



HAL
open science

Forced to face the truth: A meta-analysis on the effectiveness of moral reminders

Constance Frohly, Roberto Galbiati, Emeric Henry, Nicolas Jacquemet

► **To cite this version:**

Constance Frohly, Roberto Galbiati, Emeric Henry, Nicolas Jacquemet. Forced to face the truth: A meta-analysis on the effectiveness of moral reminders. 2026. ⟨halshs-05456784⟩

HAL Id: halshs-05456784

<https://shs.hal.science/halshs-05456784v1>

Preprint submitted on 13 Jan 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Forced to face the truth: A meta-analysis on the effectiveness of moral reminders*

Constance Frohly[†] Roberto Galbiati[‡] Emeric Henry[§] Nicolas Jacquemet[¶]

January 2026

Abstract

Moral reminders, also referred to as moral appeals or moral nudges, are widely used by governments, companies, and NGOs to promote pro-social behavior. These appeals function by either increasing the salience of moral concerns or the cost of diverting attention away from relevant information on payoffs or social norms. Drawing on over 400 studies across psychology, sociology, management and economics, we present a meta-analysis of their effects. Our findings reveal that, on average, moral reminders are effective, with an effect size (Hedge’s g) of 0.24 in a random-effects model, but with significant backfiring occurring in 12% of studies. We identify sources of heterogeneity based on disciplinary focus and design choices. Crucially, we introduce a taxonomy of moral reminders: we distinguish those that provide information on consequences, those that highlight descriptive or injunctive norms, and those that prime moral awareness. Our analysis shows that all of these instruments are effective, particularly those providing information on consequences, whereas information on injunctive norms is more likely to backfire.

Keywords: Meta-analysis; Behavioral ethics; Moral reminders; Pro-social behavior.

*CEPR DP 21025. Authors are listed in alphabetical order, and all contributed equally. We thank Abel Baret, Alexandre Grit, Anna Guellaen-Mignard, Prateek Jain, Emir Tarik Dakin and Alexandre Vessereau for their excellent research assistance, as well as Nick Zubanov, Urs Fishbacher, Christina Felfe and participants to the 2024 Workshop “*Behavioural Science Meets Metascience*” (Oxford), the 2024 workshop “*Visualizations and behavioral strategies for sustainability*” (Sienna), the UniCa C-BECSS 2025 Workshop (Cagliari), the TWI 2025 Micro Workshop (Konstanz), the 2025 *Future of behavioral Ethics* workshop (Berlin), the 2025 *Behavioral Insight for Policy* workshop (Chisinau), the 2025 CNEE workshop (Copenhagen) and seminars at U. Lille and U. Genoa, for insightful feedback. Financial support from the French National Research Agency (ANR-17-EURE-0001 and ANR-21-CE26-0022-02) is gratefully acknowledged.

[†]Sciences Po. Department of Economics, 28 rue des Saints-Pères 75007 Paris, France. constance.frohly@sciencespo.fr

[‡]CNRS, Sciences Économiques SciencesPo and CEPR. Department of Economics, 28 rue des Saints-Pères 75007 Paris, France. roberto.galbiati@sciencepo.fr

[§]Sciences Po and CEPR. Department of Economics, 28 rue des Saints-Pères 75007 Paris, France. emeric.henry@sciencepo.fr

[¶]Paris School of Economics and University Paris 1 Panthéon-Sorbonne. Centre d’Economie de la Sorbonne, 106 Bd de l’hôpital, 75013 Paris, France. nicolas.jacquemet@univ-paris1.fr

1 Introduction

To encourage pro-social actions, governments, companies and NGOs often resort to moral reminders, interventions that increase the salience of moral concerns or make it costly to ignore relevant information, as an alternative to material incentives. These reminders, variously labeled moral appeals, moral nudges, or moral suasion¹, have attracted substantial attention in psychology, sociology, and economics. Yet, probably due to the large variety of these interventions, evidence remains fragmented regarding their average effectiveness, the possibility of backfiring, and the conditions under which they succeed. To provide robust evidence on the impact of these interventions, our meta-analysis reviews a large set of studies (415 treatment effects) investigating the consequences of moral reminders on pro-sociality. Our study thus covers a large range of methods, applications and outcome measures.

Theory and evidence on individuals' motivation to act pro-socially provide mixed predictions about how effective these interventions may be. It is now well established that individuals derive utility not only from material payoffs but also from acting consistently with internalized values and from conforming to social norms (Rabin, 1998; Bicchieri, 2006). If preferences already incorporate non-material motives, the need for reminders might appear limited. However, to preserve a positive self-image when not acting pro-socially, individuals may distort or selectively ignore information on moral payoffs or social norms (Bénabou and Henkel, 2025, for a review of the literature), or neglect non-material considerations at the moment of choice. The literature has documented several ways in which individuals preserve their self image when not acting pro-socially. This can involve avoiding information about the choice situation to create “moral wiggle room” (Dana, Weber, and Kuang, 2007; Vu, Soraperra, Leib, van der Weele, and Shalvi, 2023).² Another strand of literature has shown that the active search for excuses serves as tools to maintain a positive self-image when individuals choose not to engage in prosocial behavior. For instance, Exley (2020) shows, through a series of experiments, that information on charity performance metrics can be actively used as an excuse not to give. Regardless of the method used to preserve self-image, exposing individuals to information contained in moral reminders may increase the costs of avoidance strategies. Moral reminders therefore work by forcibly providing information that become costly to ignore, or alternatively by raising the salience of moral concerns.

¹Other terms include moral framing, moral prompts, and moral-identity activation.

²In the canonical experiment of Dana, Weber, and Kuang (2007), dictators are offered two options (A and B) that describe an allocation of money between themselves and another subject. A gives marginally more money than B to the dictator, and what it implies for the other subject is uncertain. When offered the possibility to resolve the uncertainty, only half of the participants choose to do it, the others *a priori* attempting to preserve their “moral wiggle room”.

Table 1: Classification of moral reminders

Type of reminder	Description	Example
<i>Priming</i>	exposure to a task (word decoding, text reading) aimed at making salient the ethical drivers of a decision	“Participants had to write words from a list of positive traits words such as caring, generous, fair, and kind and then had to write a story using those words” (Sachdeva, Iliev, and Medin, 2009, dictator game).
<i>Consequences</i>	provide information about the consequences of the available choices	“Your tax payment contributes to the funding of publicly financed services in education, health and other important sectors of society” (Bott, Cappelen, Sørensen, and Tungodden, 2020).
<i>Descriptive norms</i>	provide information on the descriptive norm (what others do)	“The vast majority of participants were honest about the reported number in the dice task” (Dimant, Van Kleef, and Shalvi, 2020, dice-rolling task).
<i>Injunctive norms</i>	provide information on the injunctive norm (what others think should be done)	“The vast majority of participants said that one should be honest about reported number in the dice task.” (<i>ibid.</i>).

An information based taxonomy of moral reminders. This perspective motivates the classification of moral reminders we propose. We distinguish four classes of moral reminders, summarized in Table 1, where we define classes and provide examples drawn from studies included in the meta-analysis. The first category is **priming**. This type of intervention does not supply new information, but makes moral dilemmas more salient. Salience is activated, for example, by memory or imagination tasks (Sachdeva, Iliev, and Medin, 2009) or by exposure to broad moral principles (e.g., reading the Ten Commandments; Dal Bó and Dal Bó, 2014).

The remaining three classes are information-based. First we consider reminders using information on **consequences** of actions, specifically, for pro-social actions, the consequences experienced by others. This includes for instance information on how taxes are used to fund public goods (Bott, Cappelen, Sørensen, and Tungodden, 2020) or the amount of freshwater and land required for meat consumption (Palomo-Vélez, Tybur, and Van Vugt, 2018). The two other classes of moral reminders involve providing information on social norms, either on the **descriptive norm** or on the **injunctive norm**. The descriptive norm states what others typically do in comparable situations while the injunctive norm describes what people typically think should be done. For instance, both descriptive and injunctive norms are provided regarding levels of truth-telling in Dimant, Van Kleef, and Shalvi, 2020, and regarding levels of donations in Raihani and McAuliffe, 2014.

The present meta-analysis. Many studies suggest that moral reminders may have limited effects since individuals actively manage information and moral considerations in their decision-making process. Indeed, a substantial literature shows that individuals process information in a strategic and self-serving manner, a phenomenon known as motivated reasoning (Kunda, 1990; Epley and Gilovich, 2016).³ Motivated reasoning differs from the motivated search for information mentioned above, which refers to the tendency to selectively seek information aligned with preferred conclusions (Dana, Weber, and Kuang, 2007; Vu, Soraperra, Leib, van der Weele, and Shalvi, 2023). It operates at the stage of interpretation rather than acquisition of information. These results suggest that moral reminders that merely present information, about consequences or about norms, may exert limited influence on behavior, since the information will be strategically processed. Related work on motivated memory shows that individuals recall successes more readily than failures, and norm adherence more readily than norm violations (Amelio and Zimmermann, 2023). These results cast doubt on priming-based moral reminders, especially those that rely on memory activation: such interventions may raise the cognitive cost of recalling inconvenient facts but need not fundamentally shift behavior.

Yet, our meta-analysis shows that reminders do work: they have a significant positive effect, with a Standardized Mean Difference (SMD) of 0.24 — i.e., individuals’ observed pro-sociality increases by nearly a fourth of a standard deviation following a moral reminder — indicating that their practical use is often justified. However, the effect is not uniform in magnitude across studies, nor is it unequivocally positive: a non-negligible share of interventions backfire. This variability is reflected in the rather large 95% prediction interval $[-0.975; 1.457]$. From a policy perspective, this result is essential for designing effective interventions.

Importantly, we show that the average effectiveness of moral reminders is consistently observed across studies using different methodologies. For example, we find a positive SMD for field experiments, although it is significantly lower than that observed in laboratory settings. We also assess other methodological moderators, namely whether decisions are incentivized and whether reminders target intentions, behavior, or judgments. We find that judgments are the easiest to shift, behavior moderately so, while intentions are the hardest to influence, presumably because baseline intentions tend to be more prosocial. Moral reminders therefore appear to narrow the intention-behavior gap.

We examine additional factors that researchers and policymakers can manipulate to enhance the effectiveness of moral reminders. These interventions remain effective regardless of the source of the reminder (e.g., experimenters, external authorities, religious leaders). Although reminders work across domains, they are particularly effective in changing environmental and social behaviors and far less effective in improving tax compliance, where they are more prone to backfire. We also find that the presence of observers substantially amplifies the effect of moral reminders, raising

³See Bénabou and Henkel (2025) for a review.

the average impact from 0.19 to 0.57 standard deviations.

Finally, we evaluate the relative influence of the four classes of moral reminders introduced above. We find that reminders providing information about consequences perform better and are less prone to backfiring than norm-based appeals. However, substantial heterogeneity and notable outliers within classes suggest a high level of context-dependency that should be accounted for in practical implementations.

