goal-setting interventions have shown moderate to large effects on self-reported behavior ($d = 0.524$, Carrero et al., 2025) and observed behavior (e.g., $d = 0.65$, Osbaldiston and Schott, 2012; $d = -0.48$, Nisa et al., 2019). Though observed-behavior effects are notably smaller ($d = 0.334$ observed vs. $d = 0.524$ self-reported; Carrero et al., 2025). Evidence using objective behavioral outcomes remains limited. An example of a goal-setting intervention is provided by Legault et al. (2020), who combine a personalized motivational reflection exercise with goal setting, supplemented with continuous usage feedback in a field experiment with meter-recorded electricity and water use across 77 university apartments. The goal-setting component reduced hot water use by 21%, though the electricity effect (20%) only emerged when goal setting was paired with real-time feedback, illustrating that their effectiveness may depend on complementary supports like feedback.

Decision structural nudges: Decision structural nudges, also called default nudges, are a subgroup of nudging strategies that target intuitive decision making by modifying the choice environment. One example of a decision-structural nudge, in a non-environmental setting, is provided by Wansink (2004), in which reducing plate size in a cafeteria setting led to a decrease in caloric intake. Similar interventions were identified among the studies included in the meta-analysis. Kallbekken and Sælen (2013) implemented plate-size reduction in a hotel buffet and found that this intervention reduced food waste by

approximately 20%. Likewise, Giaccherini et al. (2021) examined the effect of making doggy bags the default option for customers with leftovers in restaurants in Turin. This made it an opt-out decision not to take leftovers home, thereby encouraging food waste reduction. They found a positive yet non-significant effect.

Overall, individual studies often report encouraging effects of behavioral interventions on resource conservation. However, meta-analytical evidence paints a more cautious picture, revealing relatively modest average effects and substantial heterogeneity, particularly concerning the RCT studies, in which self-selection is less common. While previous research has primarily focused on differences between intervention strategies, the behavioral setting in which interventions are implemented has received little attention as a potential moderator. Given theoretical arguments that habit strength varies across settings, examining whether intervention effectiveness differs between more and less habitual environments may yield important insights for designing more effective interventions.

3 Data and Methodology

The following section describes the methodology employed in this meta-analysis. It first outlines the study identification and selection process, including the literature search and the AI-assisted screening procedure, followed by the extraction and harmonization of effect sizes. Subsequently, it outlines the statistical methods used for the meta-analysis, including the random-effects model, the three-level meta-regression model, and the omnibus Wald tests used to assess moderator effects.

3.1 Literature Search and Screening

The Scopus database was used as the primary source for identifying relevant papers. A Boolean search string was constructed to capture a comprehensive range of literature across multiple disciplines. The used search string is presented in Table A1 and yielded 13,481 results. The search was conducted on 28 October 2025. After applying filters to include only English-language publications from 2000 onward and removing duplicate records, a total of 11,720 records were exported for further screening. The screening process begins with reviewing titles and abstracts. For this, the ASReview LAB v2, an open-source software application developed by the DISC-AI lab at Utrecht University, was used. ASReview employs active learning algorithms that iteratively refine the screening process based on prior classification decisions, thereby enhancing efficiency while maintaining a semi-automated, human-in-the-loop approach. The platform has been shown to perform effectively in simulation studies by Van de Schoot et al. (2021). The ELASu4

model configuration was applied, with a stopping criterion of 2% of all identified records; that is, screening continued until 234 consecutive title-abstract pairs were classified as irrelevant. No prior knowledge was introduced at the onset of the screening process. In cases of uncertainty regarding a paper’s relevance, it was conservatively marked as relevant for subsequent full-text review. This precautionary approach was adopted due to substantial heterogeneity in terminology and variation in abstract-writing conventions across disciplines. Following the title and abstract screening, 233 studies were identified as eligible and considered for the full-text screening.

The full-text screening and data extraction were assisted by a custom R/Shiny application developed for this thesis. The application uses a human-in-the-loop interface that employs a publicly available large language model (LLM) to extract structured information from PDF documents. The application’s design was inspired by the approach of the open-source AI-Data-Extraction tool by Schroeder (2024), though the implementation is independent. A screenshot of the application is provided in the appendix. As public LLMs, either the model Claude Sonnet 4 model or the latest Mistral large model was used. The use was based on the availability of the model², and no significant performance difference was noticed.

The application connects to a public LLM via an API. Before screening, the reviewer develops prompts resembling the inclusion criteria, which were subsequently applied to each full-text article. The LLM-generated responses and their source page are displayed alongside the corresponding article. The reviewer holds full responsibility for inclusion, and the LLM outputs are used as decision support only. A screenshot of the application interface is provided in the appendix. Of the 233 papers assessed in the full-text screening, 77 papers met the eligibility criteria and were included in the meta-analysis.

3.2 Inclusion Criteria

Studies were included if they met all of the following criteria. First, the study employed a randomized controlled trial in a natural (field) setting, with individual-level interventions. This aims to permit clean control-treatment comparison while ensuring that the measured effect is attributable to a manipulation of an individual decision-making process rather than to group-level dynamics, such as peer pressure. Second, the outcome was an objectively measured behavior rather than self-reports or behavioral intentions, reducing the risk of social desirability and outcome-reporting bias. Third, the study reported sufficient level-scale information to compute a standardized mean difference. Effects reported only on a proportional scale (e.g., log-transformed outcomes) were excluded. Combining ratio- and difference-based measures would introduce heterogeneity arising from differences in

²As of the time of analysis: mistral-large-2512

scale rather than in intervention effectiveness (Higgins et al., 2024).

The paper selection process has been documented, using the PRISMA flow diagram (Page et al., 2021), Figure A1

3.3 Coding of Moderators

This thesis uses seven categories to code behavioral interventions: information and education; feedback; appeals, nudges, and reminders; social comparison; decision structural nudges; goal setting; and social norm messages³. This taxonomy has been inspired by previous meta-analyses (Nisa et al., 2019; Mi et al., 2021) but departs from earlier work in theoretically motivated ways, reflecting the scope and specific data of the present thesis.

Information and Education interventions provide factual information about the environmental or health effects of certain behaviors, tips on how to preserve energy, or how to reduce consumption. The aim of these interventions is primarily to update knowledge and support deliberate decision-making. Examples include fliers, educational seminars, and door hangers with tips on energy conservation.

Feedback interventions provide individuals and households with information about their past consumption of electricity, gas, water, or waste. The feedback is commonly provided through smart metering technology or consumption reports. The feedback category is distinguished by its self-referential nature and the absence of a deeper intention to influence individuals' behavior beyond enabling self-monitoring and awareness.

Goal-setting interventions further encourage individuals to not only track their behavior but also to commit to specific objectives through behavioral changes. For instance, households might aim to cut their water use by 10% compared to the previous month. These interventions consistently include feedback to help participants see their progress toward the goal. This approach extends beyond basic feedback by actively prompting individuals to improve.

The **social comparison** category includes interventions that provide individuals with a specific quantitative reference point based on others' behavior. Therefore, it works by updating beliefs about one's identity and consumption patterns compared to a normative benchmark. Examples include reporting that a household's energy consumption is above the average for similar households in the neighborhood. Similar to goal-setting interventions, social comparison requires feedback to function properly. To be considered as a social comparison intervention, the study does not require any direct request or motivational framing beyond the benchmark itself.

Appeals, prompts, and reminders are behavioral interventions that increase salience

³Although gamification was initially considered as an intervention strategy, no eligible studies were identified. Therefore, this strategy was not included in the review.

and/or the motivation to perform a behavior, commonly through prompts or requests. Appeals prompt engagement with the individual by addressing values and social responsibility, but they do not convey information about others' behavior. Examples include "please help us keep the beach clean," and messages to reduce food waste to the benefit of an animal's health or to help the hotel achieve its goal of reducing water use. A direct message to the recipient qualifies as an appeal if it does not convey behavioral information about others.

Social norm messages leverage social norms to instill a sense of moral duty in the recipient. They explicitly encourage a specific behavior by emphasizing that others are already engaging in it or are supportive of it. This approach heightens the psychological pressure to conform to the norm. Examples include messages indicating that most neighbors have cut down their energy use (descriptive norm), or that guests at the hotel should help reduce towel usage (injunctive norm). Unlike appeals, they derive their motivational force from the behavior or mentioned expectations of others.

Decision structural nudges aim to subtly guide behavior without limiting an individual's choices. They are considered in Thaler and Sunstein's (2008) concept of nudging, which seeks to influence behavior rather than impose paternalistic control. While "default nudges" and "nudging" are often used interchangeably, this thesis distinguishes interventions that modify default options without direct communication with individuals from those that communicate and target underlying thought processes. Choice architectural interventions are defined here as interventions that influence the participant's available options without informing the individual about the process. They can thus be seen as a more technical approach, as they change the design of the underlying behavior. Examples include lowering thermostat settings in winter, capping air conditioning levels, and using smaller plates in restaurants.

A key methodological challenge in meta-analysis concerns the coding of intervention studies combining multiple behavioral stimuli. The approach taken in the present thesis is to code studies according to their primary theorized mechanism, that is, the intervention that a paper argues is the one being examined. For most categories, this isolates one active ingredient. For goal-setting and social comparison interventions, however, usually multiple treatments are deployed simultaneously. Coding such a study to its single primary category is therefore not a reduction to a single component; rather, the category label itself denotes the bundle. Therefore, calculated effects should not be understood as a single effect of a single intervention but rather as the bundled effect.

For the moderator analysis, the settings in which the interventions are employed are divided into three categories: tourism, households (including dormitories), and other settings. This thesis hypothesizes that the setting may act as a proxy for the degree of

contextual stability associated with the targeted behavior. This rationale is grounded in habit theory, which proposes that behaviors repeated in stable contexts become increasingly automatic and guided by contextual cues (Wood and R  nger, 2016). Consequently, household settings are expected to be associated with well-established routines, whereas tourism settings are hypothesized to provide fewer consistent contextual cues and greater disruption to established routines.

Additionally, study characteristics, such as conservation behavior and sample size, are coded. Conservation behavior was coded *ex post* on the basis of the behaviors observed in each study, yielding five categories: water, energy (including electricity and other energy uses, such as gas used for heating), waste, food waste, and towel reuse. Sample size was grouped into three buckets: fewer than 100, 100 to 500, and more than 500 participants.

Notably, most studies conduct their analyses after a short period, typically less than 12 months; therefore, the effect sizes should be interpreted as short-term and do not provide substantial information about the persistence of the effect.

3.4 Harmonization of Effect Sizes

Meta-analyses aim to combine results from various studies that report their results across different units (such as kg/day, l/day, percentages) and reporting methods (such as mean changes and percentage of uptake). To harmonize, the reported metrics are converted into Cohen’s *d* values. Cohen’s *d* yields a dimensionless quantity that represents the magnitude of an effect independent of the original measurement scale, thereby enabling direct comparison of intervention effects (Borenstein et al., 2021). The resulting Cohen’s *d* values are then corrected for small-sample bias, applying the Hedges’ *g* small-sample correction (Hedges, 1981).

One challenge that arises when combining studies across fields is the diversity in experimental design and reporting standards. Because the inclusion criteria require an RCT design, randomization should, in theory, equate the treatment and control groups at baseline, the assumption underlying the common practice of computing between-subject standardized mean differences (SMDs) from post-intervention data alone (Nisa et al., 2019; Mi et al., 2021). In practice, however, baseline imbalance can persist despite randomization, so when studies report pre- and post-intervention statistics, these are used to capture the intervention effect net of residual baseline differences (Skvarc and Fuller-Tyszkiewicz, 2024).

Based on this rationale, the most appropriate formula for each study was selected from the available information. To reduce the analyst’s discretion, formulas were evaluated in a fixed order of preference. For continuous outcomes: pre- and post-means with pre-SD⁴ or

⁴Following the recommendation in Morris (2008), pre-test standard deviations are used to minimize

regression coefficients with baseline adjustments (difference-in-differences estimates, ordinary least squares (OLS) with baseline adjustment). If no change statistics are available, post-means with post-SD or t-statistics are used. For binary outcomes, the logit method was applied. The formulas used are from the Campbell Collaboration (see Wilson, 2023). The variance was calculated for each SMD using the following formula:

$$v_d = \frac{n_c + n_t}{n_c n_t} + \frac{d^2}{2(n_c + n_t)} \quad (1)$$

$$J = 1 - \frac{3}{4(n_c + n_t - 2) - 1} \quad (2)$$

$$v_g = J^2 v_d \quad (3)$$

Where v_d is the variance of the Cohen's d estimate calculated using the control group size n_c and treatment group size n_t , d^2 being the squared Cohen's d. Moreover, J is the small-sample bias correction factor, which is also used for converting Cohen's d into Hedges' g, v_g is the variance of the Hedges' g estimate.

3.5 Meta-Analysis using a Random-Effects Model

Expanding on the concept of a literature review, meta-analyses combine reported effect sizes to compile the underlying effect size. The random-effects model assumes that effect sizes differ in their underlying effect and respective distribution (Borenstein et al., 2010). The included studies in this thesis differ not only in intervention type, stringency, and aim but also in the underlying population. Thus, the random-effects model reliably combines effect sizes into a weighted mean.

The formula from Borenstein et al. (2021, p. 72) gives us the mean effect-size calculation for a random-effects model as:

$$M = \frac{\sum_{i=1}^k W_i Y_i}{\sum_{i=1}^k W_i} \quad (4)$$

Which is the sum of the weighted effect sizes divided by the sum of weights. The weights are the inverse of the study-specific variance, V_Y , and the estimated between-study variance caused by possibly inflated post-intervention SD.

ance T^2 .

$$W_i = \frac{1}{V_Y + T^2} \quad (5)$$

For a more detailed derivation of the between-study variance estimate T^2 , see Borenstein et al. (2021, p. 73)

Furthermore, the random-effects model assumes that effect sizes are independent of each other. This assumption is violated in the given dataset, where studies contribute multiple effect sizes and the same control group is used as a reference across intervention arms. This leads to high correlation between effect sizes, and including all such effects would artificially increase the precision of the estimates (Borenstein et al., 2021). One suggested approach to deal with the issue of dependency is to include only one estimate per study. Accordingly, the overall random-effects models are estimated using a single estimate per study⁵. For simplicity reasons, the first estimated effect size per study is used. A more sophisticated approach will be discussed below.

The between-study variance is estimated using the restricted maximum likelihood method (REML; Viechtbauer, 2005). Statistical inference was based on the Knapp-Hartung adjustment for standard errors and test statistics (Knapp and Hartung, 2003). The analysis was conducted in R using the metafor package (Viechtbauer, 2010).