Our meta-analysis joins a growing literature assessing the policy relevance of behaviorally informed micro-interventions. This broad research agenda highlights key questions about the scalability of nudges (DellaVigna and Linos, 2022), their complementarity with traditional price incentives (Chetty, Saez, and Sandor, 2014), and their relative cost-effectiveness compared to standard policy instruments (Benartzi, Beshears, Milkman, Sunstein, Thaler, Shankar, Tucker-Ray, Congdon, and Galing, 2017). Part of the literature focuses on specific domains, such as behavioral climate interventions (Vlasceanu *et al.*, 2024), food waste (Zhang, Huang, Zhu, and Zhao, 2023), health behavior (Papakonstantinou, Flecke, Edmunds, Cross, Tran, and Gold, 2025) or tax-compliance (Antinyan and Asatryan, 2024). Other studies impose constraints on the implementation methods, such as commitment devices (Zickfeld, Karg, Engen, Gonzalez, Michael, and Mitkidis, 2024), observability manipulations (Bradley, Lawrence, and Ferguson, 2018), and even meditation-based interventions (Kreplin, Farias, and Brazil, 2018). In our meta-analysis, we chose not to restrict the domain of application or the methodological approach. We however put more restrictions on the targeted behavior and on the class of interventions compared to for instance Mertens, Herberz, Hahnel, and Brosch (2022); Szaszi, Palinkas, Palfi, Szollosi, and Aczel (2018); Hummel and Maedche (2019) who focus on nudges at large. Specifically, we focus on moral reminders targeted directly at promoting pro-social actions. A contribution of the paper is to propose a taxonomy of such moral reminders.⁴

2 Methods

In contrast with most meta-analytic studies, we focus on a class of interventions for which no standard keyword can be easily associated. We thus defined the list of keywords based on key papers in the moral reminders literature (see the online Supplementary Information (SI), Section A.1, Table A). Our systematic search based on these keywords targeted academic articles, published before 2020, providing an experimental measure of the effect of moral reminders on pro-social outcomes — hence excluding studies focusing on the effect of moral reminders on outcomes that

⁴The class of reminders we call *Consequences* resembles what Mertens, Herberz, Hahnel, and Brosch (2022) calls *Decision information* (increase the availability, comprehensibility, and/or personal relevance of information). In our study however, the information is on consequences on others rather than on oneself, which can explain why they appear to work better than in Mertens, Herberz, Hahnel, and Brosch (2022) since it leaves a larger scope for motivated reasoning in the absence of reminders.

do not rely on pro-sociality (e.g. sovereign debt default risk). We excluded interventions based on commitment (already reviewed in Zickfeld, Karg, Engen, Gonzalez, Michael, and Mitkidis, 2024) or explicitly designed to increase selfishness rather than pro-sociality. We also excluded social comparisons, i.e., interventions that explicitly position individual’s action within a group, as opposed to descriptive norms that focus on group information. We further complemented the literature search by an open call for papers to the community (see the PRISMA chart in the SI, Section A.1, Figure A, for detailed information about inclusion decisions; and the SI, Section E, for the full list of references).

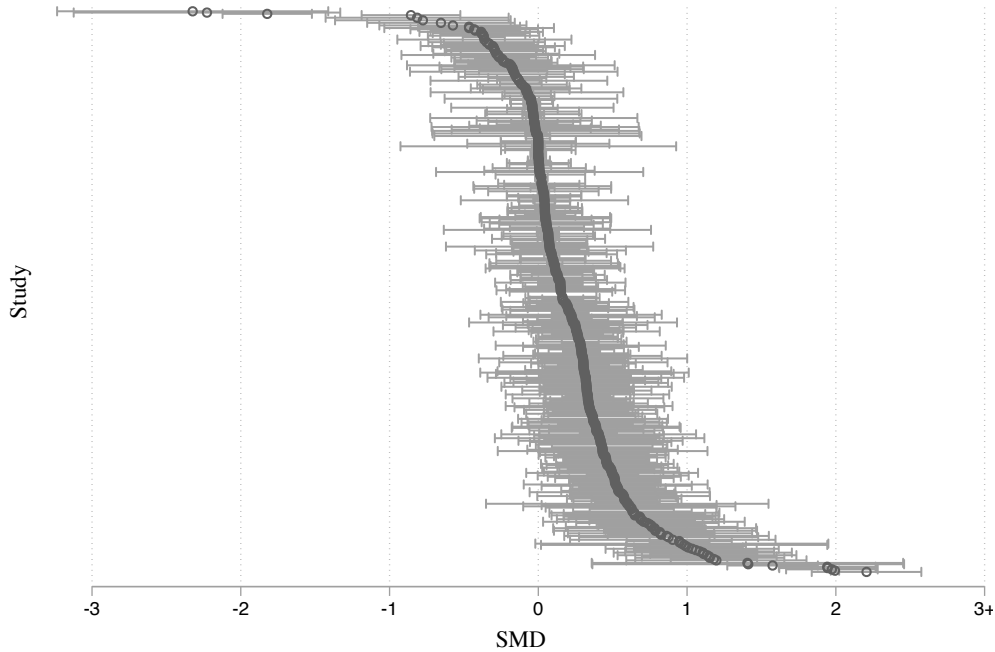
For each included paper, we keep track of each relevant study (i.e., an experiment within a paper measuring the effect of a given intervention) and define for each of them as many effect sizes (k) as the unique combinations of (i) an outcome measure, (ii) a moral reminder intervention and (iii) a control group (see the SI, Section A.3, for the list of variables and their definitions). Whenever available, we extracted the relevant summary statistics from the published version of the paper (i.e., sample sizes, outcome mean and standard deviation in the treated and control groups as well as treatment effects). We also collected the experiment datasets online when available and individually contacted all authors of included studies otherwise. We used this dataset to retrieve the missing information from the articles, which gives us access to the full meta-analytic information of 241 studies. Last, we managed to include 174 additional studies by relying on assumptions which allow to infer missing values from the partial information available in the paper (e.g., assuming the sample is equally split between groups when only the total sample size is available; see Appendix A, Section A.2).

For each study, we also recorded a large set of variables aimed at describing the characteristics of the moral reminder, the target measure of prosociality and its field of application as well as the characteristics of the experimental implementation (see the SI, Section A.4, for the complete list and formal definitions). These variables were first coded based on an extensive definition of each category. The resulting classification was then double checked using an AI algorithm (see the SI, Section A.5), which validated 76% of the initial classification (disagreements were re-investigated by the research team). Last, we also classified studies into disciplines based on the Scimago classification of the journal in which the paper was published (see the Appendix, Table D).

The resulting dataset covers 88 articles from which 122 studies are included, resulting in $K = 415$ observations of effect sizes (over $N \approx 1.03$ million observations). Each treatment effect is measured with Hedge’s g in order to correct for small sample bias, as 24 effect sizes in the meta-analysis are estimated on samples where $N \leq 30$. The Hedge’s g requires measures of sample sizes, standard deviations and mean outcomes in the treated and control groups (see, e.g., Schwarzer, Carpenter, and Rucker, 2015):

$$\hat{\theta}_k = \left(1 - \frac{3}{4N_k}\right) \frac{\hat{y}_{kT} - \hat{y}_{kC}}{\sqrt{[(N_{kT} - 1)\hat{\sigma}_{kT}^2 + (N_{kC} - 1)\hat{\sigma}_{kC}^2] / (N_k - 2)}} \quad (1)$$

Figure 1: Overall Forest Plot



Note. Forest plot of effect sizes ($K = 415$) included in the meta-analysis with their corresponding 95% confidence intervals. Extracted Hedge’s g values ranged from -9.17 to 10.40 standard deviations. $\widehat{\text{SMD}} = 0.24$, 95% CI: $[0.18, 0.30]$, $p < .001$.

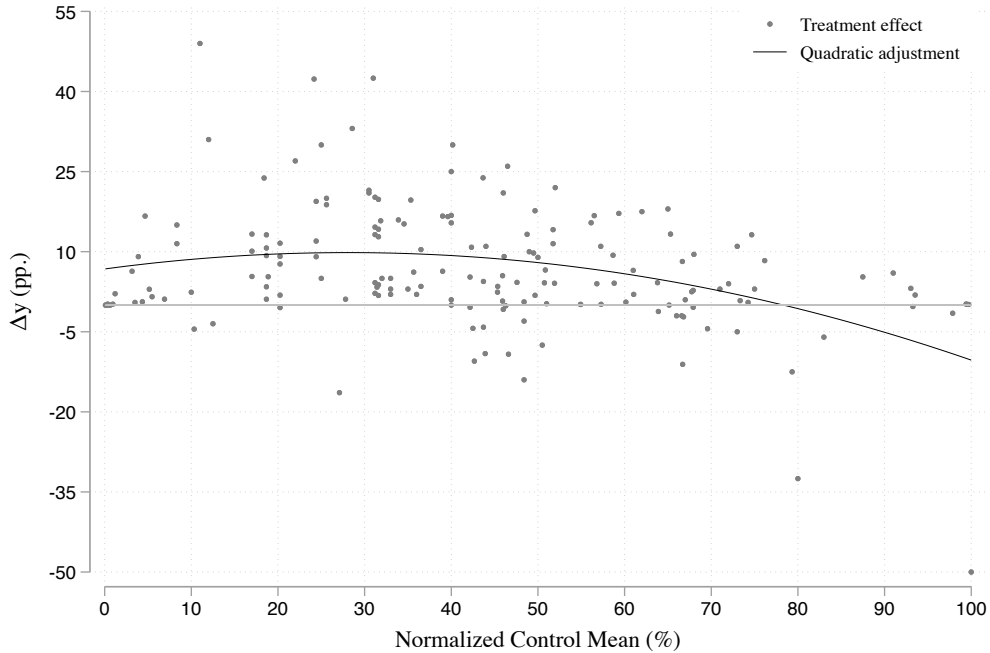
Finally, in order to estimate the average treatment effect in our dataset, we assume a random effect model to account for the heterogeneity across studies. Indeed, the true treatment effect underlying each study may differ across studies (see the SI, Section B.1 for more details).

3 Results

3.1 Moral reminders increase pro-social behavior

Our meta-analysis shows that, on average, moral reminders significantly increase pro-social behavior. Figure 1 presents the empirical distribution of effect sizes together with the best-fitting normal adjustment. The average effect size is $\widehat{\text{SMD}} = 0.24$, 95% CI: $[0.18, 0.30]$, $p < .001$. Compared to meta-studies of similar scope, this effect is of comparable magnitude, albeit on the lower end (e.g., 0.43 in Mertens, Herberz, Hahnel, and Brosch, 2022, analyzing the efficacy of general nudges, and 0.24 in Antinyan and Asatryan, 2024, focusing on tax-compliance nudges). The effect size is consistent across several robustness checks. Figures D and E (see the the SI, Section C.2) report additional leave-one-out analyses at the individual effect size level and the study level. These exercises confirm the effect of moral reminders to be robust to the exclusion of any one effect

Figure 2: Treatment effect as a function of take-up in the control group



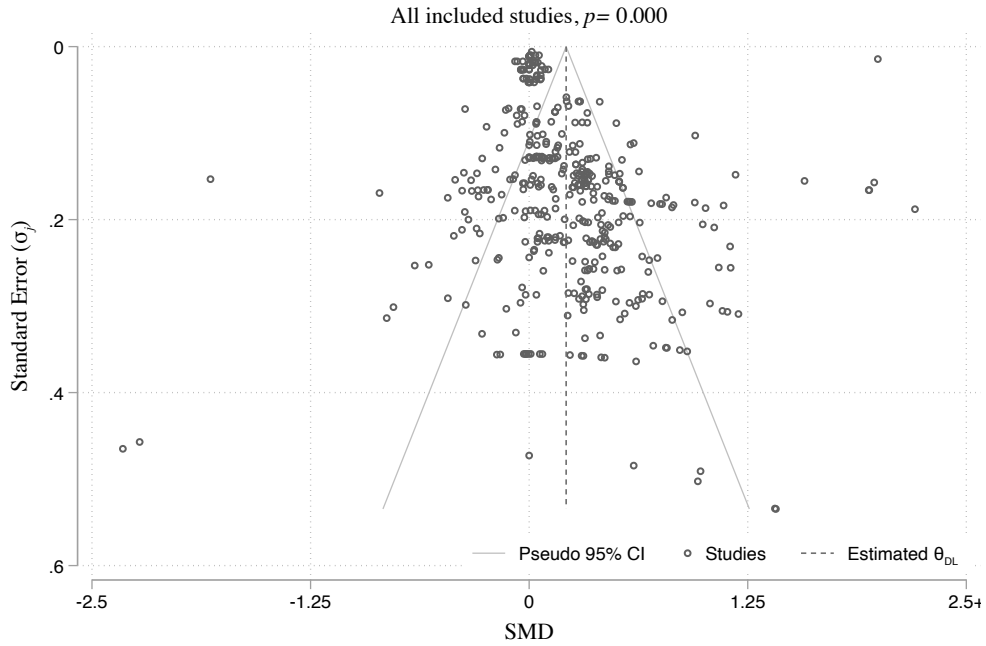
Note. Plot of the treatment effect on the treated group relative to the control group (difference in percentage points), as a function of the average pro-sociality in the control group ($N = 175$, see the Appendix, Section B.2). The control mean is normalized between 0 and 100%, where 100% is the most prosocial outcome feasible. The maximum is not systematically available, either because it was not reported in the paper or because the outcome had no limit, e.g., for charity donation or energy saving tasks.

size or study, with effects ranging from 0.23 to 0.26 and all $p < .001$. Throughout the paper, we also perform a systematic robustness analysis on trimmed data (provided in the SI, Section D), removing the top and bottom 2% of the distribution of effect sizes. In particular, Figure F reports the distribution of effect sizes when we remove outliers ($\widehat{\text{SMD}}_{\text{trim}} = 0.16$, 95% CI: [0.14, 0.17], $p < .001$).

Not surprisingly given the scope of our meta-analysis (see the SI, Table F, for descriptive statistics on the sample’s heterogeneity) there is substantial variability across studies, as indicated by the 95% prediction interval, $[-0.975; 1.457]$, which covers both positive and negative effects of quite large magnitude. Overall, 54% of studies find non-significant effects, 33.4% find significantly positive effects, and 12.2% significantly negative effects. This possible backfiring, and the conditions under which it occurs, will be examined later in the paper. When focusing on studies reporting positive effects, the average effect size is $\widehat{\text{SMD}} = 0.70$, 95% CI: [0.58, 0.82], whereas among studies reporting negative effects, the average is $\widehat{\text{SMD}} = -1.10$, 95% CI: $[-1.85, -0.32]$.

One of the dimensions that can explain the variability across studies, at least differences between the non-significant and positive impact, is the baseline rate of pro-social behavior. As argued, e.g., by DellaVigna and Linos (2022), it is typically hard for any intervention to shift

Figure 3: Funnel plot



Note. The funnel plot shows the average treatment effect and standard error of each study to test for small-study (or publication) bias, along with the p-value of the Egger statistic for small sample bias (i.e., statistical significance of the slope of the regression of effect sizes on standard errors Egger, Smith, Schneider, and Minder, 1997). The dotted lines show the pseudo 95% confidence interval of the overall treatment effect ($\widehat{sMD} = 0.24$) as the standard error varies. Under no small-study bias, we expect to see studies distributed symmetrically around the overall treatment effect, and closer to the \widehat{sMD} when precision is high.

upwards a behavior whose baseline rate is already high; while a very low baseline level indicates a target behavior that is very hard to induce and on which interventions will likely have a relatively small effect. The largest effect of behavioral interventions is thus expected to happen for intermediate values of the baseline target behavior. To examine this relationship, we draw on a metric that ensures cross-study comparability. We define the outcome as a percentage of maximally prosocial behavior—which is possible only for studies in which the target variable is bounded both above and below. Figure 2 plots the treatment effect against the normalized mean in the control group, for the subsample of 204 treatment effects for which such a normalization is feasible (see the Appendix, Section B.2), along with a quadratic adjustment of the data. As expected, we find an overall inverted U-shape relationship, with a decrease in the effect size that becomes particularly noticeable when the baseline behavior is above 60%.

Publication bias. Meta-analyses often uncover evidence of publication bias, whereby researchers tend to report successful rather than null or unsuccessful findings. Our meta-analysis is no exception. Figure 3 shows the distribution of the treatment effects relative to the average effect observed

in the meta-analysis, in terms of both their effect size and their precision. In the absence of small-study effects, the plot should look symmetrical as studies should be equally likely to find results above or below the overall treatment effect. In addition, we expect studies with lower standard error to be more closely concentrated around the overall effect. The symmetrical funnel shape shows the pseudo 95% confidence intervals in terms of the standard error, i.e. where we expect to find 95% of the treatment effects observed in the meta-analysis. There is an over-representation of positive effect sizes among studies with relatively low statistical power. Moreover, we observe a number of studies that precisely estimate a null effect at the top of Figure 3. This publication bias is confirmed by the corresponding Egger test, which finds a positive association between effects sizes and standard errors ($\hat{\beta} = 1.91, p < .001$; which corresponds to an almost 0.2 increase in the treatment effect when the standard-deviation increases by 10%).

Interestingly this publication bias seems to vary according to the type of methodological design employed. As shown in the SI, Figure C, publication bias is a particularly important phenomenon in laboratory and online experiments and much less so in field experiments. The Egger test finds evidence of small-study bias among online and laboratory experiments at the 1% significance level, among classroom experiments at the 5% level but fails to reject the null among field experiments ($p = .120$).⁵ While this may be explained by the higher cost of collecting additional observations in field experiments compared to laboratory settings, this pattern has, to the best of our knowledge, not been documented in previous meta-studies. This constitutes an added value of our meta-analysis, which integrates a wide range of methodological approaches.

3.2 Moderator Analyses

Our results show that exposure to moral reminders generally increases pro-social behavior, although such interventions can occasionally backfire. Given the considerable diversity of studies included in our analysis, it is important to address two subsequent questions: under what circumstances do moral reminders succeed, and when might they backfire? To provide a comprehensive picture of the factors that may influence the impact of moral reminders on pro-sociality, we next discuss how different moderators contribute to their effectiveness. We first examine whether the effect of reminders depends on methodological features of the studies in which they are implemented, with a particular focus on lab versus field settings, incentivized versus non-incentivized designs, and the specific measure of pro-sociality targeted. Since one of our main objectives is to inform the practical design of these interventions, we then analyze implementation moderators, in-

⁵There is also a difference across fields, with a publication bias particularly strong in economics and psychology. According to the Egger test, we find evidence of small-study bias at the 1% level within business ($\hat{\beta} = 0.91, p = .009$), economics ($\hat{\beta} = 3.28, p = .000$), psychology ($\hat{\beta} = 0.98, p = .004$) and others ($\hat{\beta} = 4.46, p < 0.001$), not among the studies categorized broadly as social sciences ($\hat{\beta} = 0.19, p = .780$). The presence of small-study bias is robust to including the methodology as a moderator for all disciplines (interestingly the p-value for social sciences decreases from 0.8 to 0.1).

Table 2: Methodological moderators

Study	K		Hedges's g with 95% CI	p-value
Elicitation method				
Field experiment	151		0.059 [-0.047, 0.165]	0.274
Class experiment	24		0.274 [0.191, 0.357]	0.000
Online experiment	82		0.245 [0.188, 0.302]	0.000
Laboratory experiment	158		0.422 [0.307, 0.537]	0.000
Test of group differences: $Q_b(3) = 21.41, p = 0.00$				
Incentives				
No incentive	258		0.140 [0.059, 0.221]	0.001
Incentivised	157		0.378 [0.312, 0.445]	0.000
Test of group differences: $Q_b(1) = 19.75, p = 0.00$				
Target measure of pro-sociality				
Intentions	35		0.040 [-0.007, 0.087]	0.093
Behavior	306		0.240 [0.165, 0.315]	0.000
Judgment	74		0.330 [0.235, 0.425]	0.000
Test of group differences: $Q_b(2) = 39.27, p = 0.00$				

Note. Effect sizes (\widehat{SMD}) across the categories listed in row (see the SI, Table G for descriptive statistics on the sample's distribution). The second column on the left corresponds to the number of effect sizes in each category. The columns on the right provide the standardized mean difference, the 95% confidence interval and the p-value at the 5% significance level.

cluding the source of the reminder and the field of application, as well as the presence of observers of participants' behavior. This can inform policy makers on the most effective way of designing these reminders.

Methodological moderators. Table 2 presents the effect of moral reminders across different study implementation settings. The effect of moral reminders appears stronger in laboratory experiments. However, the standardized mean difference remains positive in field experiments as well, and becomes statistically significant when we trim the data ($\widehat{SMD}_{\text{trim}} = 0.02, 95\% \text{ CI: } [0.01, 0.03], p = .002$). When we turn to comparing the effects of moral reminders in incentivized vs non-incentivized experiments, we observe a stronger effect for incentivized experiments. This finding may be due to a sharper moral conflict when material rewards are at stake. Indeed, monetary incentives typically reduce baseline cooperation by shifting attention toward self-interest (Bowles, 1998), but they also make the ethical dimension of the choice more salient, highlighting the conflict between material and moral motives in prosocial decisions. In this context, moral reminders become particularly effective, reactivating prosocial motives that incentives tend to suppress. As a result, although overall cooperation may remain lower in incentivized settings, the marginal

Table 3: Implementation moderators

Study	K		Hedges's g with 95% CI	p-value
Source of the reminder				
Experimenter	253		0.265 [0.164, 0.365]	0.000
External authority	106		0.181 [0.140, 0.222]	0.000
Other participants	11		0.138 [-0.105, 0.381]	0.264
Personal	38		0.185 [0.099, 0.270]	0.000
Religious authority	7		0.109 [0.022, 0.196]	0.014
Test of group differences: $Q_b(4) = 5.39, p = 0.25$				
Applications				
Decontextualized	104		0.361 [0.287, 0.436]	0.000
Social	58		0.317 [0.242, 0.392]	0.000
Environment	127		0.225 [0.147, 0.303]	0.000
Ressources	46		0.036 [0.006, 0.066]	0.020
Tax	41		0.065 [0.021, 0.108]	0.003
Finance	39		-0.034 [-0.648, 0.581]	0.914
Test of group differences: $Q_b(5) = 108.69, p = 0.00$				
Observers				
No observer	366		0.191 [0.124, 0.257]	0.000
Observer	49		0.566 [0.458, 0.674]	0.000
Test of group differences: $Q_b(1) = 33.64, p = 0.00$				
Overall			0.240 [0.178, 0.303]	0.000

Note. Effect sizes (\widehat{SMD}) across the categories listed in row (see the SI, Table H for descriptive statistics on the sample's distribution). The second column on the left corresponds to the number of effect sizes in each category. The columns on the right provide the standardized mean difference, the 95% confidence interval and the p-value at the 5% significance level.

impact of the moral reminder is amplified.