3.6 Moderator Analytical Methodology

In addition to estimating the overall pooled effect, a moderator analysis is conducted to examine how effectiveness varies across study characteristics.

Initially, a subgroup analysis, as outlined by Borenstein et al. (2021), is undertaken. For this, the categorical variables conservation behavior, intervention strategy, and, in line with the study’s aim, the context variable are analyzed. The procedure is undertaken as outlined in the previous section. Resulting in differing sample sizes by moderator, based on independent effect sizes.

Recent advances in the literature argue against relying on subgroup analysis alone, given its limited statistical power, specifically if an exclusion of effect sizes is performed, and its inability to account for dependence among effect sizes (Tanner-Smith and Tipton, 2014; Cuijpers et al., 2021).

Hence, the subsequent analysis uses meta-regressions. Meta-regression extends the standard random-effects model by using study-level variables as moderators of between-study heterogeneity (Borenstein et al., 2021).

⁵A study is hereafter understood as an experiment, with one independent sample.

Furthermore, multilevel meta-regression models following Van den Noortgate et al. (2013) can account for the nested effect-size structure present in the dataset.

The multilevel regression follows:

$$g_{ij} = \beta_0 + \sum_{p=1}^P \beta_p X_{pij} + e_{ij} + \varepsilon_{ij} + u_i, \quad (6)$$

where g_{ij} is the j -th Hedges' g within an independent sample i , X_{pij} denotes the P moderator variables with coefficients β_p , u_i captures between-sample heterogeneity, e_{ij} captures the residual within-sample variation, and ε_{ij} is the level-1 sampling error with known variance v_{ij} .

Nonetheless, the multi-level regression depends on the frequency of effect sizes across moderators and levels (Van den Noortgate et al., 2013; Cheung, 2014). Without an adequate number in each cell, the model's predictive capacity diminishes. Two supplementary data manipulations are employed to enhance the reliability of inference within the multi-level approach. First, a minimum requirement of 10 effect sizes per intervention strategy is established. This excludes the decision structural nudge ($k = 2$) and goal-setting ($k = 9$) strategies. These exclusions remove eleven effect sizes and two studies, yielding an analysis sample of $k = 158$ effect sizes from $ns = 97$ studies. Secondly, the conservation behaviors related to food waste and towel usage are aggregated into the broader categories of waste and water.

Statistical inference on the moderators is based on cluster-robust variance estimation (CR2) with Satterthwaite degrees-of-freedom approximation (Pustejovsky and Tipton, 2018), clustering on the independent sample, to guard against downward bias in standard errors arising from the small number of clusters (Hedges et al., 2010).

Whether each moderator holds explanatory value on its own is tested using the Wald test (Pustejovsky, 2025). The omnibus Wald test evaluates whether the coefficients associated with a moderator are jointly different from zero, equivalently, whether mean effect sizes differ across the moderator's levels. A significant test, therefore, provides evidence that the moderator accounts for part of the between-study heterogeneity. The accompanying R^2_{pseudo} quantifies this as the proportional reduction in the residual variance components relative to the intercept-only null model, i.e., the share of heterogeneity explained by the moderator.

3.7 Publication Bias

The meta-analytic framework is well-suited to detect publication bias, which occurs because statistically significant results are more likely to be published (Thornton and Lee, 2000). This bias arises when journals favor such results or when researchers perceive sig-

nificant findings as more important, often leading to the neglect of insignificant studies, a phenomenon known as the file drawer effect (Rosenthal, 1979). Hence, by analyzing the papers used in this analysis with a funnel plot, it is possible to gauge whether a publication bias exists. The funnel plot graphs the effect size (Hedges' g) against its precision (standard error (SE)). A visual check is performed to see if the distribution of effect sizes is symmetrical around the pooled effect size. Typically, more precise and larger studies are near the pooled estimate, whereas less precise effect sizes tend to deviate more, forming an inverted V-shaped pattern. When there is no publication bias, the funnel plot is expected to be symmetric around the pooled effect size; otherwise, there is an empty area in the bottom right corner of the graph. The funnel plot was generated using independent effect sizes. Additionally, the publication bias was assessed using Egger's regression test (Egger et al., 1997).

4 Results

This section reports the meta-analytical findings. It begins with descriptive results summarizing key characteristics of the included studies. The next part presents the results of the random-effects models, including the overall effect estimate, indicators of heterogeneity, and related subgroup analyses. This is followed by the final subsection, which details the meta-regression results and explores whether study-level moderators explain variation in effect sizes. Table A2 in the appendix provides the final dataset, presenting the used studies with their respective coding and Hedges' g values.

4.1 Descriptive Results

Following the literature search and screening process, 77 papers were included in the final sample. This sample yielded a total of 169 effect sizes, 99 of which were based on independent control groups.

Although the literature search covered studies published from 2000 onward, the earliest study in the final sample was published in 2008, and the latest in early 2026⁶, a paper first published online in 2023. The temporal distribution of papers and the intervention strategies explored through the years are presented in Figure 1. The largest number of effect sizes was published in 2011, attributable to the Opower study by Allcott (2011), which contributes 14 independent samples. The figure also shows that the interventions were implemented consistently over time, rather than being concentrated at specific points. The geographical distribution of the evidence base was concentrated in

⁶The most recent study is a pre-print version of: Paunov and Grüne-Yanoff (2026).

Europe, which accounted for 62 effect sizes (36)⁷, followed by Asia with 48 (21) and North America with 42 (33). Comparatively few effect sizes originated from Africa, Oceania, and Latin America, with 6 (3), 6 (3), and 5 (3), respectively. With respect to intervention strategies, social comparison interventions were the most frequently studied category, accounting for 69 effect sizes, of which 51 involved independent control samples. Appeal prompts and reminders represented the second largest category, comprising 34 effect sizes (22). Decision-structural nudges were the least represented intervention type, with only two effect sizes (2), both related to food waste reduction.

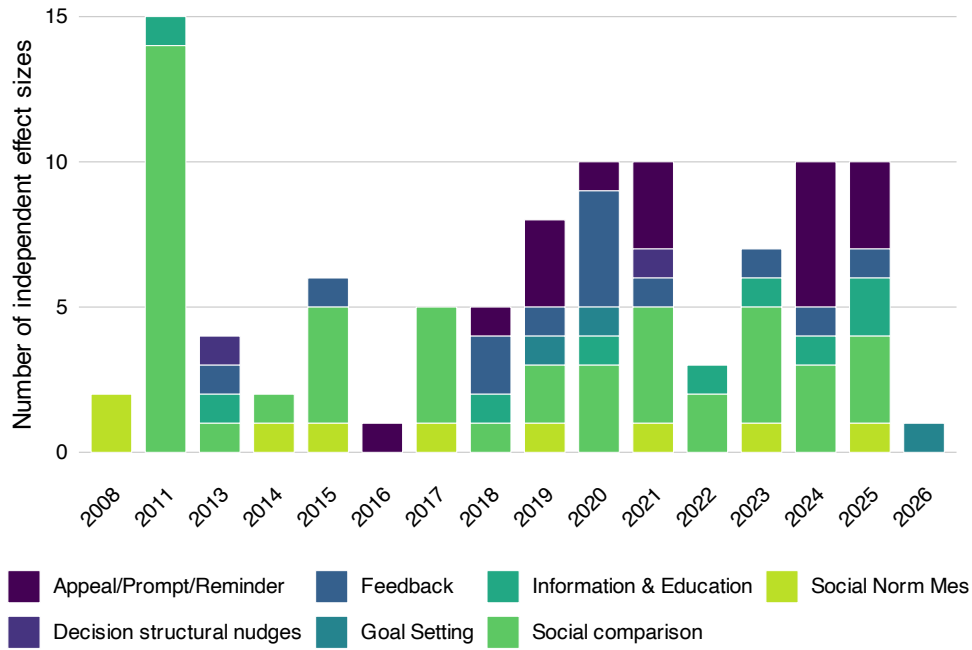


Figure 1: Number of effects per intervention type, by year

Figure 2 illustrates the distribution of conservation behaviors across intervention strategies. Energy conservation constitutes the largest share of the evidence base, with diverse intervention types tested. In particular, the prevalence of social comparison interventions is largely driven by studies targeting energy-use behavior, accounting for 43 (32) of the 69 effect sizes. In contrast, decision-structural nudges were identified exclusively in studies addressing food waste reduction. Informational interventions accounted for 10 effect sizes. This is in contrast to the higher number of informational interventions identified in the study by Nisa et al. (2019), which is attributable to differences in inclusion criteria. Their coding scheme is defined more broadly and, among other things, includes investment decision studies.

⁷Figures in parentheses report the number of effect sizes based on an independent control sample.

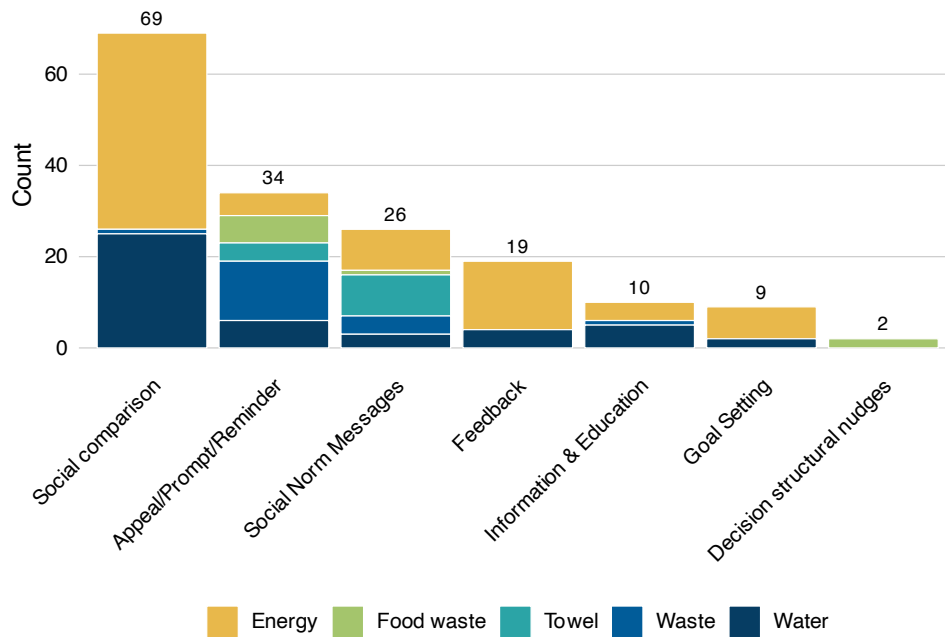


Figure 2: Number of effects per intervention type, broken down by conservation behavior, using the full sample of 169 effect

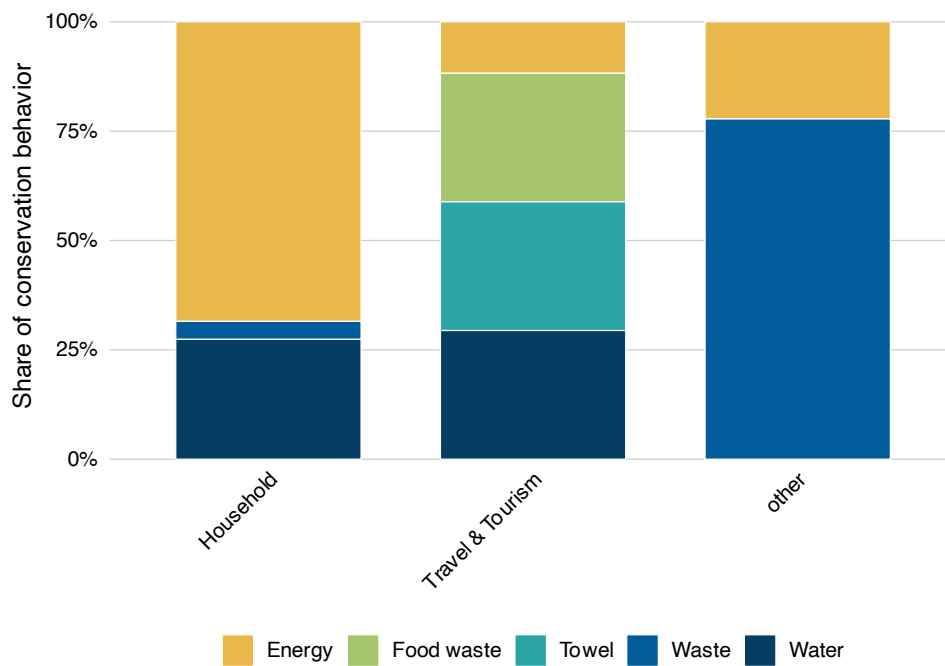


Figure 3: Share of conservation behavior by setting

Studies included in the final sample were primarily conducted in household settings (118, 73), including student dormitories, as well as in tourism contexts (32, 17) such as hotels, restaurants, and natural attractions, and in other settings (19, 9), including

university campuses, Christmas markets, and more. This can also be seen in Figure 3.

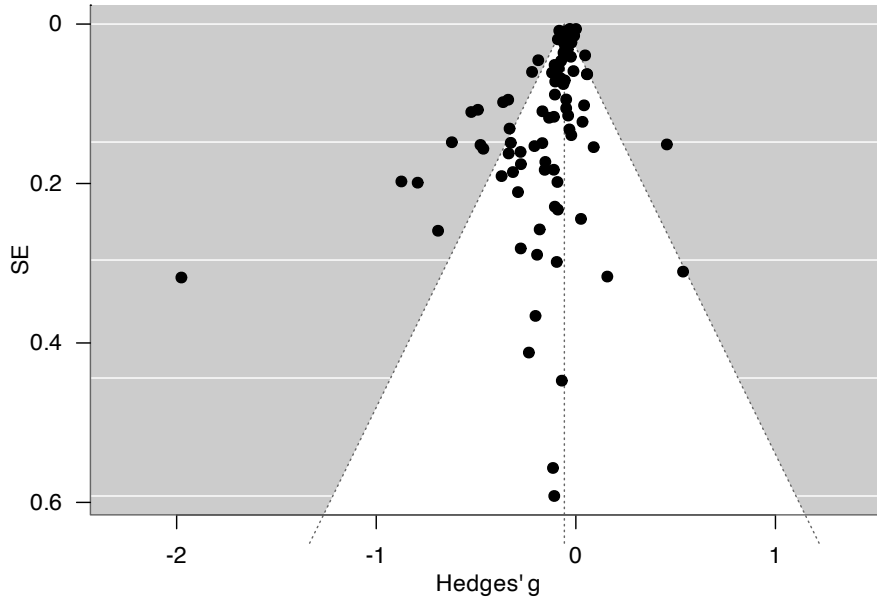


Figure 4: Funnel plot presenting the relationship between precision, measured as standard error, and effect size. Calculated as Hedges' g

The funnel plot for effect sizes with independent control groups is presented in Figure 4. A concentration of high-precision studies with close-to-zero or negative effect sizes is observed at the top, indicating small statistically significant effects. Moreover, effect sizes appear more frequently on the left side of the pooled effect estimate than on the right. As described in the methodology section, a symmetrical funnel plot indicates the absence of publication bias, whereas asymmetry may suggest small-study effects or publication bias. The visual asymmetry observed in Figure 4 was further examined using Egger's regression test, which supports funnel asymmetry ($b = -0.0253$, $p < 0.0001$), consistent with small-study effects. Despite the general conclusion about the effectiveness of behavioral interventions on conservation behavior remaining, the overrepresentation of favorable estimates in the study sample implies that the real effect may be smaller than the estimated pooled effect size presented. To mitigate potential small-sample bias, the random-effects models and the three-level meta-regression are estimated using the restricted maximum likelihood method. Statistical inference is subsequently adjusted for by using the Knapp-Hartung modifications for random-effects models and the small-sample corrections recommended by Tipton (2015).

In addition to the small-sample bias observed, Figure 5 and Figure 6 show that intervention strategies and conservation behaviors are not independently distributed: several strategies appear almost exclusively in combination with a single behavior (decision struc-

tural nudges for food waste and goal setting for energy) while some behaviors, such as towel and waste, have been studied with only a limited number of strategies. Notably, all studies that utilize social comparison as an intervention strategy are conducted in household settings.

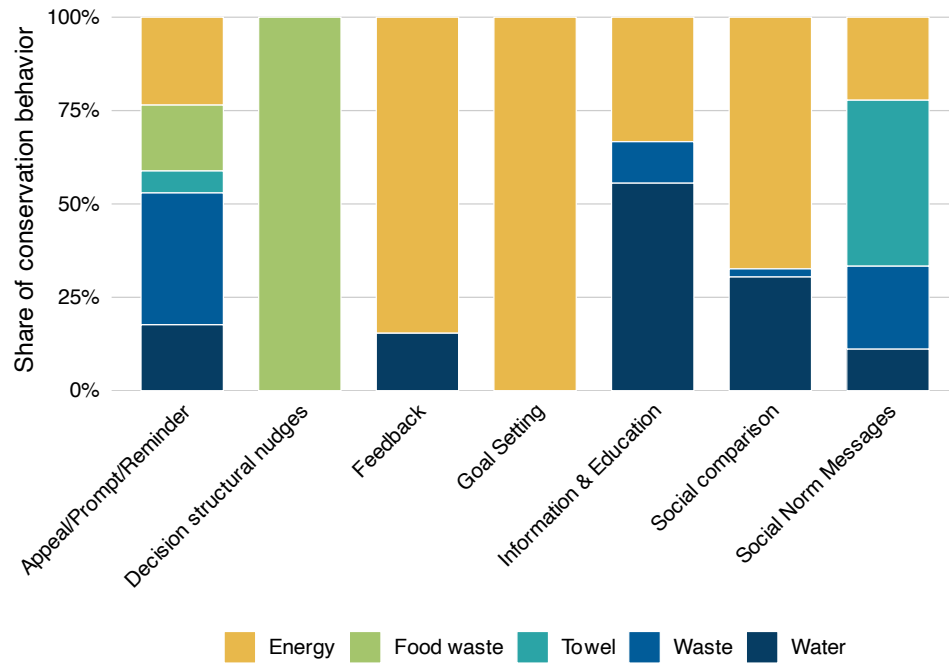


Figure 5: Share of conservation behavior by intervention type

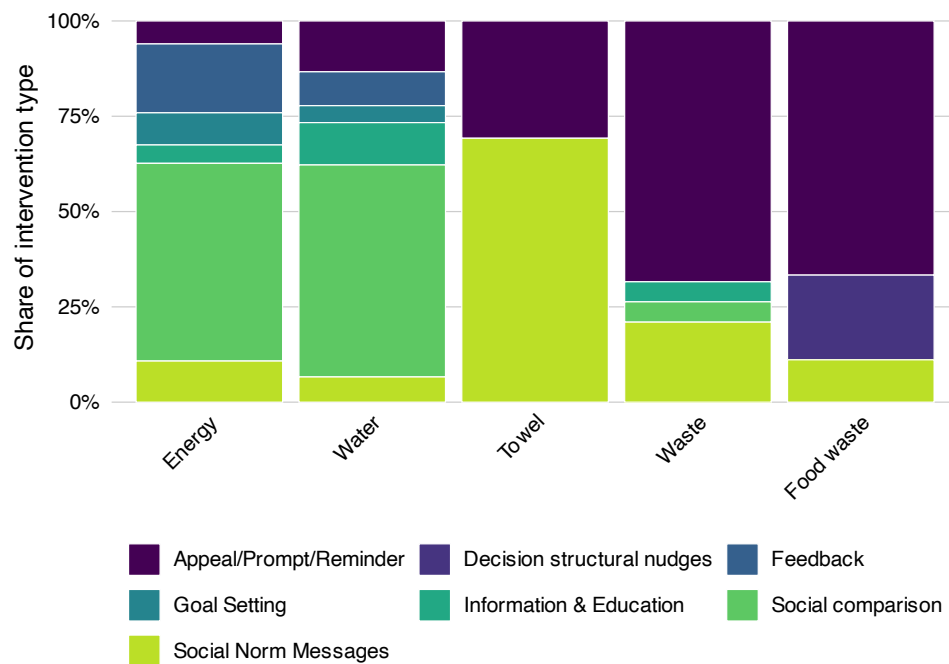


Figure 6: Intervention-type mix within each conservation behavior

4.2 Meta Analytical Results

Table 1 presents the results of the simple random-effects model and the sample obtained after applying the elimination strategy. The overall pooled effect size is -0.054 (95% CI: -0.071, -0.038). Although statistically significant, this effect size is considered very small according to the standard reference, which classifies absolute effects around 0.2 as small, around 0.5 as medium, and around 0.8 as large (Cohen, 1988). Table 1 further reports the results of the subgroup analysis. The intervention subgroups showed considerable variation in their estimated effect sizes. Information and education interventions ($g = -0.039$, 95% CI: -0.199, 0.122) had a very small, statistically non-significant effect, whereas social norm messages had a small and statistically significant effect ($g = -0.305$, 95% CI: -0.474, -0.136). Social comparison interventions have been used most frequently and have an effect size slightly below the overall pooled effect, at $g = -0.051$ (95% CI: -0.063, -0.039). The largest effect in terms of magnitude is in the appeal, prompt, and reminder subgroup ($g = -0.368$; 95% CI: -0.582, -0.155). Feedback interventions have an estimated effect size of $g = -0.139$ (95% CI: -0.219, -0.059), whereas goal-setting interventions have an effect size of $g = -0.215$, which is not distinguishable from zero given its wide confidence interval (95% CI: -0.733, 0.304). Except for the feedback intervention subgroup (I^2 : 68%), each intervention subgroup exhibits substantial heterogeneity I^2 ranging from 77% to 98%. The resource domain also exhibits some variation in its estimated effect sizes. While the towel and food waste subgroups have negative point estimates, neither effect is statistically significant ($g = -0.296$, 95% CI: -0.622, 0.030; $g = -0.500$, 95% CI: -1.503, 0.502). Conversely, energy and water are statistically significant, with similar effect sizes: $g = -0.051$ (95% CI: -0.068, -0.033) and $g = -0.041$ (95% CI: -0.065, -0.017). Lastly, studies examining waste reduction have an effect size of $g = -0.242$ (95% CI: -0.410, -0.073). For the resource domain subgroups, residual heterogeneity was moderate to substantial for the energy, water, and towel subgroups (66%, 63%, 62%); only the waste and food waste groups show very high I^2 values of 97% and 98%, respectively.

Table 1: Meta-analytic estimates by moderator subgroup.

Moderator	k	Effect size g (95% CI)	τ	I^2
Overall Model	99	-0.054*** (-0.071, -0.038)	0.030	66.9%
Setting				
Household	73	-0.044*** (-0.055, -0.033)	0.020	56.4%
Travel & Tourism	17	-0.240* (-0.475, -0.006)	0.380	95.2%
Other	9	-0.317** (-0.461, -0.174)	0.124	35.0%
Intervention type				
Appeal/Prompt/Reminder	22	-0.368** (-0.582, -0.155)	0.419	97.8%
Feedback	17	-0.139** (-0.219, -0.059)	0.113	68.2%
Information & Education	10	-0.039 (-0.199, 0.122)	0.187	98.0%
Social Comparison	51	-0.051*** (-0.063, -0.039)	0.028	76.8%
Social Norm Messages	17	-0.305** (-0.474, -0.136)	0.289	92.8%
Goal Setting	6	-0.215 (-0.733, 0.304)	0.428	92.3%
Behavior domain				
Energy	55	-0.051*** (-0.068, -0.033)	0.026	65.9%
Water	26	-0.041** (-0.065, -0.017)	0.026	62.9%
Towel	5	-0.296 (-0.622, 0.030)	0.185	61.8%
Waste	10	-0.242* (-0.410, -0.073)	0.206	97.1%
Food waste	5	-0.500 (-1.503, 0.502)	0.757	97.9%
Sample size				
<100	17	-0.236 (-0.501, 0.029)	0.411	65.9%
100-499	32	-0.226*** (-0.320, -0.133)	0.210	70.2%
500+	50	-0.043*** (-0.054, -0.033)	0.020	66.3%

Note. k = number of effect sizes; τ = standard deviation of true effects; I^2 = proportion of variance due to heterogeneity; * $p < .05$, ** $p < .01$, *** $p < .001$.

Coming to the setting variable. As described earlier, most studies are conducted in a household setting ($k = 73$); the estimated effect is statistically significant ($g = -0.044$, 95% CI: -0.055, -0.033). Furthermore, the household setting exhibits moderate heterogeneity, with the residual heterogeneity I^2 at 56.4%. In the travel and tourism subgroup, the effect size is at $g = -0.240$ (95% CI: -0.475, -0.006), with a very high level of residual heterogeneity. The other setting had the largest effect size ($g = -0.317$, 95% CI: -0.461, -0.174), despite the smaller number of effects and lower heterogeneity ($I^2 = 35\%$).

Notably, the subgroup analysis by sample size reveals that the estimated effect size decreases as sample size increases. Studies with fewer than 100 participants show the largest but least precise estimate ($g = -0.236$, 95% CI: -0.501, 0.029), which is non-significant, whereas the 100-499 band is significant at a similar magnitude ($g = -0.226$, 95% CI: -0.320, -0.133). Studies using samples with 500+ subjects yielded an effect size of $g = -0.043$ (95% CI: -0.054, -0.033). This inverse relationship is in line with the small-study effects already indicated by the funnel plot and Egger's test.

As discussed in the methodology section, the subgroup analysis is complemented by a three-level meta-regression to assess the robustness of the findings. Unlike the subgroup analysis, which was restricted to independent effect sizes and therefore retained only slightly more than 50% of the available data, the three-level meta-regression incorporates all effect sizes while accounting for their dependence.

Table 2: Three-level meta-analytic estimates by moderator subgroup

Moderator	k	Effect size g (95% CI)	df
Overall	169	-0.144*** (-0.186, -0.101)	91.0
Setting			
Household	110	-0.082*** (-0.105, -0.058)	63.2
Travel & Tourism	29	-0.284** (-0.473, -0.096)	14.0
Other	19	-0.392** (-0.631, -0.153)	6.6
Intervention type			
Appeal/Prompt/Reminder	34	-0.221** (-0.375, -0.066)	10.2
Feedback	19	-0.123** (-0.205, -0.041)	11.5
Information & Education	10	-0.124 (-0.287, 0.038)	4.9
Social Comparison	69	-0.096*** (-0.141, -0.052)	43.9
Social Norm Messages	26	-0.245* (-0.432, -0.057)	11.1
Behavior domain			
Energy	76	-0.118*** (-0.156, -0.079)	44.2
Water	56	-0.115*** (-0.160, -0.070)	17.8
Waste	26	-0.310** (-0.525, -0.095)	11.6
Sample size			
<100	31	-0.349* (-0.613, -0.085)	10.9
100–499	54	-0.243*** (-0.350, -0.135)	30.1
500+	73	-0.068*** (-0.091, -0.046)	49.6

Note. Three-level meta-analytic estimates by moderator subgroup. Using REML on the full analysis sample ($k = 158$, $n_s = 97$ studies). Inference uses cluster-robust variance estimation (CR2) with Satterthwaite degrees of freedom (df). k = number of effect sizes. * $p < .05$, ** $p < .01$, *** $p < .001$.

Meta regression results

Table 2 presents the subgroup estimates from the three-level meta-regression with cluster-robust variance estimation (CR2). The overall estimated effect size is $g = -0.144$ (95% CI: -0.186, -0.101), which is considerably larger in magnitude than the estimate obtained from the simple random-effects model ($g = -0.054$). This difference reflects the inclusion of dependent effect sizes in the three-level model, whereas the simple random-effects model was based on the reduced sample resulting from the elimination strategy. Despite the larger overall estimate, the relative pattern of subgroup effects remains consistent across the two approaches. Across intervention types, social norm messages continue

to show the largest estimated effect ($g = -0.245$, 95% CI: $-0.432, -0.057$), closely followed by appeals, prompts, and reminders ($g = -0.221$, 95% CI: $-0.375, -0.066$) and feedback interventions ($g = -0.123$, 95% CI: $-0.205, -0.041$). However, information and education interventions remain statistically indistinguishable from zero ($g = -0.124$, 95% CI: $-0.287, 0.038$), while social comparison interventions yield the smallest estimated effect size ($g = -0.096$, 95% CI: $-0.141, -0.052$). Compared with the simple random-effects model, the estimated effects change in magnitude across all intervention types, although their relative ordering remains largely unchanged. A similar pattern is observed across the resource domains. Using the combined waste subgroup, it exhibits the largest estimated effect ($g = -0.310$, 95% CI: $-0.525, -0.095$), whereas energy and water conservation show comparable effects of $g = -0.118$ (95% CI: $-0.156, -0.079$) and $g = -0.115$ (95% CI: $-0.160, -0.070$), respectively. The results for the setting moderator are also consistent with the previous analysis. The household setting yields the smallest estimated effect size ($g = -0.082$, 95% CI: $-0.105, -0.058$), whereas larger effects are observed for travel and tourism ($g = -0.284$, 95% CI: $-0.473, -0.096$) and the "other" setting ($g = -0.392$, 95% CI: $-0.631, -0.153$). Thus, although the three-level model produces larger effect estimates overall, the relative differences between settings closely resemble those observed in the simple random-effects model. Finally, the inverse relationship between sample size and effect size persists in the three-level analysis. Studies with fewer than 100 participants show the largest estimated effect ($g = -0.349$, 95% CI: $-0.613, -0.085$), followed by studies with sample sizes between 100 and 499 ($g = -0.243$, 95% CI: $-0.350, -0.135$). Whilst studies with more than 500 participants yield a substantially smaller estimated effect ($g = -0.068$, 95% CI: $-0.091, -0.046$), reinforcing the inverse association between sample size and effect magnitude observed in the simple random-effects analysis. Additionally, to the estimated effect sizes, Satterthwaite-approximated degrees of freedom are reported, to indicate the reliability of the CR2 small-sample inference. With the exception of the information and education subgroup ($df = 4.9$), all subgroup estimates exceed the commonly recommended minimum of approximately five degrees of freedom for reliable inference (Tipton, 2015). Nevertheless, several subgroups are based on relatively few degrees of freedom (e.g., Other, $df = 6.6$; Social Norm Messages, $df = 11.1$), and the corresponding confidence intervals and hypothesis tests should therefore be interpreted with some caution.