The third panel of Table 2 reports the effects of moral reminders across the different measures of pro-sociality they may target. We follow the classical distinction in moral psychology between judgments (the ability to recognize a decisions situation as morally loaded), intentions (the stated willingness to behave pro-socially in a given situation), and behaviors (Ellemers, van der Toorn, Paunov, and van Leeuwen, 2019). Interventions that influence judgments exhibit, on average, larger effect sizes, whereas changing actual behavior appears more difficult. This panel reveals an interesting result regarding the large literature in psychology about the behavior-intention gap (Ajzen, 1985; Sheeran, 2002): moral reminders have a stronger effect on behavior than on intentions. This finding is consistent with intentions being less prone to self-manipulation, and hence less sensitive to the increased salience of morality implemented by moral reminders.

Implementation moderators. Understanding the role of implementation moderators, that can be adjusted by designers, is particularly relevant for policy design. The effect of different types of implementation is summarized in Table 3. The first panel examines whether the source of the reminder influences its effectiveness. A striking result emerges: regardless of who delivers the reminder (e.g., the experimenter or an external authority), the standardized mean difference remains significantly positive.⁶ The effect is somewhat stronger when the reminder is voiced by the experimenter ($\widehat{\text{SMD}} = 0.265, 95\% \text{ CI: } [0.164, 0.365]$). This might raise concerns that the results are primarily driven by potential experimenter demand effects. However, this does not appear to be the case, as reminders issued by external authorities⁷ produce comparable effects ($\widehat{\text{SMD}} = 0.183, 95\% \text{ CI: } [0.141, 0.225]$). This finding is confirmed by the multivariate analysis reported in Table I, which shows that the source of the reminder does not have a statistically significant impact.⁸ Interestingly, when the source of moral reminders is a religious authority, the effect isn't larger than when it is issued by other external authorities, ($\widehat{\text{SMD}} = 0.109, 95\% \text{ CI: } [0.022, 0.196]$ vs. $\widehat{\text{SMD}} = 0.181, 95\% \text{ CI: } [0.140, 0.222]$).

From a policy perspective, it is important to understand whether the effectiveness of moral reminders depends on the type of application. Significant differences emerge in the second panel of Table 3. Reminders are particularly effective in promoting environmental and societal behaviors (such as voluntary civic engagement, e.g. Dean, Fielding, and Wilson, 2019), but substantially less so when aimed at resource conservation (e.g., recycling, water or energy use, such as in Pellerano, Price, Puller, and Sánchez, 2017) or tax compliance (see e.g., Bott, Cappelen, Sørensen, and Tungodden, 2020). Even if there are some important differences across fields, it is worth noting that moral reminders always have a positive and significant effect on average. In the case of tax applications, the average effect is $\widehat{\text{SMD}} = 0.07, 95\% \text{ CI: } [0.02; 0.11]$. The relatively smaller effect of moral reminders targeted at increasing tax compliance is in line with the findings of Antinyan and Asatryan (2024), who find that tax reminders are an effective tool for increasing pro-social behavior, but that reminders that appeal to morality or recall social norms are much less effective than reminders that leverage a deterrence effect by recalling the consequences in terms of sanctions for those who do not pay taxes.

A particularly strong moderator is the presence of observers. When participants know that their decision is observed by others, the average treatment effect of a moral reminder more than doubles ($\widehat{\text{SMD}} = 0.191, 95\% \text{ CI: } [0.1240, 0.257]$ vs. $\widehat{\text{SMD}} = 0.566, 95\% \text{ CI: } [0.458, 0.674]$). Observers are, in general, other participants in the experiment who do not take part in the task and observe

⁶Examples could include reminders delivered by researchers in laboratory settings or by institutional actors in field experiments.

⁷For instance, tax authorities in field experiments using official letters, or hotel management in experiments promoting towel reuse.

⁸While our dataset is large by the standards of meta-analyses, the number of available studies remains limited for models including many covariates.

Table 4: Classes of Moral Reminders

Study	K		Hedges's g with 95% CI	p-value
Priming	87		0.277 [0.205, 0.348]	0.000
Consequences	60		0.322 [0.182, 0.463]	0.000
Descriptive norm	135		0.182 [0.142, 0.222]	0.000
Injunctive norm	133		0.200 [-0.017, 0.417]	0.071
Test of group differences: $Q_b(3) = 7.64, p = 0.05$				
Overall			0.240 [0.178, 0.303]	0.000

0 .2 .4 .6

Note. Effect sizes ($\widehat{\text{SMD}}$) across the categories listed in row (see Table 1 for formal definitions). The second column on the left corresponds to the number of effect sizes in each category. The columns on the right provide the standardized mean difference, the 95% confidence interval and the p-value at the 5% significance level.

the decisions of the players, in particular whether they behave pro-socially. This effect remains stable across studies, even though the number of observers ranges from one to ten, indicating that the mere presence of an observer matters more than the number or intensity of observation (see, for example, Shariff, 2015; Rege and Telle, 2004). Taken together, these findings indicate that the effectiveness of moral reminders varies substantially across contexts and highlights ways to increase their impact.

3.3 A classification of moral reminders

In the previous sections, we documented a sizable average positive effect of moral reminders on prosocial behavior. As discussed in the introduction, this result was not a foregone conclusion, given individuals' tendency to process information and memory in a strategic manner. We now return to the classification of moral reminders introduced earlier, which distinguishes between reminders that operate by increasing the salience of moral motives and those that rely on the provision of morally relevant information (Table 1). There is currently no unified theoretical framework that predicts which classes of moral reminders should increase pro-social behavior and which are most likely to fail or even backfire. However, we provide some insights from the literature suggesting how these different tools could operate.

First, regarding information based instruments, there is strong evidence that people strategically avoid information about the consequences of their actions (Dana, Weber, and Kuang, 2007), while there is far less evidence that they avoid information about social norms. In everyday settings, norms are continuously reinforced through social interaction and are therefore highly salient; ignoring them carries reputational costs. By contrast, the consequences of actions are often uncertain or less cognitively accessible, and therefore easier to disregard. From this perspective,

Table 5: Multivariate analysis

	(1)	(2)	(3)	(4)
Consequences	0.075 (0.529)	0.164 (0.214)	0.325** (0.027)	0.436** (0.017)
Descriptive norm	-0.055 (0.572)	-0.053 (0.609)	0.057 (0.606)	0.148 (0.299)
Injunctive norm	-0.084 (0.395)	-0.072 (0.518)	-0.040 (0.734)	-0.026 (0.849)
Observer				0.302* (0.085)
Controls		✓	✓	✓
Methodological moderators			✓	✓
Implementation moderators				✓
Observations	415	415	415	415

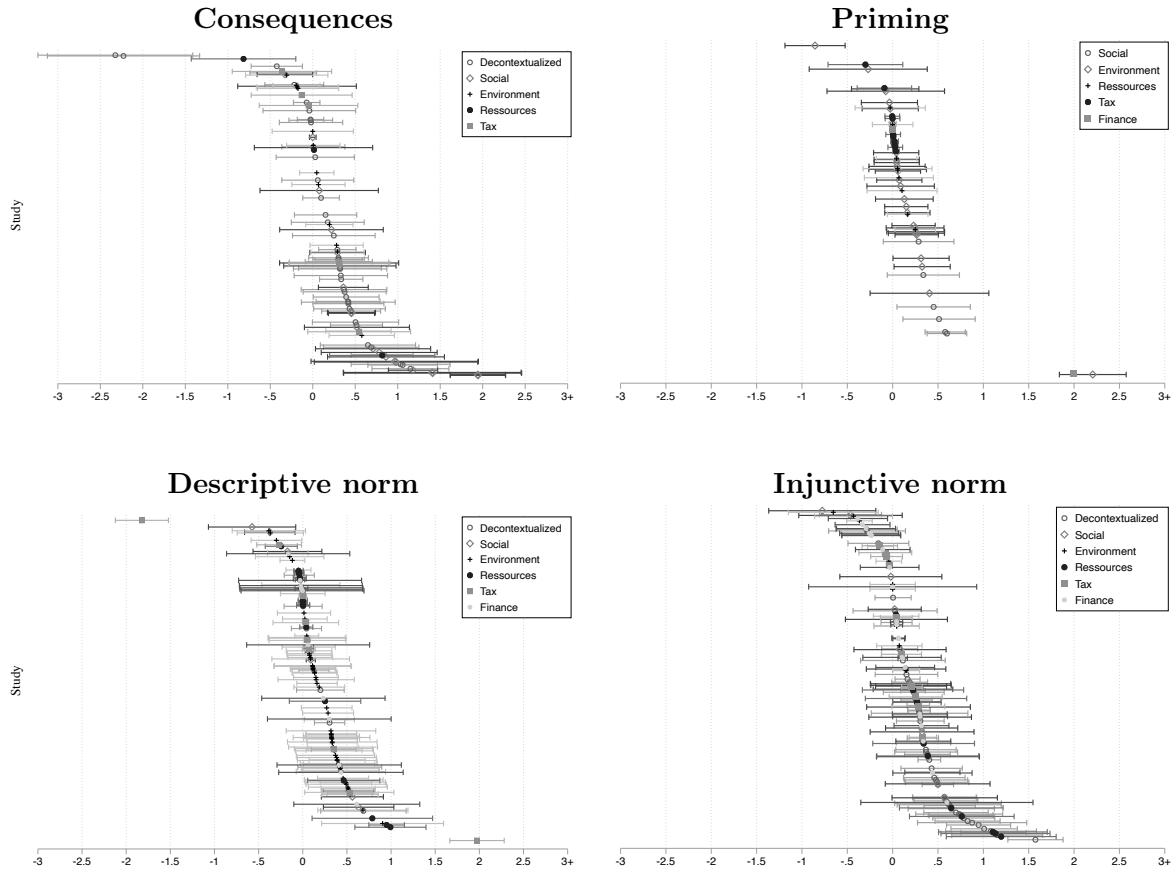
Note. Meta regression of the average treatment effect (see the SI, Section B.1, for a formal presentation of the model). Controls include the country where the study took place and the discipline to which the journal is affiliated; methodological moderators are defined in Table 2; implementation moderators are defined in Table 3 (full results available in the SI, Table I). See the SI, Table J, for a replication using a multi-level meta regression that allows for correlations of the sampling noise at the paper and the study levels. *Legend:* * $p < .10$, ** $p < .05$, *** $p < .01$.

reminders highlighting consequences should have more leverage than those that simply inform on social norms. Second, the literature also points out that part of the success of information-based reminders might be due to an increased salience of moral dimensions, rather than to a correction of beliefs.⁹ If this salience effect is large, we would expect information-based and priming reminders to have similar effects.