The consistently larger effect estimates obtained from the three-level model suggest that the elimination strategy used in the simple random-effects analysis may influence the magnitude of the pooled estimates. Table 3 presents the Wald tests on the analysis sample ($k = 158$, studies (ns) = 97), examining whether each moderator explains between-study heterogeneity. Unlike Nisa et al. (2019), who estimated a single model including all moderators simultaneously, this study estimates separate models for each moderator.

This approach was adopted because the household setting and the social comparison intervention are perfectly confounded in the analysis sample, making it inappropriate to estimate their unique effects within a single model. The omnibus Wald test is statistically significant for the setting moderator, $F(2, 15.2) = 6.83$, $p = 0.008$, $R^2_{pseudo} = 41.0\%$, and for sample size, $F(2, 27.5) = 7.08$, $p = 0.003$, $R^2_{pseudo} = 43.3\%$. Intervention type does not reach statistical significance ($p = 0.522$) despite explaining a nontrivial share of between-study heterogeneity R^2_{pseudo} of more than 25%. The behavior domain has low explanatory power ($p = 0.357$, $R^2_{pseudo} = 5.0\%$). These results provide no evidence of detectable differences in effect sizes across intervention types or behavior domains. This null finding should be interpreted with caution: both the low degrees of freedom in the denominator and the substantial within-subgroup heterogeneity documented in the subgroup analysis (I^2 up to 98%) reduce the power of the omnibus test to detect between-level differences, so the absence of a significant result does not imply that the moderators are unrelated to the observed heterogeneity. In comparison, in their single-model setup Nisa et al. (2019, p. 6) found significant moderating effects for two intervention strategies (social comparison and nudges), the sample size variable, and the waste resource domain.

Table 3: Marginal omnibus Wald tests by moderator

Moderator	F	df_1	df_2	p	R^2_{pseudo}
Setting	6.83	2	15.2	0.008	41.0%
Intervention type	0.87	4	7.4	0.522	25.7%
Behavior domain	1.44	2	3.2	0.357	5.0%
Sample size	7.08	2	27.5	0.003	43.3%

Note. Marginal omnibus Wald tests were estimated using CR2 cluster-robust variance estimation with the Hotelling–Zhang (HTZ) approximation on the full analysis sample ($k = 158$, $n_s = 97$). Models include the social comparison intervention category. R^2_{pseudo} values are calculated relative to the null model.

Because all social comparison interventions were conducted in household settings, a second set of omnibus Wald tests was estimated after excluding the social comparison subgroup to assess whether the observed setting effect persisted once this confounding was removed. The results, presented in Table 4, show that setting remains significant, $F(2, 20.7) = 3.92$, $p = 0.036$, but with a substantially lower R^2_{pseudo} (10.3%). The setting moderator, therefore, remains statistically significant after excluding the social comparison studies, although both its explanatory power and statistical significance are substantially reduced. The sample size moderator no longer reaches significance, $F(2, 19.1) = 2.57$, $p = 0.103$, with a R^2_{pseudo} of 15%. The loss of significance should be interpreted cautiously. Because the social comparison studies are closely associated with both the household setting and larger sample sizes, removing them simultaneously alters several characteristics of the

dataset. Accordingly, the differences between Table 3 and Table 4 should be interpreted as reflecting changes in the composition of the analysis sample.

Table 4: Marginal omnibus Wald tests by moderator (social comparison excluded).

Moderator	F	df_1	df_2	p	R^2_{pseudo}
Setting	3.92	2	20.7	0.036	10.3%
Intervention type	1.69	3	12.2	0.221	2.8%
Behavior domain	1.19	2	31.1	0.317	1.8%
Sample size	2.57	2	19.1	0.103	15.0%

Note. Marginal omnibus Wald tests were estimated using CR2 cluster-robust variance estimation with the HTZ approximation. The subsample consists of $k = 89$, $n_s = 53$. R^2_{pseudo} values are calculated relative to the social-comparison excluded null model.

Overall, the results of the three-level meta-regression indicate that behavioral interventions are associated with a modest reduction in resource use, with an estimated overall effect size $g = -0.144$. The moderator analyses identify setting as the most consistent source of variation in effect sizes, with both the subgroup analyses and the omnibus Wald tests providing evidence of differences across settings. In contrast, although the subgroup analyses suggest that some intervention types and resource domains are associated with larger estimated effects than others, the omnibus Wald tests provide no evidence that these two moderators explain a significant proportion of the observed between-study heterogeneity. Notably, although the three-level meta-regression yielded a larger effect size estimate, the recognized pattern within subgroups remained stable across the applied methods. Finally, when interpreting the results of the moderator analyses, one should acknowledge the substantial confounding among several study characteristics, particularly between intervention type, setting, and sample size, which limits the extent to which the observed moderator effects can be interpreted independently.

5 Limitations

The present thesis is subject to several limitations relating both to its methodology and the interpretation of its findings that warrant discussion.

First, in the evidence synthesis literature, it has become common practice to rely on multiple reviewers. The use of multiple independent reviewers reduces the risk of selection bias, subjective interpretations, and coding errors. Considering the nature of this thesis, this approach was not feasible. To mitigate the limitations, it was decided to enhance the documentation effort to increase transparency and reproducibility. Nevertheless, even

though each step of the reviewing process has been carefully documented in line with accepted standards (e.g., flow diagram), conducting a meta-analysis is inherently subject to the author’s judgment. Consequently, inclusion and exclusion criteria, as well as moderator coding, may be subject to alternative interpretations. The decisions taken were guided by the objective of ensuring that the included interventions and contextual settings were sufficiently comparable to support a meaningful meta-analysis.

The second limitation concerns the use of AI and machine learning tools to assist with screening and data extraction. While AI-assisted evidence synthesis can substantially improve efficiency, it also introduces two main caveats. Firstly, the dependence on the models employed. AI was used in two phases of the study review process and in data extraction. Regarding title and abstract screening, ASReview has been shown to reliably assist in identifying relevant studies. To minimize the risk of prematurely excluding relevant studies, a conservative stopping criterion was adopted. For the full-text screening, a human-in-the-loop approach was employed to mitigate the risk of overreliance on publicly available LLM results. The data extraction was similar to the full-text screening: AI-assisted but human-in-the-loop. Secondly, behavioral sciences have warned against underestimating the effects of anchoring (Tversky and Kahneman, 1974). The application used for the full-text screening and data extraction is currently designed as an opt-out tool: AI suggestions are presented before the reviewer has read the paper, potentially biasing the reviewer toward the suggested decision/data. Nevertheless, residual dependence on model performance cannot be entirely ruled out. As AI-assisted evidence synthesis becomes more widespread, standardized reporting conventions should be developed for more transparent use and evaluation of AI-assisted workflows. Further, as with reporting inter-reviewer agreement in conventional synthesis studies, future AI-assisted reviews could report an AI-reviewer agreement statistic.

Third, another key limitation concerns the confounding among moderators. The proposed setting moderator is strongly correlated with sample size, conservation behavior, and intervention strategy, making it difficult to attribute the observed effect to the setting rather than to the factors it covaries with. Several model specifications were estimated to probe the robustness of the results, and setting remains a significant moderator even after adjusting for sample size. Nonetheless, the association cannot be cleanly attributed to the setting alone. Achieving this requires primary research in which the setting varies independently, such as large-scale tourism studies or smaller household studies, rather than the present confounded distribution. This scarcity of data linking intervention type to behavior domain makes it difficult to attribute effects to the intervention strategy itself and, therefore, to recommend any single intervention type.

Finally, the proposed explanation that intervention effectiveness varies across settings

because contextual factors serve as proxies for habit strength should be interpreted with caution. The present meta-analysis identifies an association at the study level rather than examining behavioral mechanisms underlying individual responses. Consequently, the argument provides indirect support for the hypothesis that stronger habits reduce responsiveness to behavioral interventions, but does not constitute a direct test of the mechanism. Confirming this relationship would require individual-level studies that directly measure habit strength and examine whether it moderates the effectiveness of interventions across different settings.

6 Conclusion

This thesis synthesizes field-experimental evidence on the effectiveness of non-price behavioral interventions for resource conservation and examines whether effectiveness depends on the targeted behavior’s habitual nature. For screening and data extraction, AI tools assist the review process. First, can non-price interventions cultivate behavioral change and encourage resource conservation? Yes, in total, 169 effect sizes from 99 studies taken from 77 papers are synthesized, and the overall effect size is given as a Hedges’ g of -0.144. This is after a variety of state-of-the-art meta-analytical tools are applied to correct for underlying biases. Nonetheless, due to small-study bias, this result should be understood as an upper-bound estimate. The small magnitude is consistent with a broader convergence toward a smaller effect size in the literature on environmental behavioral interventions, particularly as more rigorous experimental evidence with larger sample sizes accumulates (Nisa et al., 2019). Generally, there is high heterogeneity across intervention types and contexts, which reveals that there is no silver bullet. Although the pooled effect is small, this should not be taken as evidence that behavioral interventions are without value: small effects can add up to meaningful change, particularly when implemented frequently (Funder and Ozer, 2019). Rather than discouraging their use, this modest magnitude argues for greater rigor in implementing such interventions, moving beyond isolated, one-off designs toward approaches that explicitly account for known limitations, such as the risk of boomerang effects or the need to calibrate psychological distance in social-comparison interventions. Second, is the effectiveness of given interventions moderated by the contextual setting in which they are implemented? The main moderators of the effectiveness of behavioral interventions to foster conservation behavior are sample size and the contextual setting in which the intervention is implemented. The results of a three-level meta-regression indicate a greater effectiveness of non-price interventions in tourism settings than in household settings, but the result is confounded with sample size. Household studies are regularly performed with larger samples.

Nonetheless, interventions appear less effective in contexts where behavior is habitual. This pattern aligns with the moderating role of habit strength in shaping behavior, as proposed by Verplanken and Aarts (1999). The setting variable, which could be interpreted as a proxy for habitual behavior, may be considered indirect evidence that habit strength differs with setting. Yet, because setting serves as an indicator of habit strength rather than a direct measure, and is confounded by the type of intervention, the behavior domain, and, most importantly, sample size, the causal inference regarding habit remains tentative. This thesis provides exploratory evidence that situational context is associated with the effectiveness of behavioral interventions. Moreover, to gain deeper insight into whether a setting may actually be a good proxy for habit strength, studies are needed that directly measure the underlying mechanisms and may use a broader range of behaviors. Notably, a closely related strand of the literature is the Habit Discontinuity Hypothesis (Verplanken et al., 2008, 2018), which posits that during periods of change, habitual patterns are altered and the likelihood of adopting new habits increases. Studies examining different life events, e.g., going to college or changing jobs, and their effects on routines support the notion that timed interventions may be promising (Verplanken and Roy, 2016). Future meta-analysis could synthesize the evidence to evaluate whether interventions are indeed more effective and lasting when timed to such transitions. In conclusion, the findings indicate that the efficacy of non-price interventions is influenced more by the context in which they are implemented than by the specific strategy selected. This underscores the importance of focusing on the targeting and timing of interventions, rather than seeking a singular optimal strategy, as a more promising direction for future research.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211.
- Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, 32(4):665–683.
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10):1082–1095.
- Asensio, O. I. and Delmas, M. A. (2015). Nonprice incentives and energy conservation. *Proceedings of the National Academy of Sciences*, 112(6):E510–E515.
- Bergquist, M., Thiel, M., Goldberg, M. H., and 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.
- Bhanot, S. P. (2017). Rank and response: A field experiment on peer information and water use behavior. *Journal of Economic Psychology*, 62:155–172.
- Bonan, J., Cattaneo, C., d’Adda, G., Galliera, A., and Tavoni, M. (2025). Social norms and tariff salience: An experimental study on household waste management. *Journal of Environmental Economics and Management*, 130:103124.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2):97–111.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.
- Brülisauer, M., Goette, L., Jiang, Z., Schmitz, J., and Schubert, R. (2020). Appliance-specific feedback and social comparisons: Evidence from a field experiment on energy conservation. *Energy Policy*, 145:111742.
- Carrero, I., Valor, C., Redondo, R., and Vilà, I. (2025). A meta-analysis of the effectiveness of implementation intentions for pro-environmental behavior adoption. *Sustainable Production and Consumption*, 55:63–75.
- Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, 19(2):211–229.

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge, New York, 2nd edition.
- Cuijpers, P., Griffin, J. W., and Furukawa, T. A. (2021). The lack of statistical power of subgroup analyses in meta-analyses: A cautionary note. *Epidemiology and Psychiatric Sciences*, 30:e78.
- Delmas, M. A., Fischlein, M., and Asensio, O. I. (2013). Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012. *Energy Policy*, 61:729–739.
- Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109):629–634.
- Funder, D. C. and 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.
- Giaccherini, M., Gilli, M., Mancinelli, S., and Zoli, M. (2021). Nudging food waste decisions at restaurants. *European Economic Review*, 135:103722.
- Greene, D., Zinn, A. K., Demeter, C., and Dolnicar, S. (2025). “hi, i’m terri towel. please reuse me.” can anthropomorphizing towels prompt tourists to reuse them? *Journal of Travel Research*, page 00472875251349233.
- Hammar, H. and Åkerfeldt, S. (2015). CO2 taxation in Sweden: Experiences of the past and future challenges. United Nations, Financing for Development Office. Available at https://www.un.org/esa/ffd/wp-content/uploads/2016/12/13STM_Article_CO2-tax_AkerfeldtHammar.pdf.
- Hedges, L. V. (1981). Distribution theory for glass’s estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2):107–128.
- Hedges, L. V., Tipton, E., and Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1):39–65.
- Higgins, J. P. T., Li, T., and Deeks, J. J. (2024). Chapter 6: Choosing effect measures and computing estimates of effect. In Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., and Welch, V. A., editors, *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane, version 6.5 edition. Chapter last updated August 2023. Available from <https://www.cochrane.org/handbook>.

- IPCC (2022). Summary for policymakers. In Shukla, P., Skea, J., Slade, R., Al Khourdajie, A., van Diemen, R., McCollum, D., Pathak, M., Some, S., Vyas, P., Fradera, R., Belkacemi, M., Hasija, A., Lisboa, G., Luz, S., and Malley, J., editors, *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA.
- Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5):1449–1475.
- Kallbekken, S. and Sælen, H. (2013). ‘Nudging’ hotel guests to reduce food waste as a win–win environmental measure. *Economics Letters*, 119(3):325–327.
- Knapp, G. and Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22(17):2693–2710.
- Legault, L., Bird, S., Powers, S. E., Sherman, A., Schay, A., Hou, D., and Janoyan, K. (2020). Impact of a motivational intervention and interactive feedback on electricity and water consumption: A smart housing field experiment. *Environment and Behavior*, 52(6):666–692.
- Lemmel, F. K., Hoppen, T. H., and Morina, N. (2025). Long-term effects of social norm interventions on pro-environmental behaviours: A systematic review and meta-analysis. *Journal of Environmental Psychology*, page 102784.
- Loschelder, D. D., Siepelmeyer, H., Fischer, D., and Rubel, J. A. (2019). Dynamic norms drive sustainable consumption: Norm-based nudging helps café customers to avoid disposable to-go-cups. *Journal of Economic Psychology*, 75:102146.
- Mi, L., Gan, X., Sun, Y., Lv, T., Qiao, L., and Xu, T. (2021). Effects of monetary and nonmonetary interventions on energy conservation: A meta-analysis of experimental studies. *Renewable and Sustainable Energy Reviews*, 149:111342.
- Morgan, M. and Chompreeda, K. (2015). The relative effect of message-based appeals to promote water conservation at a tourist resort in the gulf of Thailand. *Environmental Communication*, 9(1):20–36.
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2):364–386.
- Nemati, M. and Penn, J. (2020). The impact of information-based interventions on conservation behavior: a meta-analysis. *Resource and Energy Economics*, 62:101201.

- Nisa, C. F., Bélanger, J. J., Schumpe, B. M., and Faller, D. G. (2019). Meta-analysis of randomised controlled trials testing behavioural interventions to promote household action on climate change. *Nature Communications*, 10(1):4545.
- Osbaldiston, R. and Schott, J. P. (2012). Environmental sustainability and behavioral science: Meta-analysis of proenvironmental behavior experiments. *Environment and Behavior*, 44(2):257–299.
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.
- 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., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., and Moher, D. (2021). The prisma 2020 statement: an updated guideline for reporting systematic reviews. *Journal of Clinical Epidemiology*, 134:178–189.
- Paunov, Y. and Grüne-Yanoff, T. (2026). Boosting vs. nudging sustainable energy consumption: a long-term comparative field test in a residential context. *Behavioural Public Policy*, 10(1):42–67.
- Pellerano, J. A., Price, M. K., Puller, S. L., and Sánchez, G. E. (2017). Do extrinsic incentives undermine social norms? evidence from a field experiment in energy conservation. *Environmental and Resource Economics*, 67(3):413–428.
- Pigou, A. C. (1932). *The Economics of Welfare*. Macmillan & Co., London, 4th edition.
- Pustejovsky, J. E. (2025). *clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections*. R package version 0.6.0.
- Pustejovsky, J. E. and Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 36(4):672–683.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3):638.
- Schroeder, N. (2024). AIDE: AI-Assisted data extraction. R package, GitHub repository.

- Skvarc, D. R. and Fuller-Tyszkiewicz, M. (2024). Calculating repeated-measures meta-analytic effects for continuous outcomes: a tutorial on pretest–posttest-controlled designs. *Advances in Methods and Practices in Psychological Science*, 7(1):25152459231217238.
- Tanner-Smith, E. E. and Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, 5(1):13–30.
- Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.
- Thornton, A. and Lee, P. (2000). Publication bias in meta-analysis: its causes and consequences. *Journal of Clinical Epidemiology*, 53(2):207–216.
- Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., and Staake, T. (2018). Overcoming salience bias: How real-time feedback fosters resource conservation. *Management Science*, 64(3):1458–1476.
- Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, 20(3):375.
- Tonke, S. (2024). Providing procedural knowledge: A field experiment to encourage resource conservation in Namibia. *Journal of Development Economics*, 166:103202.
- Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157):1124–1131.
- Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481):453–458.
- Tversky, A. and Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, 59(4):S251–S278.
- United Nations Environment Programme (2022). Plastic pollution. <https://www.unep.org/plastic-pollution>. Accessed: 2025-09-03.
- Van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdemans, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G.-J., et al. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*, 3:125–133.

- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., and Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, 45(2):576–594.
- Verplanken, B. and Aarts, H. (1999). Habit, attitude, and planned behaviour: is habit an empty construct or an interesting case of goal-directed automaticity? *European Review of Social Psychology*, 10(1):101–134.
- Verplanken, B., Aarts, H., van Knippenberg, A., and Moonen, A. (1998). Habit versus planned behaviour: a field experiment. *British Journal of Social Psychology*, 37(1):111–128.
- Verplanken, B. and Roy, D. (2016). Empowering interventions to promote sustainable lifestyles: Testing the habit discontinuity hypothesis in a field experiment. *Journal of Environmental Psychology*, 45:127–134.
- Verplanken, B., Roy, D., and Whitmarsh, L. (2018). Cracks in the wall: Habit discontinuities as vehicles for behaviour change. In *The psychology of habit: Theory, mechanisms, change, and contexts*, pages 189–205. Springer.
- Verplanken, B., Walker, I., Davis, A., and Jurasek, M. (2008). Context change and travel mode choice: Combining the habit discontinuity and self-activation hypotheses. *Journal of Environmental Psychology*, 28(2):121–127.
- Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30(3):261–293.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3):1–48.
- Wang, J., Wen, L., Li, Y., Liu, J., and Che, S. (2025). Uncovering the impacts of social comparative information on water saving behavior among urban residents: A field experiment in China. *Journal of Cleaner Production*, 492:144863.
- Wansink, B. (2004). Environmental factors that increase the food intake and consumption volume of unknowing consumers. *Annual Review of Nutrition*, 24(1):455–479.
- Wilson, D. B. (2023). Practical meta-analysis effect size calculator. <https://www.campbellcollaboration.org/calculator/>. Version 2023.11.27. Campbell Collaboration. Accessed: 2025-12-21.
- Wood, W. and Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, 67(1):289–314.

Appendix

Prisma Flow Diagram

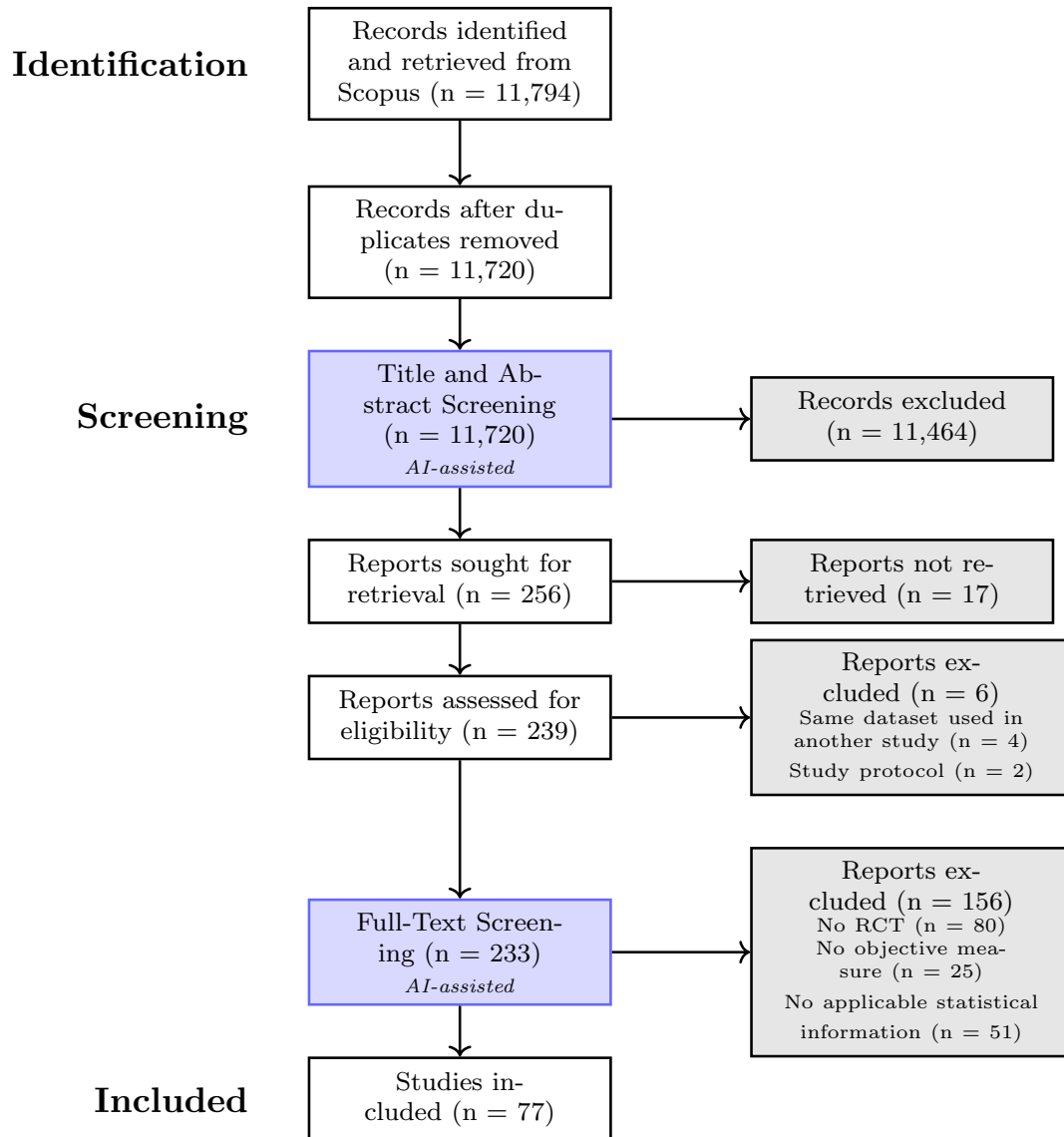


Figure A1: PRISMA flow diagram illustrating the study selection process

Boolean Search String

Table A1: Search strategy used for the literature search in Scopus

Concept	Search Terms
Experimental Design	"randomized controlled trial" OR "field stud*" OR "field experiment*" OR "control group" OR (random* AND (select OR alloc*))
Topic	energy OR electricity OR gas OR water OR towel OR heating OR cooling OR waste OR dispos*
Behavioral Intervention	"behavior* intervention*" OR "behaviour* intervention*" OR "non-price intervention" OR nudge* OR "choice architecture" OR "cognitive dissonance" OR "social comparison" OR "social norm*" OR labeling OR feedback OR information* OR "goal setting" OR "commitment device*" OR ((descriptive OR injunctive) AND norm*)
Behavior	reduc* OR decreas* OR limit* OR mitigat* OR control* OR conserv* OR wast* OR excess OR discard* OR surplus OR leftover OR resourc* OR consum* OR recycl* OR reuse OR ((increas* OR improv* OR promot* OR foster) AND (efficien* OR sav*))

Note. The four groups above (Experimental Design, Topic, Behavioral Intervention, Behavior) are connected with AND. The wildcard character (*) was used to capture variations of search terms

Screening Prompts

The following prompts were applied during the full-text screening to determine whether a paper is eligible for inclusion. Each prompt is an inclusion/exclusion check.

Random assignment. Does the study describe random assignment of participants, households, or other individual-level units to treatment and control/comparison groups? Include studies in which randomization occurs at a small aggregate level (e.g., building).

Objective outcome measurement. Is an observed behavior or action measured through direct observation, sensor data, administrative records, or other objective verification?

Extractable effect size. Does the paper report statistics that allow calculation of a standardized mean difference effect size (e.g., Cohen's d)? This includes: means and standard deviations with sample sizes; regression coefficients with standard errors; t -statistics, F -statistics, or p -values with sample sizes; odds ratios or risk ratios with confidence intervals; sample proportions; or pre-calculated effect sizes.

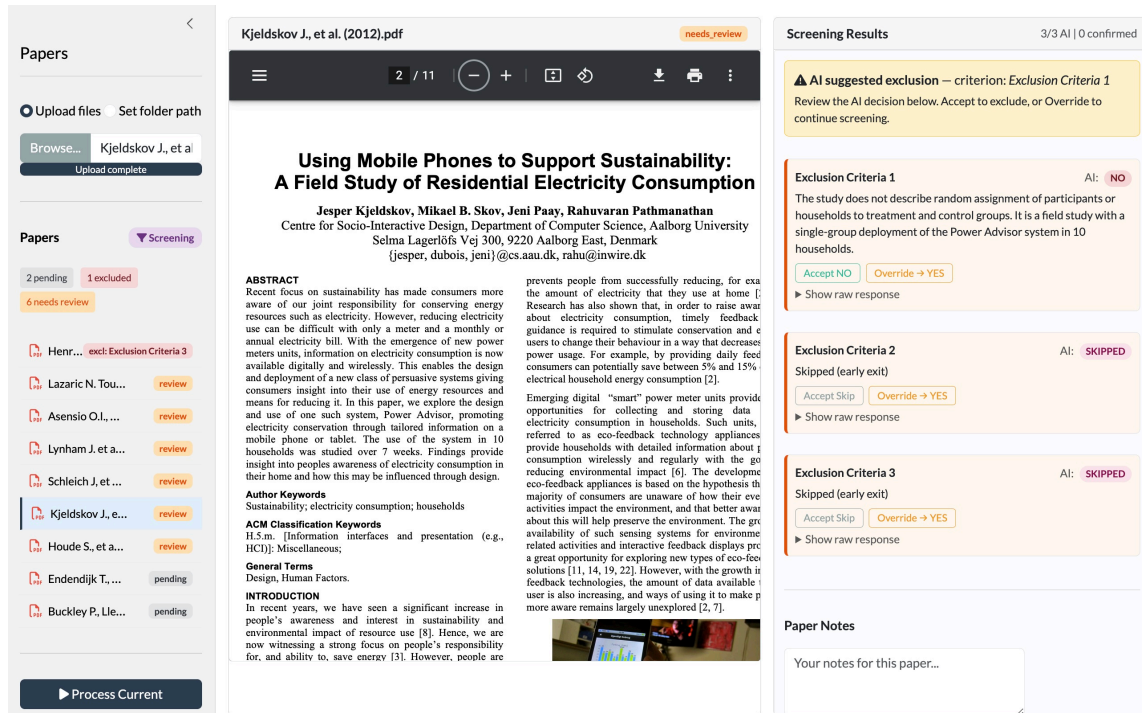


Figure A2: Screenshot showing the built application used for the full-text screening, screening mode

Data Extraction Prompts

The following prompts were applied in the data extraction process to extract study characteristics and statistical data.