Table 4 provides a comparison of the effects of the different type of moral reminders introduced in Table 1. The first main finding is that all these different classes of reminders have a significant positive impact on pro-sociality (although it is only significant at the 10% level for injunctive norms). Priming ($\widehat{\text{SMD}} = 0.277, 95\% \text{ CI: } [0.205; 0.348]$) and providing information on consequences of pro-social behavior ($\widehat{\text{SMD}} = 0.322, 95\% \text{ CI: } [0.182; 0.463]$) work better than providing information on descriptive norms ($\widehat{\text{SMD}} = 0.182, 95\% \text{ CI: } [0.142; 0.222]$) and on injunctive norms ($\widehat{\text{SMD}} = 0.200, 95\% \text{ CI: } [-0.1017; 0.417]$), a result that fits with the theoretical considerations presented above. Table 5 suggests that these differences are in fact significant and that information on consequences dominates other classes of moral reminders as a tool to promote pro-sociality. This message should however be interpreted with caution, since as we show in Figure 4,

⁹In the same spirit, Guriev, Henry, Marquis, and Zhuravskaya (2025) show that fact-checking reduces the spread of disinformation not because it supplies corrective information, but because it makes accuracy concerns more salient.

Figure 4: Forest plot, by tool and application



Note. Forest plot of effect sizes by tools (refer to Table 4 for the average effect sizes), distinguishing the applications.

where we plot the distribution of SMD by instrument, there are important outliers. Moreover, the 95% prediction intervals (PI) for all four tools remain wide and, crucially, all include zero (PI= $[-0.237, 0.790]$ for priming, PI= $[-0.893, 1.220]$ for consequences, PI= $[-0.134, 0.374]$ for descriptive norms and PI= $[-1.178, 1.718]$ for injunctive norms). This implies that although some tools have higher average effects, we cannot exclude the risk of backfiring for any of these moral reminders.

Moderators of backfiring We explore in this section what type of reminder or implementation moderator are most conducive to backfiring, an effect observed in 12% of the included studies. Results are presented in Table 6, where the dependent variable is an indicator variable taking value 1 if the SMD is significantly smaller than zero. First, we find that studies trying to encourage tax payments are much more likely to backfire, an increase of 12 percentage points. This result

Table 6: Predictors of a (significantly) negative effect

	(1)	(2)	(3)	(4)
Consequences	-0.037 (0.203)	-0.025 (0.437)	-0.038 (0.221)	-0.025 (0.595)
Descriptive norm	-0.002 (0.959)	0.002 (0.952)	-0.004 (0.901)	0.001 (0.975)
Injunctive norm	0.030 (0.371)	0.042 (0.245)	0.038 (0.216)	0.058* (0.090)
Observer				-0.079* (0.052)
Controls		✓	✓	✓
Methodological moderators			✓	✓
Implementation moderators				✓
Observations	415	415	415	415

Note. Linear regressions of the dummy variable indicating whether a study identifies en significantly negative effect. Robust standard in parenthesis. Controls include the country where the study took place and the discipline to which the journal is affiliated; methodological moderators are defined in Table 2; implementation moderators are defined in Table 3 (full results available in the SI, Table K). *Legend:* * $p < .10$, ** $p < .05$, *** $p < .01$.

is coherent with the tax compliance experiment of Neve, Imbert, Spinnewijn, Tsankova, and Luts (2021), who show that moral reminders treatments backfire whereas reminders on tax enforcement increase tax compliance. Second, consistently with the evidence about their average effect we find relevant differences in the impact of different classes of moral reminders on backfiring. Interventions that inform participants about the injunctive norm are substantially more likely to backfire than other classes of reminders. One possible interpretation is that the communicated norm lies below participants' average expectations, thereby discouraging pro-social behavior. This pattern is not observed for interventions providing information on descriptive norms, suggesting that they are probably better assessed *ex-ante* by experimental subjects. Finally, we find that including observers significantly decreases backfiring. The effect size roughly compensates the effect of providing information on the injunctive norm rather than on consequences. This is coherent with the fact that this backfiring seems to mostly operate through exposition to social norms.

4 Discussion

Our meta-study shows that moral reminders work and that this effectiveness is maintained across fields of application, design choices and types of interventions. It does point out however that this overall effectiveness comes along with backfiring in certain circumstances. When implemented to promote tax payments, moral reminders are less likely to succeed and more prone to backfiring. On

the contrary, moral reminders implemented with observers are more likely to succeed and less likely to backfire — suggesting that social image concerns play an important role in the effectiveness of moral reminders.

We also provide a comparison of the different classes of reminders as classified in Table 1. The first and main finding is that both information based instruments and tools priming morals work on average. The discussion in the introduction suggested that this was far from obvious. The results suggest that the cognitive frictions created by the reminders may be sufficient to overcome motivated reasoning (in the case of information-based tools) or motivated memory (in the case of priming). These reminders are likely creating costs that appear sufficient to prevent these mechanisms from fully operating.

A meta-analysis is however not designed to uncover mechanisms. We cannot tell if information-based reminders work through the information they provide, or just because they make morality more salient. What the work does show is that all instruments work, in particular providing information about consequences of actions. It also points out that information on the injunctive norm is more likely to backfire and that the effects of reminders are sensitive to the context of application, suggesting that policy makers should be particularly careful about implementation and contextual circumstances. Further work on these topics is needed, to uncover more clearly the mechanisms determining the role of these different moral appeals in promoting pro-social behavior.

Data availability

The data from our experiment, as well as the codes used to generate the statistical analysis, will be provided as part of our replication package, available at https://osf.io/tcn5p/overview?view_only=5efd77eb240a461880a55197d454dc72, once final. Further heterogeneity analyses can be investigated using the online data visualisation tool published at <https://sites.google.com/view/so-self-meta-analysis/home?authuser=5>, which also presents additional results and analysis of the datasets collected for this study.

References

- AJZEN, I. (1985): “From Intentions to Actions: A Theory of Planned Behavior,” in *Action Control*, pp. 11–39. Berlin Heidelberg: Springer.
- AMELIO, A., AND F. ZIMMERMANN (2023): “Motivated Memory in Economics—A Review,” *Games*, 14(1), 15, Publisher: Multidisciplinary Digital Publishing Institute.
- ANTINYAN, A., AND Z. ASATRYAN (2024): “Nudging for Tax Compliance: A Meta-Analysis,” *Economic Journal*, 135(668), 1033–1068.

- BÉNABOU, R., A. FALK, AND J. TIROLE (2020): “Narratives, Imperatives, and Moral Persuasion,” *University of Bonn, mimeo*.
- BÉNABOU, R., AND L. HENKEL (2025): “Identity as Self-Image,” *NBER WP*, (34297).
- BENARTZI, S., J. BESHEARS, K. L. MILKMAN, C. R. SUNSTEIN, R. H. THALER, M. SHANKAR, W. TUCKER-RAY, W. J. CONGDON, AND S. GALING (2017): “Should Governments Invest More in Nudging?,” *Psychological Science*, 28(8), 1041–1055.
- BICCHIERI, C. (2006): *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.
- BOTT, K. M., A. W. CAPPELEN, E. Ø. SØRENSEN, AND B. TUNGODDEN (2020): “You’ve Got Mail: A Randomized Field Experiment on Tax Evasion,” *Management Science*, 66(7), 2801–2819.
- BOWLES, S. (1998): “Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions,” *Journal of economic literature*, 36(1), 75–111.
- BRADLEY, A., C. LAWRENCE, AND E. FERGUSON (2018): “Does Observability Affect Prosociality?,” *Proceedings of the Royal Society B: Biological Sciences*, 285(1875), 20180116.
- CHETTY, R., E. SAEZ, AND L. SANDOR (2014): “What Policies Increase Prosocial Behavior? An Experiment with Referees at the Journal of Public Economics,” *Journal of Economic Perspectives*, 28(3), 169–188.
- DAL BÓ, E., AND P. DAL BÓ (2014): ““Do the Right Thing:” The Effects of Moral Suasion on Cooperation,” *Journal of Public Economics*, 117, 28–38.
- DANA, J., R. A. WEBER, AND J. X. KUANG (2007): “Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness,” *Economic Theory*, 33(1), 67–80.
- DEAN, A. J., K. S. FIELDING, AND K. A. WILSON (2019): “Building Community Support for Coastal Management — What Types of Messages Are Most Effective?,” *Environmental Science & Policy*, 92, 161–169.
- DELLAVIGNA, S., AND E. LINOS (2022): “RCTs to Scale: Comprehensive Evidence From Two Nudge Units,” *Econometrica*, 90(1), 81–116.
- DIMANT, E., G. A. VAN KLEEF, AND S. SHALVI (2020): “Requiem for a Nudge: Framing Effects in Nudging Honesty,” *Journal of Economic Behavior & Organization*, 172, 247–266.
- EGGER, M., G. D. SMITH, M. SCHNEIDER, AND C. MINDER (1997): “Bias in Meta-Analysis Detected by a Simple, Graphical Test,” *BMJ*, 315(7109), 629–634.
- ELLEMERS, N., J. VAN DER TOORN, Y. PAUNOV, AND T. VAN LEEUWEN (2019): “The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017,” *Personality and Social Psychology Review*, 23(4), 332–366.

- EPLEY, N., AND T. GILOVICH (2016): “The Mechanics of Motivated Reasoning,” *Journal of Economic Perspectives*, 30(3), 133–140.
- EXLEY, C. L. (2020): “Using Charity Performance Metrics as an Excuse Not to Give,” *Management Science*, 66(2), 553–563.
- GURIEV, S., E. HENRY, T. MARQUIS, AND E. ZHURAVSKAYA (2025): “Curtailling False News, Amplifying Truth,” *available at SSRN*.
- HUMMEL, D., AND A. MAEDCHE (2019): “How Effective Is Nudging? A Quantitative Review on the Effect Sizes and Limits of Empirical Nudging Studies,” *Journal of Behavioral and Experimental Economics*, 80, 47–58.
- KREPLIN, U., M. FARIAS, AND I. A. BRAZIL (2018): “The Limited Prosocial Effects of Meditation: A Systematic Review and Meta-Analysis,” *Scientific Reports*, 8(1), 2403.
- KUNDA, Z. (1990): “The Case for Motivated Reasoning,” *Psychological Bulletin*, 108(3), 480–498.
- MERTENS, S., M. HERBERZ, U. J. J. HAHNEL, AND T. BROSCH (2022): “The Effectiveness of Nudging: A Meta-Analysis of Choice Architecture Interventions across Behavioral Domains,” *Proceedings of the National Academy of Sciences*, 119(1), e2107346118.
- NEVE, J.-E. D., C. IMBERT, J. SPINNEWIJN, T. TSANKOVA, AND M. LUTS (2021): “How to Improve Tax Compliance? Evidence from Population-Wide Experiments in Belgium,” *Journal of Political Economy*, 129(5), 1425–1463.
- PALOMO-VÉLEZ, G., J. M. TYBUR, AND M. VAN VUGT (2018): “Unsustainable, Unhealthy, or Disgusting? Comparing Different Persuasive Messages against Meat Consumption,” *Journal of Environmental Psychology*, 58, 63–71.
- PAPAKONSTANTINOY, T., S. L. FLECKE, C. E. R. EDMUNDS, R. CROSS, A. TRAN, AND N. GOLD (2025): “A Systematic Review and Meta-Analysis of the Effectiveness of Social Norms Messaging Approaches for Improving Health Behaviours in Developed Countries,” *Nature Human Behaviour*, pp. 1–19.
- PELLERANO, J. A., M. K. PRICE, S. L. PULLER, AND G. E. SÁNCHEZ (2017): “Do Extrinsic Incentives Undermine Social Norms? Evidence from a Field Experiment in Energy Conservation,” *Environmental and Resource Economics*, 67(3), 413–428.
- RABIN, M. (1998): “Psychology and Economics,” *Journal of Economic Literature*, 36(1), 11–46.
- RAIHANI, N. J., AND K. MCAULIFFE (2014): “Dictator Game Giving: The Importance of Descriptive versus Injunctive Norms,” *PLoS ONE*, 9(12), e113826.
- REGE, M., AND K. TELLE (2004): “The Impact of Social Approval and Framing on Cooperation in Public Good Situations,” *Journal of Public Economics*, 88(7-8), 1625–1644.
- SACHDEVA, S., R. ILIEV, AND D. L. MEDIN (2009): “Sinning Saints and Saintly Sinners: The Paradox of Moral Self-Regulation,” *Psychological Science*, 20(4), 523–528.

- SCHWARZER, G., J. R. CARPENTER, AND G. RÜCKER (2015): *Meta-Analysis with R*, vol. 4784. Springer.
- SHARIFF, A. F. (2015): “Does Religion Increase Moral Behavior?” *Current Opinion in Psychology*, 6, 108–113.
- SHEERAN, P. (2002): “Intention—Behavior Relations: A Conceptual and Empirical Review,” *European Review of Social Psychology*, 12(1), 1–36.
- SZASZI, B., A. PALINKAS, B. PALFI, A. SZOLLOSI, AND B. ACZEL (2018): “A Systematic Scoping Review of the Choice Architecture Movement: Toward Understanding When and Why Nudges Work,” *Journal of Behavioral Decision Making*, 31(3), 355–366.
- VLASCEANU, M., K. C. DOELL, J. B. BAK-COLEMAN, B. TODOROVA, M. M. BERKEBILE-WEINBERG, S. J. GRAYSON, Y. PATEL, D. GOLDWERT, Y. PEI, A. CHAKROFF, E. PRONIZIUS, K. L. VAN DEN BROEK, D. VLASCEANU, S. CONSTANTINO, M. J. MORAIS, P. SCHUMANN, S. RATHJE, K. FANG, S. M. AGLIOTI, M. ALFANO, A. J. ALVARADO-YEPEZ, A. ANDERSEN, F. ANSEEL, M. A. J. APPS, C. ASADLI, F. J. AWUOR, F. AZEVEDO, P. BASAGLIA, J. J. BÉLANGER, S. BERGER, P. BERTIN, M. BIAŁEK, O. BIALOBRZESKA, M. BLAYA-BURGO, D. N. M. BLEIZE, S. BØ, L. BOECKER, P. S. BOGGIO, S. BORAU, B. BOS, A. BOUGUETTAYA, M. BRAUER, C. BRICK, T. BRIK, R. BRIKER, T. BROSCHE, O. BUCHEL, D. BUONAURO, R. BUTALIA, H. CARVACHO, S. A. E. CHAMBERLAIN, H.-Y. CHAN, D. CHOW, D. CHUNG, L. CIAN, N. COHEN-EICK, L. S. CONTRERAS-HUERTA, D. CONTU, V. CRISTEA, J. CUTLER, S. D’OTTONE, J. DE KEERSMAECKER, S. DELCOURT, S. DELOUVÉE, K. DIEL, B. D. DOUGLAS, M. A. DRUPP, S. DUBEY, J. EKMANIS, C. T. ELBAEK, M. ELSHERIF, I. M. ENGELHARD, Y. A. ESCHER, T. W. ETIENNE, L. FARAGE, A. R. FARIAS, S. FEUERRIEGEL, A. FINDOR, L. FREIRA, M. FRIESE, N. P. GAINS, A. GALLYAMOVA, S. J. GEIGER, O. GENSCHOW, B. GJONESKA, T. GKIPOPOULOS, B. GOLDBERG, A. GOLDENBERG, S. GRADIDGE, S. GRASSINI, K. GRAY, S. GRELE, S. M. GRIFFIN, L. GRIGORYAN, A. GRIGORYAN, D. GRIGORYEV, J. GRUBER, J. GUILARAN, B. HADAR, U. J. HAHNEL, E. HALPERIN, A. J. HARVEY, C. A. P. HAUGESTAD, A. M. HERMAN, H. E. HERSHFIELD, T. HIMICHI, D. W. HINE, W. HOFMANN, L. HOWE, E. T. HUAMAN-CHULLUNCUY, G. HUANG, T. ISHII, A. ITO, F. JIA, J. T. JOST, V. JOVANOVIĆ, D. JURGIEL, O. KÁCHA, R. KANKAANPÄÄ, J. KANTOROWICZ, E. KANTOROWICZ-REZNICHENKO, K. KAPLAN MINTZ, I. KAYA, O. KAYA, N. KHACHATRYAN, A. KLAS, C. KLEIN, C. A. KLÖCKNER, L. KOPPEL, A. I. KOSACHENKO, E. J. KOTHE, R. KREBS, A. R. KROSCHE, A. P. KROUWEL, Y. KYRYCHENKO, M. LAGOMARSINO, C. LAMM, F. LANGE, J. LEE CUNNINGHAM, J. LEES, T. Y. LEUNG, N. LEVY, P. L. LOCKWOOD, C. LONGONI, A. LÓPEZ ORTEGA, D. D. LOSCHELDER, J. G. LU, Y. LUO, J. LUOMBA, A. E. LUTZ, J. M. MAJER, E. MARKOWITZ, A. A. MARSH, K. L. MASCARENHAS, B. MBILINGI, W. MBUNGU, C. MCHUGH, M. H. MEIJERS, H. MERCIER, F. L. MHAGAMA, K. MICHALAKIS, N. MIKUS, S. MILLIRON, P. MITKIDIS, F. S. MONGE-RODRÍGUEZ, Y. L. MORA, D. MOREAU, K. MOTOKI, M. MOYANO, M. MUS, J. NAVAJAS, T. L. NGUYEN, D. M. NGUYEN, T. NGUYEN, L. NIEMI, S. R. R. NIJSSEN, G. NILSONNE, J. P. NITSCHKE, L. NOCKUR, R. OKURA, S. ÖNER, A. A. ÖZDOĞRU, H. PALUMBO, C. PANAGOPOULOS, M. S. PANASITI, P. PÄRNAMETS, M. PARUZEL-CZACHURA, Y. G. PAVLOV, C. PAYÁN-GÓMEZ, A. R. PEARSON, L. PEREIRA DA COSTA, H. M. PETROWSKY, S. PFATTHEICHER, N. T. PHAM, V. PONIZOVSKIY, C. PRETUS, G. G. RÊGO, R. REIMANN, S. A. RHOADS, J. RIANO-MORENO, I. RICHTER, J. P. RÖER, J. ROSA-SULLIVAN, R. M.

- ROSS, A. SABHERWAL, T. SAITO, O. SARRASIN, N. SAY, K. SCHMID, M. T. SCHMITT, P. SCHOENEGGER, C. SCHOLZ, M. G. SCHUG, S. SCHULREICH, G. SHREEDHAR, E. SHUMAN, S. SIVAN, H. SJÅSTAD, M. SOLIMAN, K. SOUD, T. SPAMPATTI, G. SPARKMAN, O. SPASOVSKI, S. K. STANLEY, J. A. STERN, N. STRAHM, Y. SUKO, S. SUL, S. SYROPOULOS, N. C. TAYLOR, E. TEDALDI, G. TINGHÖG, L. D. T. HUYNH, G. A. TRAVAGLINO, M. TSAKIRIS, Í. TÜTER, M. TYRALA, Ö. M. ULUĞ, A. URBANEK, D. VALKO, S. VAN DER LINDEN, K. VAN SCHIE, A. VAN STEKELENBURG, E. VANAGS, D. VÄSTFJÄLL, S. VESELY, J. VINTR, M. VRANKA, P. O. WANGUCHE, R. WILLER, A. D. WOJCIK, R. XU, A. YADAV, M. ZAWISZA, X. ZHAO, J. ZHAO, D. ŽUK, AND J. J. VAN BAVEL (2024): “Addressing Climate Change with Behavioral Science: A Global Intervention Tournament in 63 Countries,” *Science Advances*, 10(6), eadj5778.
- VU, L., I. SORAPERRA, M. LEIB, J. VAN DER WEELE, AND S. SHALVI (2023): “Ignorance by Choice: A Meta-Analytic Review of the Underlying Motives of Willful Ignorance and Its Consequences.,” *Psychological Bulletin*, 149(9-10), 611.
- ZHANG, J., Y. HUANG, J. ZHU, AND L. ZHAO (2023): “A Meta-Analysis on the Effectiveness of Food-Waste Reducing Nudges,” *Food Policy*, 120, 102480.
- ZICKFELD, J. H., S. T. S. KARG, S. S. ENGEN, A. S. R. GONZALEZ, J. MICHAEL, AND P. MITKIDIS (2024): “Committed (Dis)Honesty: A Systematic Meta-Analytic Review of the Divergent Effects of Social Commitment to Individuals or Honesty Oaths on Dishonest Behavior,” *Psychological Bulletin*, pp. 586–620.

Online Supplementary Information

A Data Construction

A.1 Data Collection

Literature search. A systematic review of the literature on moral reminders with the software Rayyan helped identify potentially relevant articles.¹⁰ Rayyan is an artificial intelligence program dedicated to meta-analyses and widely used in the literature (particularly in medical fields). The software performs a keyword analysis of the titles and abstract of references available on google scholar and then uses reviewers decision to include/exclude studies to generate a 1-5 star ratings of the likelihood of inclusion for each study. Rayyan requires at least 50 inclusion/exclusion decisions in order to compute reliable ratings. We classified all articles rated 3+ stars which corresponds to almost 2,500 articles. The list of keywords and the references used to initialize the inclusion decisions stemmed from two key sources: we started from the experimental papers referenced in the recent comprehensive review of the literature contained in Bénabou, Falk, and Tirole (2020), along with the secondary references in the corresponding papers, from which we extracted the list of keywords presented in Table A.