Number of independent studies

How many separate, independent studies are reported in this paper? Count studies as independent only if they use non-overlapping samples and non-overlapping control or comparison groups. Provide a single integer. If no number of studies different from one is reported, answer: 1. Example: 2.

Primary behavior targeted

What is the primary behavior targeted by the intervention? Select ONE category that best fits: Energy consumption (electricity, gas, heating, cooling); Water consumption (shower, towel); Food waste; Household waste. If the intervention targets a behavior that does not fit any of these, state a new category in no more than 4 words. If a secondary behavior is targeted, list each. Example: Water consumption.

Intervention strategy

Which type(s) of behavioral intervention strategy was used? Select all that apply from: Information/education; Social comparison; Feedback; Choice architecture; Appeal/Prompt/Reminder; Social norm message; Competition/Gamification. If multiple strategies are combined

within one treatment arm, list all separated by '+'. Example: Social norm + Information.

Country of data collection

In which country or countries was the data collection conducted? Use full English country names. If multiple countries, list all separated by commas. If multiple studies were conducted in different countries, indicate per study. Example: United States; or: Study 1: Germany, Study 2: Netherlands.

Setting

In what setting was the intervention delivered? Select ONE of the following categories: Household; Public Space; Restaurant/Hotel; Workplace; School/University; Retail/Supermarket. If the setting clearly does not fit any of these, propose a new category in a single word or short phrase. Example: Household.

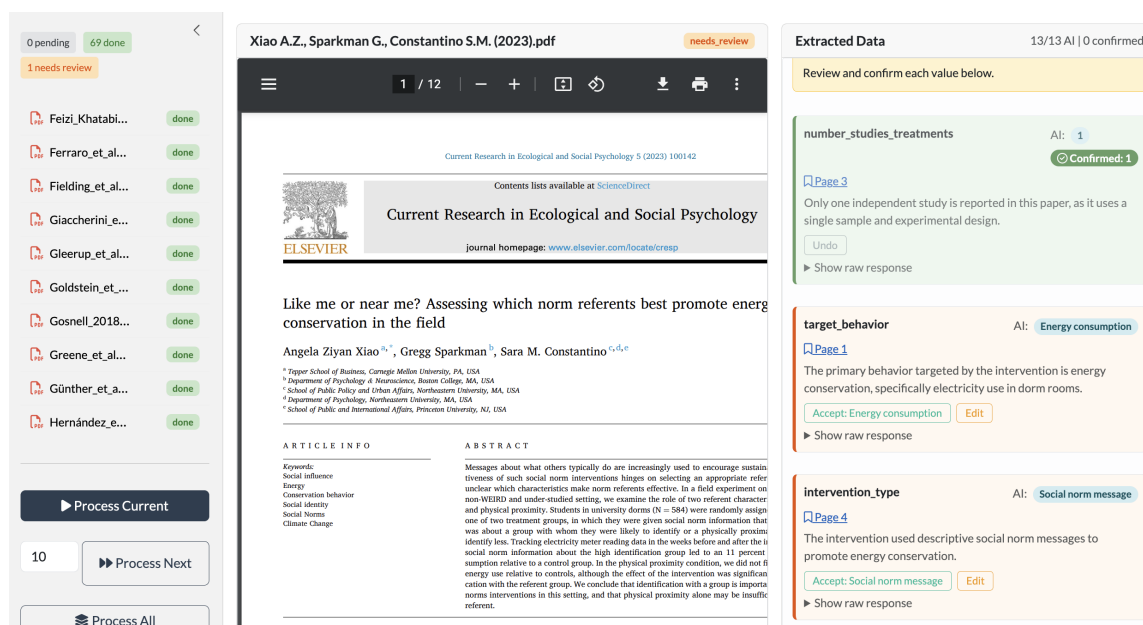


Figure A3: Screenshot showing the built application used for the data extraction, data extraction mode

Number of treatment arms

For each study, how many treatment arms (intervention conditions) are tested, excluding the shared control or comparison group? Provide a single integer. If only one study arm is tested, answer 1. Example: 2.

Post-intervention statistics

Report the post-intervention mean and standard deviation of the primary outcome variable, per experimental group (including the control). If only the mean is reported, state the mean and note "SD nr". If post-intervention mean and SD are not reported for any

group, report “NR”. Example: Control: $M=7.5$, $SD=2.1$; Treatment 1: $M=7.7$, $SD=2.3$. Or: Sample: $M=7.45$, SD not reported.

Sample size per arm

Report the sample size for each experimental arm, including the control group. State the unit of analysis (individuals, households, classrooms, etc.). If the unit is something other than individuals, also report the total number of individuals if available. Example: Control: 50 households; Treatment 1: 48 households; Treatment 2: 52 households (unit: household).

Declaration of Generative AI in Scientific Writing

During the preparation of this work, Claude (Anthropic) was used to improve the manuscript’s readability and language, to review and debug R code, and to assist in developing an R Shiny application for screening and data extraction, as detailed in the methodology section. After using this tool, the author reviewed, tested, and edited all outputs as needed and takes full responsibility for the content of the publication.

Meta-Analysis References

- Ahsanuzzaman, Eskander, S., Islam, A., and Wang, L. C. (2024). Non-price energy conservation information and household energy consumption in a developing country: Evidence from an RCT. *Journal of Environmental Economics and Management*, 127:103022.
- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10):1082–1095.
- Andor, M. A., Gerster, A., Peters, J., and Schmidt, C. M. (2020). Social norms and energy conservation beyond the US. *Journal of Environmental Economics and Management*, 103:102351.
- Asensio, O. I. and Delmas, M. A. (2015). Nonprice incentives and energy conservation. *Proceedings of the National Academy of Sciences*, 112(6):E510–E515.
- Asmare, F., Jaraitė, J., and Kažukauskas, A. (2021). The effect of descriptive information provision on electricity consumption: Experimental evidence from Lithuania. *Energy Economics*, 104:105687.
- Bator, R. J., Phelps, K., Tabanico, J., Schultz, W. P., and Walton, M. L. (2019). When it is not about the money: Social comparison and energy conservation among residents who do not pay for electricity. *Energy Research & Social Science*, 56:101198.
- Bergquist, M. and Nilsson, A. (2016). I saw the sign: Promoting energy conservation via normative prompts. *Journal of Environmental Psychology*, 46:23–31.
- Bergquist, M., Nilsson, A., and Ejelöv, E. (2019). Contest-based and norm-based interventions:(how) do they differ in attitudes, norms, and behaviors? *Sustainability*, 11(2):425.
- Bhanot, S. P. (2017). Rank and response: A field experiment on peer information and water use behavior. *Journal of Economic Psychology*, 62:155–172.
- Bhanot, S. P. (2021). Isolating the effect of injunctive norms on conservation behavior: New evidence from a field experiment in California. *Organizational Behavior and Human Decision Processes*, 163:30–42.
- Bohner, G. and Schlüter, L. E. (2014). A room with a viewpoint revisited: Descriptive norms and hotel guests’ towel reuse behavior. *PLOS ONE*, 9(8):e104086.

- Bonan, J., Cattaneo, C., d’Adda, G., Galliera, A., and Tavoni, M. (2024). Widening the scope: the direct and spillover effects of nudging water efficiency in the presence of other behavioral interventions. *Journal of Environmental Economics and Management*, 127:103037.
- Bonan, J., Cattaneo, C., d’Adda, G., Galliera, A., and Tavoni, M. (2025). Social norms and tariff salience: An experimental study on household waste management. *Journal of Environmental Economics and Management*, 130:103124.
- Boomsma, M., Vringer, K., and van Soest, D. (2025). The impact of real-time energy consumption feedback on residential gas and electricity usage. *Journal of Environmental Economics and Management*, 132:103163.
- Brandon, A., List, J. A., Metcalfe, R. D., Price, M. K., and Rundhammer, F. (2019). Testing for crowd out in social nudges: Evidence from a natural field experiment in the market for electricity. *Proceedings of the National Academy of Sciences*, 116(12):5293–5298.
- Brent, D. A., Cook, J. H., and Olsen, S. (2015). Social comparisons, household water use, and participation in utility conservation programs: Evidence from three randomized trials. *Journal of the Association of Environmental and Resource Economists*, 2(4):597–627.
- Brülisauer, M., Goette, L., Jiang, Z., Schmitz, J., and Schubert, R. (2020). Appliance-specific feedback and social comparisons: Evidence from a field experiment on energy conservation. *Energy Policy*, 145:111742.
- Chen, C.-W., Chen, J. I., and Lin, M.-J. (2023). Performance feedback in a group contest: A field experiment on electricity conservation. *Journal of Economics & Management Strategy*, 32(3):558–582.
- Chen, V. L., Delmas, M. A., Locke, S. L., and Singh, A. (2017). Information strategies for energy conservation: A field experiment in India. *Energy Economics*, 68:215–227.
- Cozzio, C., Tokarchuk, O., and Maurer, O. (2021). Minimising plate waste at hotel breakfast buffets: an experimental approach through persuasive messages. *British Food Journal*, 123(9):3208–3227.
- Crago, C. L., Spraggon, J. M., and Hunter, E. (2020). Motivating non-ratepaying households with feedback and social nudges: A cautionary tale. *Energy Policy*, 145:111764.

- Delmas, M. A. and Lessem, N. (2014). Saving power to conserve your reputation? the effectiveness of private versus public information. *Journal of Environmental Economics and Management*, 67(3):353–370.
- Essl, A., Steffen, A., and Staehle, M. (2021). Choose to reuse! the effect of action-close reminders on pro-environmental behavior. *Journal of Environmental Economics and Management*, 110:102539.
- Fang, X., Goette, L., Rockenbach, B., Sutter, M., Tiefenbeck, V., Schoeb, S., and Staake, T. (2023). Complementarities in behavioral interventions: Evidence from a field experiment on resource conservation. *Journal of Public Economics*, 228:105028.
- Feizi, M. and Khatabiroudi, N. (2023). Social and environmental nudges and water usage: Evidence from a field experiment in Iran. *Water Resources and Economics*, 42:100223.
- Ferraro, P. J., Miranda, J. J., and Price, M. K. (2011). The persistence of treatment effects with norm-based policy instruments: evidence from a randomized environmental policy experiment. *American Economic Review*, 101(3):318–322.
- Fielding, K. S., Spinks, A., Russell, S., McCrea, R., Stewart, R., and Gardner, J. (2013). An experimental test of voluntary strategies to promote urban water demand management. *Journal of Environmental Management*, 114:343–351.
- Giaccherini, M., Gilli, M., Mancinelli, S., and Zoli, M. (2021). Nudging food waste decisions at restaurants. *European Economic Review*, 135:103722.
- Goette, L., Leong, C., and Qian, N. (2019). Motivating household water conservation: A field experiment in Singapore. *PLOS ONE*, 14(3):e0211891.
- Gosnell, G. K. (2018). Communicating resourcefully: a natural field experiment on environmental framing and cognitive dissonance in going paperless. *Ecological Economics*, 154:128–144.
- Greene, D., Zinn, A. K., Demeter, C., and Dolnicar, S. (2025). “hi, i’m terri towel. please reuse me.” can anthropomorphizing towels prompt tourists to reuse them? *Journal of Travel Research*, page 00472875251349233.
- Günther, S. A., Schöb, S., Coroamă, V. C., Tiefenbeck, V., Mattern, F., and Staake, T. (2024). All eyes on consumption feedback: A randomized controlled trial on the interplay of financial incentives, environmental appeals, and consumption feedback. In *Inducing Behavior Change with the Help of Information Systems: Cross-Domain Applications, Mechanisms, and Real-World Impact*, pages 126–163. Otto-Friedrich-Universität Bamberg.

- Günther, S. A., Staake, T., Schöb, S., and Tiefenbeck, V. (2020). The behavioral response to a corporate carbon offset program: A field experiment on adverse effects and mitigation strategies. *Global Environmental Change*, 64:102123.
- Idahosa, L. O. and Akotey, J. O. (2021). A social constructionist approach to managing HVAC energy consumption using social norms—a randomised field experiment. *Energy Policy*, 154:112293.
- Iwafune, Y., Mori, Y., Kawai, T., and Yagita, Y. (2017). Energy-saving effect of automatic home energy report utilizing home energy management system data in Japan. *Energy*, 125:382–392.
- Jaeger, C. M. and Schultz, W. P. (2017). Coupling social norms and commitments: Testing the underdetected nature of social influence. *Journal of Environmental Psychology*, 51:199–208.
- Jessoe, K., Lade, G. E., Loge, F., and Spang, E. (2021). Residential water conservation during drought: Experimental evidence from three behavioral interventions. *Journal of Environmental Economics and Management*, 110:102519.
- Joo, H. H., Lee, J., and Park, S. (2018). Every drop counts: A water conservation experiment with hotel guests. *Economic Inquiry*, 56(3):1788–1808.
- Juvan, E., Zhu, O. Y., Grün, B., and Dolnicar, S. (2025). On the importance of field studies for testing theory-driven behavioral change interventions in (sustainable) tourism. *Journal of Travel Research*, 64(6):1449–1463.
- Kallbekken, S. and Sælen, H. (2013). ‘nudging’ hotel guests to reduce food waste as a win–win environmental measure. *Economics Letters*, 119(3):325–327.
- Kažukauskas, A., Broberg, T., and Jaraitė, J. (2021). Social comparisons in real time: A field experiment of residential electricity and water use. *The Scandinavian Journal of Economics*, 123(2):558–592.
- Kim, J. H. and Kaemingk, M. (2021). Persisting effects of social norm feedback letters in reducing household electricity usage in post-Soviet Eastern Europe: A randomized controlled trial. *Journal of Economic Behavior & Organization*, 191:153–161.
- Kirchler, B., Kollmann, A., Sinea, A., and Volintiru, C. (2024). Behavioural insights into energy consumption in times of crisis. *Energy Research & Social Science*, 117:103744.