Criteria for inclusion. Out of the 3+ stars matches, the following exclusion criteria were applied to decide whether a given study falls in the scope of the current meta-analysis:

- There was no reminder, or the reminder had no moral component,
- The identification of the causal effect of the reminder is non-experimental (e.g. IV identification strategies are excluded, a non causal measure like, e.g., individuals who have a lower self-esteem being more pro-social, would be excluded),
- The outcome was not pro-social (e.g. sovereign debt default risk),
- There was no control group,
- The experiment focused on children,
- The article was not written in English,
- The article was published after 2020.

We further excluded studies whose design involved commitment from the participants (relying on participants' compliance with a preliminary request to change their subsequent behavior; e.g., foot-in-door interventions) or social comparisons (explicitly comparing one's behavior with the behavior of peers), and articles that were retracted by the editor. Last, we further complemented the literature search by an open call for papers, listing the keywords and the exclusion criteria that circulated in May 2025 within several professional organizations (*Economic Science Association, Society for the Advancement of Behavioral Economics*) — see the PRISMA chart in Figure A for more precise information, and Section E for the full list of references.

¹⁰<https://www.rayyan.ai/>

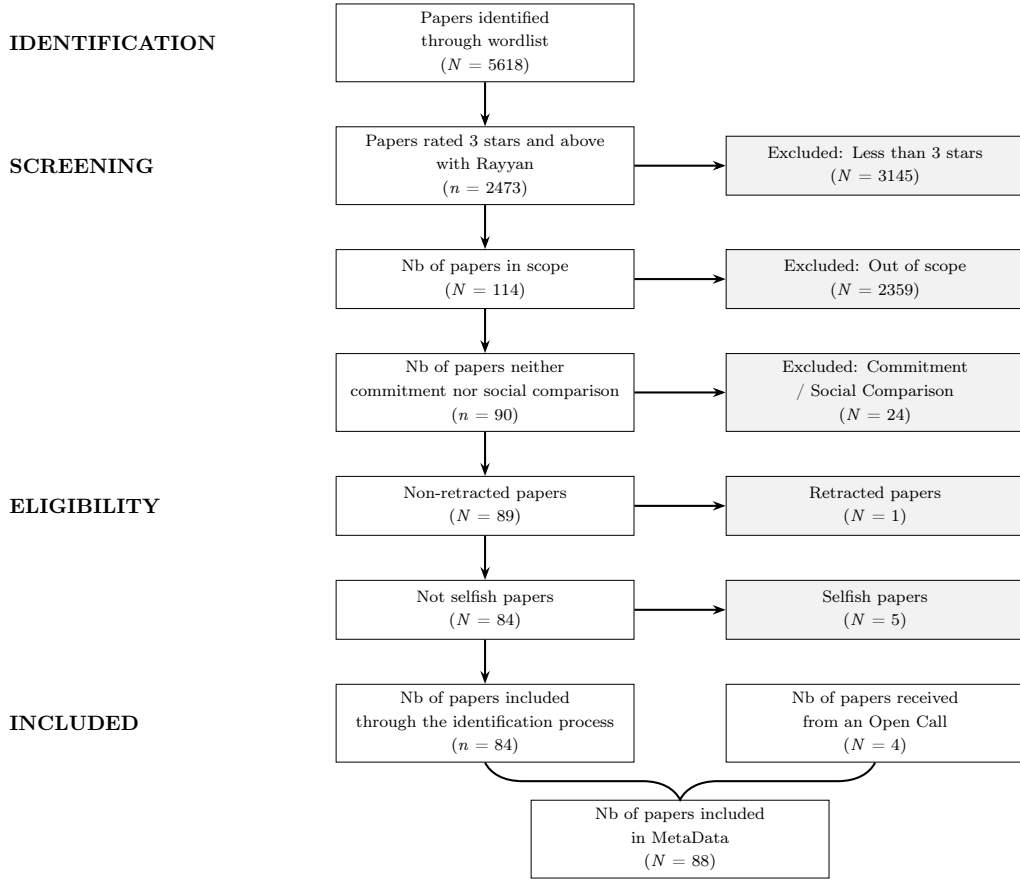
Table A: Keywords for Inclusion

Moral identity	Moral cleansing	Morality
Moral suasion	Non-binding recommendation	Tax compliance
Moral licensing	Threatened morality	Social norms
Moral framing	Normative information	Moral hypocrisy
Moral self	Moral purity	Fairness
Moral compensation	Moral standards	Cooperation
Moral credentials	Conscience appeal	Honest
Moral hypocrisy	Recommended contribution	Moral
Licensing effect	Moral persuasion	Social
Moral regulation	Experiment	Ethic
Normative appeal	Motivation	Self
Moral appearance	Framing	Norms
Moral self-worth	Public goods	Behavior
Moral appeal	Self-concept	

Meta data collection. Based on the published version of these articles, we collected information characterizing the journal, date, field and discipline of the article, the main characteristics of the experiment (including the outcomes, the samples, the type of game played, the behavior measured, the pay-offs, a detailed descriptions of the treatment and control conditions, anonymity, whether observers were present, etc). The data was recorded such that each line of data corresponds to one comparison between a treatment group and a control group. We collect information in each case that is as precise as possible on the treatment effect (means and standard deviations if possible and regression coefficients otherwise). Each comparison between one treatment group and one counter-factual group (control or placebo) is hereafter denoted an effect size (k). The resulting dataset, called Meta Data (MD), provides study-level information for each effect size.

Individual-level data collection. In addition to recording the publicly available information, we also tried to access the individual data of all 122 included studies. We found the data publicly available online for 18 studies. We then contacted authors for 95 studies asking about the availability of the data (when current contact information was either available in the published version of the paper or on professional web-pages; the contact information could not be retrieved for 6 studies). The response rate was 72% (i.e., we heard back from authors of 68 studies). In total, we obtained the data for 58 studies among which only 49 can be included (for the remaining 9, the codebook wasn't available and we did not manage to exploit the data nor to obtain clarifications from the authors). The data was not available for 28 studies either because it required external approval (e.g. from tax services), or has been lost, or the authors did not want to share them. In the end, the sample of individual data covers 49 studies out of 122.

Figure A: PRISMA chart



Note. See the main text for the full list of exclusion criteria used to consider papers as out of scope. Articles are excluded if and only if (i) all reminders push towards selfish behaviour, and/or rely on social comparison/commitment. The sample sizes cannot be interpreted as the number of articles featuring studies with selfish reminders or social comparison/commitment.

A.2 Data Consolidation

We use the resulting Data Studies (DS) dataset to fill in the missing information we were not able to extract from the published version of the article. It was in particular useful for articles which only report results as regression coefficients with covariates instead of the standard group means, standard deviations and sizes necessary for meta-analyses. Since we managed to collect the individual data for only a subset of the included studies (see Section A.1), we always give priority to the means, standard deviations and group sizes available in the MD (as it would impose a different treatment on studies for which the data has been shared with us). We (i) use the information in MD to double check the integrity of the data cleaning we applied to DS (we used the scripts shared by the authors whenever possible, but sometimes had to replicate them or complement them); and (ii) use the information available in DS to complement the missing informations in MD (in general, because the info was not reported in the manuscript).

A.2.1 Sanity Check: Comparison between MD and DS information

To check that we correctly replicated the same data cleaning as the authors for each study, we compare the information we can retrieve from DS to the information reported in the paper (stored in MD) in several steps. We start with the most conservative comparison, based both on the mean value of the outcome variable and the sample size in each treatment. After allowing for discrepancies of up to 10^{-1} between the two sources that may be due to rounding, 109 effect sizes out of 168 achieve a perfect match between the group means in the two datasets, and 40 additional studies match on identical group sizes in both sources.

This leaves 15 effect sizes out of 168 in DS that did not pass the means and group size comparisons. Discrepancies are then mainly due to regressions with controls that do not match group means. In this case, we tried to replicate as much as possible of the results reported in the paper to understand where the discrepancies came from, and contacted the authors if we could not explain the differences. In the end, we managed to replicate the results of 11 effect sizes thanks to the help of the authors (i.e., the differences were due to the specifics of the statistical treatment of the data; e.g., mean differences versus regressions with covariates).

We were unable to replicate the results for the remaining 4 effect sizes: we found either different regression coefficients (1 effect size) or different p-values (3 treatment effects). Since we don't want to penalize the studies for which individual data has been received, we keep the corresponding data in MD. Given the mismatch with the results we retrieve in DS for these studies, we however exclude them from the DS dataset (i.e., treat the corresponding studies as if no data has been received).

A.2.2 Outcome variable

Depending on the experimental design of the included studies, three types of comparisons are available in the dataset:

- Cross-Sectional (CS) comparisons between treatment and control groups (310 effect sizes);
- Before-After (BA) comparisons within group before and after treatment (23 effect sizes);
- Difference-in-Differences (DiD) comparisons between treatment and control groups both before and after treatment (82 effect sizes).

The control group used to perform these comparisons can be either a 'super control' (i.e., participants are not exposed to any intervention, 228 effect sizes) or a placebo (i.e., participants are exposed to an intervention similar to the one implemented in the main treatment but which does not contain any moral reminder, 187 effect sizes).

We compute DiD comparisons whenever information is available for both control and treatment groups both before and after intervention. Then we compute either Control-Treatment comparisons (between individuals) or Before-After comparisons (within individuals), depending on the design of the study. Note, however, that we give priority to the inclusion of studies into the quantitative meta-analysis: if, e.g., we have to choose between including a DiD comparison into the significance analysis only, or the TE comparison into the meta-analysis for the same study, we favor the TE measure. As is standard in a meta-analysis, we also favor a TE measure when means and standard deviations are available for each group, compared to any measure based on a coefficient and a standard error (because of potential additional controls in a regression or lack of information about the test used).

Table B: Assumptions necessary to infer missing information in MD

Assumption	# studies with group sizes	# studies without sizes
DiD comparisons	48	
Missing group sizes		45
Missing standard deviations		
inferred from standard errors	18	1
filled with p-values	8	17
filled with p-values' upper bounds	11	19
with p-values' UB & missing coeff.	5	2

Note. Inference from DiD comparisons is performed under (conservative) assumption that $\mathbf{Cov}(before, after) = 0$. Missing group sizes requires the assumption that $N_T = N_C = N/2$. Missing standard deviations filled with standard errors assume independent t-test with equal variances across groups. Missing standard deviations filled with p-values assume independent t-test with equal variances and that t-stat follows a normal distribution with $N - 2$ degrees of freedom. Missing standard deviations filled with p-values' upper bounds assume independent t-test with equal variances across groups, that t-stat follows a normal distribution with $N - 2$ degrees of freedom and that the p-value is equal to the upper bound. Missing coefficients are replaced with group means and assume that the p-value or standard error were observed in regressions with the same unit as in the descriptive statistics.