- Lee, E., Kang, M., Song, J., and Kang, M. (2020). From intention to action: Habits, feedback and optimizing energy consumption in South Korea. *Energy Research & Social Science*, 64:101430.
- Legault, L., Bird, S., Powers, S. E., Sherman, A., Schay, A., Hou, D., and Janoyan, K. (2020). Impact of a motivational intervention and interactive feedback on electricity and water consumption: A smart housing field experiment. *Environment and Behavior*, 52(6):666–692.
- Leoniak, K. J. and Cwalina, W. (2019). The role of normative prompts and norm support cues in promoting light-switching behavior: A field study. *Journal of Environmental Psychology*, 64:1–11.
- López-Rodríguez, S., Deyá-Tortella, B., and Lejarraga, T. (2025). Information is not enough to boost water savings in hotel rooms. *Humanities and Social Sciences Communications*, 12(1):1–10.
- Loschelder, D. D., Siepelmeyer, H., Fischer, D., and Rubel, J. A. (2019). Dynamic norms drive sustainable consumption: Norm-based nudging helps café customers to avoid disposable to-go-cups. *Journal of Economic Psychology*, 75:102146.
- Lurbé, S., Burkhardt, J., Goemans, C., Manning, D., and Hans, L. (2023). Further evidence on social comparison and residential water use. *Water Resources and Economics*, 41:100214.
- Mi, L., Ding, C., Yang, J., Yu, X., Cong, J., Zhu, H., and Liu, Q. (2019). Using goal and contrast feedback to motivate chinese urban families to save electricity actively—a randomized controlled field trial. *Journal of Cleaner Production*, 226:443–453.
- Mi, L., Qiao, L., Du, S., Xu, T., Gan, X., Wang, W., and Yu, X. (2020). Evaluating the effect of eight customized information strategies on urban households’ electricity saving: A field experiment in china. *Sustainable Cities and Society*, 62:102344.
- Miranda, J. J., Datta, S., and Zoratto, L. (2020). Saving water with a nudge (or two): Evidence from Costa Rica on the effectiveness and limits of low-cost behavioral interventions on water use. *The World Bank Economic Review*, 34(2):444–463.
- Morgan, M. and Chompreeda, K. (2015). The relative effect of message-based appeals to promote water conservation at a tourist resort in the gulf of Thailand. *Environmental Communication*, 9(1):20–36.

- Mundt, D., Batzke, M. C., Bläsing, T. M., Gomera Deaño, S., and Helfers, A. (2024). Effectiveness and context dependency of social norm interventions: Five field experiments on nudging pro-environmental and pro-social behavior. *Frontiers in Psychology*, 15:1392296.
- Nguyen, N. N., Goto, D., Tran, D., and Vu, H. T. (2024). Nudging households to save electricity with feedback: Experimental evidence from Vietnam. *Energy Policy*, 195:114378.
- Paunov, Y. and Grüne-Yanoff, T. (2026). Boosting vs. nudging sustainable energy consumption: a long-term comparative field test in a residential context. *Behavioural Public Policy*, 10(1):42–67.
- Pellerano, J. A., Price, M. K., Puller, S. L., and Sánchez, G. E. (2017). Do extrinsic incentives undermine social norms? evidence from a field experiment in energy conservation. *Environmental and Resource Economics*, 67(3):413–428.
- Pereira-Doel, P., Font, X., Wyles, K., and Pereira-Moliner, J. (2019). Showering smartly. a field experiment using water-saving technology to foster pro-environmental behaviour among hotel guests. *E-Review of Tourism Research*, 17(3).
- Qin, B. and Chen, H. (2022). Does the nudge effect persist? evidence from a field experiment using social comparison message in China. *Bulletin of Economic Research*, 74(3):689–703.
- Ramli, U. and Laffan, K. (2022). Double trouble: Concurrently targeting water and electricity using normative messages in the Middle East. *Energy Research & Social Science*, 88:102496.
- Salazar, G., Neves, J., Alves, V., Silva, B., Giger, J.-C., and Veríssimo, D. (2021). The effectiveness and efficiency of using normative messages to reduce waste: A real world experiment. *PLOS ONE*, 16(12):e0261734.
- Schleich, J., Klobasa, M., Gözl, S., and Brunner, M. (2013). Effects of feedback on residential electricity demand—findings from a field trial in Austria. *Energy Policy*, 61:1097–1106.
- Schultz, W. P., Estrada, M., Schmitt, J., Sokoloski, R., and Silva-Send, N. (2015). Using in-home displays to provide smart meter feedback about household electricity consumption: A randomized control trial comparing kilowatts, cost, and social norms. *Energy*, 90:351–358.
- Schultz, W. P., Khazian, A. M., and Zaleski, A. C. (2008). Using normative social influence to promote conservation among hotel guests. *Social Influence*, 3(1):4–23.

- Schäufele-Elbers, I., Schamel, G., and Perathoner, M. (2025). Nudging food waste off the plate? an explorative study investigating the generation of plate waste over time and the effectiveness of information nudges to bridge the attitude-behavior gap. *Journal of Foodservice Business Research*, 28(4):774–792.
- Tiberio, L., Kirchler, B., Massullo, C., Carrus, G., Haider, J., Kollmann, A., and Caffaro, F. (2025). Unveiling the power of social norms interventions: Investigating energy savings behavior in an italian energy cooperative. *Energy Research & Social Science*, 122:103989.
- Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., and Staake, T. (2018). Overcoming salience bias: How real-time feedback fosters resource conservation. *Management Science*, 64(3):1458–1476.
- Tiefenbeck, V., Staake, T., Roth, K., and Sachs, O. (2013). For better or for worse? empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy*, 57:160–171.
- Tonke, S. (2024). Providing procedural knowledge: A field experiment to encourage resource conservation in Namibia. *Journal of Development Economics*, 166:103202.
- Torres, M. M. J. and Carlsson, F. (2018). Direct and spillover effects of a social information campaign on residential water-savings. *Journal of Environmental Economics and Management*, 92:222–243.
- Wang, J., Wen, L., Li, Y., Liu, J., and Che, S. (2025). Uncovering the impacts of social comparative information on water saving behavior among urban residents: A field experiment in China. *Journal of Cleaner Production*, 492:144863.
- Warren, C., Becken, S., Nguyen, K., and Stewart, R. A. (2018). Transitioning to smart sustainable tourist accommodation: Service innovation results. *Journal of Cleaner Production*, 201:599–608.
- Williams, S. P., Thondhlana, G., and Kua, H. W. (2022). Behavioural interventions yield electricity savings among high-income households in Johannesburg, South Africa. *Energy & Environment*, 33(8):1508–1523.
- Wu, L. and Zhou, Y. (2025). Social norms and energy conservation in China. *Resource and Energy Economics*, 82:101491.
- Xiao, A. Z., Sparkman, G., and Constantino, S. M. (2023). Like me or near me? assessing which norm referents best promote energy conservation in the field. *Current Research in Ecological and Social Psychology*, 5:100142.

- Zhu, J., Li, J., Liu, P., Shang, Y., Ding, Z., and Alam, M. M. (2024). Effects of mixed normative feedback strategies on residents' energy conservation in shared spaces. *Energy and Buildings*, 305:113893.
- Zhu, J., Morshed Alam, M., Liu, R., Wang, J., Ding, Z., and Ekambaram, P. (2023). Evaluating the effect of normative feedback on energy conservation in a shared space. *Energy and Buildings*, 284:112862.

Table A2: Coded studies included in the meta-analysis

ID	Study	Region	Intervention	Behavior	Setting	Hedges' g	Variance	N (ctrl)	N (treat)	N (total)
1	Ahsanuzzaman et al. (2024)	Asia	Social comparison	Energy	Household	-0.220	0.0036	550	567	1117
2	Ahsanuzzaman et al. (2024)	Asia	Social comparison	Energy	Household	-0.182	0.0036	550	568	1118
3	Ahsanuzzaman et al. (2024)	Asia	Information & Education	Energy	Household	-0.377	0.0037	550	563	1113
4	Allcott (2011)	North America	Social comparison	Energy	Household	-0.035	0.0001	18694	18790	37484
5	Allcott (2011)	North America	Social comparison	Energy	Household	-0.072	0.0001	28160	14014	42174
6	Allcott (2011)	North America	Social comparison	Energy	Household	-0.048	0.0001	39249	19512	58761
7	Allcott (2011)	North America	Social comparison	Energy	Household	-0.090	0.0004	4358	7254	11612
8	Allcott (2011)	North America	Social comparison	Energy	Household	-0.052	0.0002	10290	16947	27237
9	Allcott (2011)	North America	Social comparison	Energy	Household	-0.041	0.0010	1034	23906	24940
10	Allcott (2011)	North America	Social comparison	Energy	Household	-0.048	0.0002	8028	9861	17889
11	Allcott (2011)	North America	Social comparison	Energy	Household	-0.033	0.0001	24863	24808	49671
12	Allcott (2011)	North America	Social comparison	Energy	Household	-0.042	0.0001	44336	17447	61783
13	Allcott (2011)	North America	Social comparison	Energy	Household	-0.023	0.0002	19641	5570	25211
14	Allcott (2011)	North America	Social comparison	Energy	Household	-0.021	0.0003	13997	3852	17849
15	Allcott (2011)	North America	Social comparison	Energy	Household	-0.039	0.0001	19673	19663	39336
16	Allcott (2011)	North America	Social comparison	Energy	Household	-0.083	0.0001	34905	24761	59666
17	Allcott (2011)	North America	Social comparison	Energy	Household	-0.074	0.0002	14390	9903	24293
18	Andor et al. (2020)	Europe	Social comparison	Energy	Household	-0.048	0.0003	5812	5808	11620
19	Asensio and Delmas (2015)	North America	Social comparison	Energy	Household	-0.090	0.0542	33	42	75
20	Asmare et al. (2021)	Europe	Feedback	Energy	Household	0.057	0.0040	632	419	1051
21	Bator et al. (2019)	North America	Social comparison	Energy	Household	-0.048	0.0089	242	208	450
22	Bergquist and Nilsson (2016)	Europe	Appeal/Prompt/Reminder	Energy	other	-0.168	0.0223	70	126	196
23	Bergquist and Nilsson (2016)	Europe	Appeal/Prompt/Reminder	Energy	other	-0.553	0.0223	70	138	208
24	Bergquist et al. (2019)	Europe	Appeal/Prompt/Reminder	Energy	Household	-0.092	0.0392	170	30	200
25	Bergquist et al. (2019)	Europe	Social Norm Messages	Energy	Household	0.080	0.0392	170	30	200
26	Bhanot (2017)	North America	Social comparison	Water	Household	0.046	0.0015	1308	1288	2596
27	Bhanot (2017)	North America	Social comparison	Water	Household	-0.006	0.0015	1308	1300	2608
28	Bhanot (2021)	North America	Social comparison	Water	Household	-0.050	0.0003	4586	18347	22933
29	Bhanot (2021)	North America	Social comparison	Water	Household	-0.040	0.0003	4586	18346	22932
30	Bhanot (2021)	North America	Social comparison	Water	Household	-0.035	0.0004	4586	4587	9173
31	Bohner and Schlüter (2014)	Europe	Social Norm Messages	Towel	Travel & Tourism	-0.105	0.0525	42	35	77
32	Bohner and Schlüter (2014)	Europe	Social Norm Messages	Towel	Travel & Tourism	-0.415	0.0544	42	34	76
33	Bohner and Schlüter (2014)	Europe	Appeal/Prompt/Reminder	Towel	Travel & Tourism	-1.115	0.0658	42	30	72
34	Bohner and Schlüter (2014)	Europe	Social Norm Messages	Towel	Travel & Tourism	-0.278	0.0521	42	36	78
35	Bohner and Schlüter (2014)	Europe	Social Norm Messages	Towel	Travel & Tourism	-0.633	0.0638	42	27	69

Table A2 – continued from previous page

ID	Study	Region	Intervention	Behavior	Setting	Hedges' g	Variance	N (ctrl)	N (treat)	N (total)
36	Bonan et al. (2024)	Europe	Social comparison	Water	Household	-0.030	0.0000	54490	54490	108980
37	Bonan et al. (2025)	Europe	Social comparison	Waste	Household	-0.050	0.0001	15585	31348	46933
38	Boomsma et al. (2025)	Europe	Feedback	Energy	Household	-0.056	0.0050	384	418	802
39	Brandon et al. (2019)	North America	Social comparison	Energy	Household	-0.035	0.0003	4641	11744	16385
40	Brent et al. (2015)	North America	Social comparison	Water	Household	-0.074	0.0021	897	992	1889
41	Brent et al. (2015)	North America	Social comparison	Water	Household	-0.062	0.0013	1547	1545	3092
42	Brent et al. (2015)	North America	Social comparison	Water	Household	-0.025	0.0017	1200	1180	2380
43	Brülisauer et al. (2020)	Asia	Social comparison	Energy	Household	-0.153	0.0299	37	352	389
44	Chen et al. (2017)	Asia	Social comparison	Energy	Household	-0.115	0.3100	7	6	13
45	Chen et al. (2023)	Asia	Social comparison	Energy	Household	-0.275	0.0309	44	126	170
46	Chen et al. (2023)	Asia	Social comparison	Energy	Household	-0.217	0.0308	44	126	170
47	Chen et al. (2023)	Asia	Social comparison	Energy	Household	-0.340	0.0310	44	126	170
48	Cozzio et al. (2021)	Europe	Appeal/Prompt/Reminder	Food waste	Travel & Tourism	-0.119	0.0037	546	532	1078
49	Crago et al. (2020)	North America	Feedback	Energy	Household	0.158	0.1003	20	20	40
50	Crago et al. (2020)	North America	Social comparison	Energy	Household	0.200	0.0959	20	22	42
51	Delmas and Lessem (2014)	North America	Social comparison	Energy	Household	-0.096	0.0890	23	22	45
52	Delmas and Lessem (2014)	North America	Social comparison	Energy	Household	-0.330	0.0923	23	21	44
53	Essl et al. (2021)	Europe	Appeal/Prompt/Reminder	Waste	Household	-0.621	0.0220	98	93	191
54	Essl et al. (2021)	Europe	Appeal/Prompt/Reminder	Waste	Household	-0.547	0.0214	98	96	194
55	Fang et al. (2023)	Europe	Feedback	Energy	Household	-0.207	0.0234	82	90	172
56	Feizi and Khatabiroudi (2023)	Asia	Social comparison	Water	Household	-0.332	0.0172	118	118	236
57	Feizi and Khatabiroudi (2023)	Asia	Information & Education	Water	Household	0.033	0.0150	133	133	266
58	Ferraro et al. (2011)	North America	Information & Education	Water	Household	-0.020	0.0001	71779	11699	83478
59	Ferraro et al. (2011)	North America	Appeal/Prompt/Reminder	Water	Household	-0.083	0.0001	71779	11695	83474
60	Ferraro et al. (2011)	North America	Social comparison	Water	Household	-0.145	0.0001	71779	11699	83478
61	Fielding et al. (2013)	Oceania	Information & Education	Water	Household	-0.181	0.0664	54	21	75
62	Fielding et al. (2013)	Oceania	Social Norm Messages	Water	Household	-0.119	0.0375	54	53	107
63	Fielding et al. (2013)	Oceania	Feedback	Water	Household	-0.079	0.0364	54	56	110
64	Giaccherini et al. (2021)	Europe	Decision structural nudges	Food waste	Travel & Tourism	-0.372	0.0364	69	47	116
65	Giaccherini et al. (2021)	Europe	Social Norm Messages	Food waste	Travel & Tourism	-0.835	0.0320	69	67	136
66	Goette et al. (2019)	Asia	Appeal/Prompt/Reminder	Water	Household	-0.049	0.0112	184	175	359
67	Goette et al. (2019)	Asia	Social comparison	Water	Household	-0.043	0.0112	184	173	357
68	Gosnell (2018)	Europe	Information & Education	Waste	Household	-0.014	0.0002	9642	9668	19310
69	Gosnell (2018)	Europe	Social Norm Messages	Waste	Household	-0.019	0.0002	9642	9660	19302
70	Greene et al. (2025)	Oceania	Appeal/Prompt/Reminder	Towel	Travel & Tourism	-0.111	0.0334	61	59	120
71	Greene et al. (2025)	Oceania	Appeal/Prompt/Reminder	Towel	Travel & Tourism	-0.086	0.0334	62	58	120

Table A2 – continued from previous page

ID	Study	Region	Intervention	Behavior	Setting	Hedges' g	Variance	N (ctrl)	N (treat)	N (total)
72	Günther et al. (2020)	Europe	Feedback	Energy	Household	-0.277	0.0257	79	78	157
73	Günther et al. (2020)	Europe	Feedback	Energy	Household	-0.235	0.0255	79	79	158
74	Günther et al. (2024)	Europe	Feedback	Water	Travel & Tourism	-0.053	0.0009	2223	2223	4446
75	Idahosa and Akotey (2021)	Africa	Appeal/Prompt/Reminder	Energy	Travel & Tourism	-0.063	0.0057	420	305	725
76	Idahosa and Akotey (2021)	Africa	Social Norm Messages	Energy	Travel & Tourism	-0.031	0.0055	420	322	742
77	Iwafune et al. (2017)	Asia	Social comparison	Energy	Household	-0.012	0.0035	637	524	1161
78	Jaeger and Schultz (2017)	North America	Social Norm Messages	Water	Household	-0.058	0.0014	1424	1375	2799
79	Torres and Carlsson (2018)	Latin America	Social comparison	Water	Household	-0.086	0.0031	655	656	1311
80	Jessoe et al. (2021)	North America	Social comparison	Water	Household	-0.058	0.0006	2782	4559	7341
81	Joo et al. (2018)	Asia	Appeal/Prompt/Reminder	Water	Travel & Tourism	-0.315	0.0345	44	88	132
82	Joo et al. (2018)	Asia	Goal Setting	Water	Travel & Tourism	-0.903	0.0387	44	79	123
83	Joo et al. (2018)	Asia	Social Norm Messages	Water	Travel & Tourism	-0.488	0.0341	44	95	139
84	Juvan et al. (2025)	Europe	Appeal/Prompt/Reminder	Food waste	Travel & Tourism	0.055	0.0039	482	537	1019
85	Juvan et al. (2025)	Europe	Appeal/Prompt/Reminder	Food waste	Travel & Tourism	-0.164	0.0042	482	483	965
86	Kallbekken and Sælen (2013)	Europe	Decision structural nudges	Food waste	Travel & Tourism	-0.235	0.1698	38	7	45
87	Kallbekken and Sælen (2013)	Europe	Appeal/Prompt/Reminder	Food waste	Travel & Tourism	-0.318	0.1703	38	7	45
88	Kažukauskas et al. (2021)	Europe	Social comparison	Energy	Household	-0.039	0.0132	315	100	415
89	Kažukauskas et al. (2021)	Europe	Social comparison	Water	Household	0.054	0.0123	315	110	425
90	Kim and Kaemingk (2021)	Europe	Social comparison	Energy	Household	-0.018	0.0001	107748	9942	117690
91	Kirchler et al. (2024)	Europe	Information & Education	Energy	Household	-0.009	0.0002	6911	14992	21903
92	Lee et al. (2020)	Asia	Feedback	Energy	Household	-0.134	0.0138	570	83	653
93	Legault et al. (2020)	North America	Goal Setting	Energy	Household	-0.690	0.0672	36	28	64
94	Leoniak and Cwalina (2019)	Europe	Appeal/Prompt/Reminder	Energy	other	-0.491	0.0116	238	142	380
95	Leoniak and Cwalina (2019)	Europe	Social Norm Messages	Energy	other	-0.766	0.0112	238	160	398
96	Leoniak and Cwalina (2019)	Europe	Social Norm Messages	Energy	other	-1.080	0.0115	238	170	408
97	Loschelder et al. (2019)	Europe	Social Norm Messages	Waste	other	-0.108	0.3504	10	4	14
98	Lurbé et al. (2023)	North America	Social comparison	Water	Household	-0.068	0.0004	6441	3713	10154
99	López-Rodríguez et al. (2025)	Europe	Information & Education	Water	Travel & Tourism	0.456	0.0228	90	90	180
100	López-Rodríguez et al. (2025)	Europe	Information & Education	Water	Travel & Tourism	0.089	0.0238	84	84	168
101	Mi et al. (2019)	Asia	Goal Setting	Energy	Household	0.026	0.0597	34	33	67
102	Mi et al. (2019)	Asia	Goal Setting	Energy	Household	-0.231	0.0592	34	34	68
103	Mi et al. (2019)	Asia	Goal Setting	Energy	Household	-0.271	0.0594	34	34	68
104	Mi et al. (2019)	Asia	Goal Setting	Energy	Household	-0.307	0.0614	34	32	66
105	Mi et al. (2020)	Asia	Information & Education	Energy	Household	-0.202	0.1340	15	15	30
106	Mi et al. (2020)	Asia	Social Norm Messages	Energy	Household	-0.361	0.1355	15	15	30
107	Mi et al. (2020)	Asia	Feedback	Energy	Household	-0.511	0.1377	15	15	30

Table A2 – continued from previous page

ID	Study	Region	Intervention	Behavior	Setting	Hedges' g	Variance	N (ctrl)	N (treat)	N (total)
108	Miranda et al. (2020)	Latin America	Social comparison	Water	Household	-0.046	0.0015	1310	1299	2609
109	Miranda et al. (2020)	Latin America	Social comparison	Water	Household	-0.042	0.0015	1310	1283	2593
110	Miranda et al. (2020)	Latin America	Goal Setting	Water	Household	-0.089	0.0015	1310	1278	2588
111	Morgan and Chompreeda (2015)	Asia	Social Norm Messages	Towel	Travel & Tourism	-0.792	0.0396	55	54	109
112	Morgan and Chompreeda (2015)	Asia	Appeal/Prompt/Reminder	Towel	Travel & Tourism	-1.515	0.0456	55	58	113
113	Morgan and Chompreeda (2015)	Asia	Social Norm Messages	Towel	Travel & Tourism	-0.938	0.0393	55	58	113
114	Mundt et al. (2024)	Europe	Appeal/Prompt/Reminder	Waste	other	-0.478	0.0230	90	89	179
115	Mundt et al. (2024)	Europe	Appeal/Prompt/Reminder	Waste	other	-0.326	0.0222	95	88	183
116	Mundt et al. (2024)	Europe	Appeal/Prompt/Reminder	Waste	other	-0.464	0.0245	86	82	168
117	Mundt et al. (2024)	Europe	Appeal/Prompt/Reminder	Waste	other	-0.398	0.0243	86	82	168
118	Mundt et al. (2024)	Europe	Appeal/Prompt/Reminder	Waste	other	-0.023	0.0194	102	104	206
119	Nguyen et al. (2024)	Asia	Feedback	Energy	Household	-0.105	0.0078	254	258	512
120	Nguyen et al. (2024)	Asia	Social comparison	Energy	Household	-0.148	0.0078	254	258	512
121	Nguyen et al. (2024)	Asia	Goal Setting	Energy	Household	-0.092	0.0078	254	258	512
122	Paunov and Grüne-Yanoff (2026)	Europe	Goal Setting	Energy	Household	0.537	0.0964	21	22	43
123	Paunov and Grüne-Yanoff (2026)	Europe	Feedback	Energy	Household	-0.377	0.0947	21	22	43
124	Pellerano et al. (2017)	Latin America	Social comparison	Energy	Household	-0.021	0.0002	9425	9359	18784
125	Pereira-Doel et al. (2019)	Europe	Feedback	Water	Travel & Tourism	-0.188	0.0021	914	1048	1962
126	Qin and Chen (2022)	Asia	Social comparison	Energy	Household	0.041	0.0104	193	192	385
127	Qin and Chen (2022)	Asia	Social comparison	Energy	Household	-0.046	0.0105	193	189	382
128	Ramli and Laffan (2022)	Asia	Social comparison	Water	Household	0.001	0.0000	52986	53010	105996
129	Ramli and Laffan (2022)	Asia	Social comparison	Water	Household	-0.002	0.0000	52986	51256	104242
130	Ramli and Laffan (2022)	Asia	Social comparison	Water	Household	0.000	0.0000	52986	51557	104543
131	Ramli and Laffan (2022)	Asia	Social comparison	Energy	Household	-0.003	0.0000	52986	53010	105996
132	Ramli and Laffan (2022)	Asia	Social comparison	Energy	Household	-0.003	0.0000	52986	51256	104242
133	Ramli and Laffan (2022)	Asia	Social comparison	Energy	Household	-0.002	0.0000	52986	51557	104543
134	Salazar et al. (2021)	Europe	Social Norm Messages	Waste	other	-0.070	0.2001	30	6	36
135	Salazar et al. (2021)	Europe	Social Norm Messages	Waste	other	-0.251	0.2009	30	6	36
136	Salazar et al. (2021)	Europe	Appeal/Prompt/Reminder	Waste	other	-0.357	0.2018	30	6	36
137	Salazar et al. (2021)	Europe	Appeal/Prompt/Reminder	Waste	other	-0.217	0.2007	30	6	36
138	Salazar et al. (2021)	Europe	Appeal/Prompt/Reminder	Waste	other	-0.190	0.2005	30	6	36
139	Salazar et al. (2021)	Europe	Appeal/Prompt/Reminder	Waste	other	-0.254	0.2343	30	5	35
140	Salazar et al. (2021)	Europe	Appeal/Prompt/Reminder	Waste	other	-0.276	0.0792	26	25	51
141	Salazar et al. (2021)	Europe	Appeal/Prompt/Reminder	Waste	other	-0.080	0.0770	26	26	52
142	Schäufele-Elbers et al. (2025)	Europe	Appeal/Prompt/Reminder	Food waste	Travel & Tourism	-1.976	0.1011	28	31	59
143	Schäufele-Elbers et al. (2025)	Europe	Appeal/Prompt/Reminder	Food waste	Travel & Tourism	-1.652	0.0995	28	26	54

Table A2 – continued from previous page

ID	Study	Region	Intervention	Behavior	Setting	Hedges' g	Variance	N (ctrl)	N (treat)	N (total)
144	Schleich et al. (2013)	Europe	Feedback	Energy	Household	-0.106	0.0026	750	775	1525
145	Schultz et al. (2008)	North America	Social Norm Messages	Towel	Travel & Tourism	-0.339	0.0090	135	655	790
146	Schultz et al. (2008)	North America	Social Norm Messages	Towel	Travel & Tourism	-0.168	0.0120	187	152	339
147	Schultz et al. (2008)	North America	Social Norm Messages	Towel	Travel & Tourism	-0.327	0.0079	187	403	590
148	Schultz et al. (2015)	North America	Feedback	Energy	Household	-0.032	0.0175	122	108	230
149	Schultz et al. (2015)	North America	Social comparison	Energy	Household	-0.164	0.0198	122	87	209
150	Tiberio et al. (2025)	Europe	Social Norm Messages	Energy	Household	-0.110	0.0135	222	111	333
151	Tiberio et al. (2025)	Europe	Social Norm Messages	Energy	Household	0.022	0.0137	222	109	331
152	Tiefenbeck et al. (2013)	North America	Social comparison	Water	Household	-0.337	0.0263	77	77	154
153	Tiefenbeck et al. (2018)	Europe	Feedback	Energy	Household	-0.364	0.0096	209	215	424
154	Tiefenbeck et al. (2018)	Europe	Feedback	Energy	Household	-0.338	0.0096	209	212	421
155	Tonke (2024)	Africa	Appeal/Prompt/Reminder	Water	Household	-0.022	0.0005	3667	3749	7416
156	Tonke (2024)	Africa	Appeal/Prompt/Reminder	Water	Household	-0.004	0.0005	3667	3709	7376
157	Tonke (2024)	Africa	Appeal/Prompt/Reminder	Water	Household	-0.000	0.0005	3667	3751	7418
158	Wang et al. (2025)	Asia	Social comparison	Water	Household	-0.077	0.0046	466	406	872
159	Wang et al. (2025)	Asia	Social comparison	Water	Household	-0.155	0.0047	466	394	860
160	Wang et al. (2025)	Asia	Social comparison	Water	Household	-0.097	0.0047	466	393	859
161	Wang et al. (2025)	Asia	Feedback	Water	Household	-0.048	0.0045	466	427	893
162	Warren et al. (2018)	Oceania	Feedback	Energy	Travel & Tourism	-0.524	0.0122	207	144	351
163	Williams et al. (2022)	Africa	Information & Education	Energy	Household	-0.290	0.0444	46	45	91
164	Wu and Zhou (2025)	Asia	Social comparison	Energy	Household	-0.103	0.0052	448	338	786
165	Xiao et al. (2023)	Asia	Social Norm Messages	Energy	Household	-0.874	0.0390	50	64	114
166	Xiao et al. (2023)	Asia	Social Norm Messages	Energy	Household	0.077	0.0382	50	55	105
167	Zhu et al. (2023)	Asia	Social comparison	Energy	Household	-0.156	0.0334	60	60	120
168	Zhu et al. (2024)	Asia	Social comparison	Energy	Household	-0.194	0.0837	24	24	48
169	Zhu et al. (2024)	Asia	Social comparison	Energy	Household	-0.099	0.0834	24	24	48