A.2.3 Inference of the remaining missing information

The information that remain missing after consolidating the MD dataset thanks to the DS data is inferred from the available information based on the following set of assumptions (summarized in Table B).

DiD comparisons. Many studies only provide the results (both point-estimates and standard errors) for before-after differences. In most instances, the covariance between a given group's pre- and post-treatment values is not reported. To infer the post-treatment variances in each group from such data, we thus assume the four groups (control/treatment combined with pre/post) are independent (hence over-estimating the variance) and the post-treatment variance is the same in both groups — allowing to infer the standard deviation of the post-treatment outcomes from the standard deviation of the difference in outcomes.

Missing group sizes. In some cases, only the total number of participants is reported. We thus assume that participants were equally distributed between the treatment and the control groups.

Missing standard deviations. For articles that report the treatment effect as a regression coefficient, we save the coefficient as the mean of the treatment group and 0 as the mean of the control group. This is equivalent to a standard mean difference in a meta-analysis if the regression does not include covariates. Regression coefficients take precedence when at least one of the means or standard deviations is missing. In this case, we draw information from the estimate of data dispersion that requires the least assumptions: first standard errors, then p-values, then p-value upper bounds. When only p-values or p-values upper bounds and means are available (while coefficients are missing), we compute standard deviations by applying the

Table C: Source of the information recorded in MD

Source	Nb. of studies (k)
Means, standard deviations and group sizes available	241
... from MD	128
... from DS	113
Means, standard deviations and group sizes inferred from MD	174
(see Table B for more details)	
Excluded due to missing information	115

Note. All categories are mutually exclusive. If studies fit in multiple categories, higher ranked categories take precedence.

p-value to the difference between means. We assume that the unit of measurement is identical between the descriptive statistics and the econometric estimations. For studies that provide only the standard errors or p-values from statistical tests, we can compute the standard deviations for t-tests and Wald tests by assuming that the treatment and control groups have identical distributions and variance. To do so, we assume that the t-tests reported in articles are independent two samples t-tests.

This process results in 415 studies for which quantitative information about effect sizes is included in the analysis. Table C describes the distribution of the included studies over these different sources of information.

A.3 Extracted variables

The following additional variables are directly extracted from the information available in the paper:

- **Control group:** Whether the control to which the treatment is compared is exposed to no intervention or to a placebo intervention.
- **Country:** For each study, the variable records the country in which the study has been run.
- **Elicitation method:** Refers to the implementation of the experiment (online/ classroom / physical lab / field).
- **Incentivized:** Whether the experiment relies on performance-based incentives.
- **Observers:** Whether the pro-social action is performed under the scrutiny of another participant.
- **Year of publication:** Year of publication of the paper.

A.4 Imputed variables

On top of the raw information extracted from the studies dataset, our data also include variables that have been created to describe the heterogeneity in the MD datasets.

A.4.1 Moral reminders classification

In order to compare treatment effects across studies, moral reminders were classified in different categories:

- The **classes of moral reminders** describe the content of the reminder:
 - Priming: exposure to a task (word decoding, text reading) aimed at making salient the ethical drivers of a decision;
 - Consequences: provide information about the consequences of the available choices;
 - Descriptive norms: provide information on the descriptive norm (what others do);
 - Injunctive norms: provide information on the injunctive norm (what others think should be done).
- The **source** describes who emits each reminder:
 - Experimenter: the reminder comes from the individual implementing the experiment. *Example: “The majority of our readers pay for their weekend newspaper.”*
 - External authority: the moral component is induced by an external authority, such as a university or a party. *Example: “I understand that this short survey falls under MIT’s honor system.”*
 - Religious authority: the moral component comes from a religious reference. *Example: Participants had to recall and write down the Ten Commandments.*
 - Other participants: the moral component comes from other participants or peers. *Example: “In 2015, 82% of all guests in room [room number] reused their towels.”*
 - Personal: the moral component comes from the participant. *Example: “Please describe an act where you helped someone else or did something for a good cause in your work or everyday life recently.”*
- The dummy variable **Selfish** equals one if the reminder aims at fostering selfishness or a lack of pro-sociality and zero if it reinforces a pro-social stance.

A.4.2 Target measure of pro-sociality

The inputted variable **Target1** aim to record the nature of the measure of pro-sociality targeted by the moral reminder. It is inspired by Ellemers, van der Toorn, Paunov, and van Leeuwen (2019)’s research themes.

- Behavior includes:
 - Actual. The measure of pro-sociality is based on actual behavior (that might or might not be incentivized). *Example: donation to charity, contribution to a public good.*
 - Self-reported. Participants are asked about their own past behavior, with no possibility to check whether they tell the truth (coded as Behavior otherwise, as intentions if the self-report is about future behavior). *Example: “How many times did you help people in need over the last month?”*

- Intention: Participants are asked about what they intend to do in a particular situation. *Example*: “Do you think you will help the next person in need you will meet after the end of the experiment”?
- Judgment includes
 - Antecedent. Personal view about what is the right thing to do or about whether what others do is right, which is at the source of moral behavior. *Example*: “What is your level of agreement with the statement ‘Everyone should contribute the maximum level to the public good’”.
 - Implications. Self-reflective statement about one’s own morality, i.e., the outcome from **self-observation** of one’s own behavior. *Example*: “On a scale from very moral to very immoral, how moral was your decision during the experiment?”; or emotional responses that characterize moral situations and used to diagnose the moral implications of different events. Typically refers to feelings of guilt and shame that people experience with regard to their own behavior, or outrage and disgust in response to the moral transgressions of others. *Example*: “To what extent do you feel angry / what is your relative culpability about your decisions during the experiment?”.

A.4.3 Field of application

The inputted variable **Application** records the sphere of social life to which the target refers.

- Decontextualized: Target behavior that is not specific to policy relevant domains.
- Tax: Compliance with taxation from public authorities.
- Finance: e.g. Debt repayment in the banking sector, voluntary payment of newspapers;
- Resources: If the target relates to the saving of any natural resource (water, energy, recycling, etc);
- Environment: Any kind of pro-environmental behavior other than natural resources saving (e.g., protecting endangered species, switching to durable goods or transportation modes, etc.);
- Social: groups applications to Health (e.g., Organ/blood donation), Politics (Voting behavior and choice of institutions; e.g. participation to an election, support for a particular policy) and Societal cohesion (individual’s support for values / principles / activities that promote societal cohesion; e.g., voluntary civil engagement, reduced discrimination against minority).

A.4.4 Measurement method

The inputted variables **Elicitation1** and **Elicitation2** keep track of the methodology used to elicit the measure of pro-sociality.

- **Elicitation1** = **Elicitation2** = Scale. Outcome measured on a scale (typically a Lickert scale, sometimes a score like when an IAT is used).
- **Elicitation1** = Experimental games. Specifically refers to elicitation methods that are lab-specific.
 - **Elicitation2** = Public good game [PGG]. Participants must decide how much resources they want to contribute to a joint project.

- **Elicitation2** = Common Dilemma [Common]. Participants must decide how much resources they want to take out from a joint project.
 - **Elicitation2** = Prisoner’s dilemma [PD].
 - **Elicitation2** = Tax Evasion Game [TEG]. Participants asked to freely report the amount of income that will be taxed.
 - **Elicitation2** = Dictator game [DG]. Participants asked whether they want to share some valuable resource (money/time/effort) with some other participant who is in turn passive (no direct return or response).
 - **Elicitation2** = Trust game [TG]. Participants send some payoff relevant quantity (typically, money) to other participant(s) who in turn decides what quantity to send back. Sometimes called “investment game”.
 - **Elicitation2** = Performance Reporting [PR]. Participants asked to self-report how they did in a task (what number they get from rolling a dice, how many matrices they solved, etc.).
 - **Elicitation2** = Others. If the game do not belong to the above classification.
- **Elicitation1** = **Elicitation2** = Field task. Any daily-life outcome. The typical example is voluntary paying for a newspaper or re-using towels in hotel rooms. This excludes lab-in-the-field experiments, classified within Experimental games.

A.4.5 Discipline

We group journals according to the field to which it belongs. To that end, we relied on the categories provided by Scimago. When this classification was set to ‘others’, we sampled papers from the journal and classified according to the department to which authors of papers published in the journal are affiliated. Table D provides the full classification of all journals in which included studies have been published.

A.5 Cross-validation using AI

To assess the robustness and internal consistency of the classification of subjective variables, a cross-validation procedure was implemented using *Le chat*, an open-source large language model developed by the French company Mistral AI. This procedure was not designed to achieve full agreement between human and AI classifications, but rather to evaluate the stability of the coding scheme and to identify potential ambiguities or borderline cases requiring further scrutiny. The AI was employed as a complementary instrument to the human annotation process, with the aim of highlighting points of interpretative uncertainty.

A.5.1 Cross-Validation Methodology

A standardized prompt was developed to guide the classification process across all included papers. It provided the AI with precise instructions, operational definitions for each target variable, and representative examples drawn from the dataset. This ensured that the model operated within a coherent interpretative framework aligned with that of the human coders. Convergence was defined as the assignment of the same main category by both the AI and the human coder. An exception was made for the *Source: Experimenter*

Table D: Classification of journals into disciplines

Journal	Discipline	K	$\widehat{\text{SMD}}$	Journal	Discipline	K	$\widehat{\text{SMD}}$
AEJ: Economic Policy	Economics	4	0.115	Journal of Political Economy	Economics	5	0.066
American Economic Journal	Economics	3	0.077	Journal of Public Economics	Economics	25	0.290
American Economic Review	Economics	3	0.023	Journal of the European Economic Association	Economics	9	1.813
American Society of Criminology	Others	4	0.007	Judgment and Decision Making	Economics	7	0.250
Australian Journal of Psychology	Psychology	1	0.081	Kidney International Reports	Others	3	0.000
Basic and Applied Social Psychology	Psychology	26	0.024	Law & Society Review	Others	2	-0.196
Contemporary Accounting Research	Business	1	0.068	Management Science	Business	5	0.031
Ecological Economics	Economics	4	-0.082	Managerial and Decision Economics	Economics	2	0.460
Economics Letters	Economics	3	0.476	National Tax Journal	Economics	4	0.003
Environment and Behavior	Others	2	3.473	Organizational Behavior and Human Decision Processes	Psychology	1	-0.423
Environmental Science and Policy	Others	3	0.148	PLoS ONE	Social Sciences	8	0.196
Ethics & Behavior	Psychology	4	0.193	Personality and Social Psychology Bulletin	Psychology	17	0.429
European Economic Review	Economics	2	0.553	Philosophy, Politics and Economics	Social Sciences	1	0.290
FinanzArchiv / Public Finance Analysis	Economics	3	-0.056	Political Behavior	Others	2	0.131
Frontiers in Psychology	Psychology	5	0.151	Psychological Science	Psychology	16	0.046
Group Processes & Intergroup Relations	Others	3	0.315	Resource and Energy Economics	Economics	7	-0.029
International Studies Quarterly	Social Sciences	2	0.589	Review of Economics and Statistics	Economics	1	0.036
Journal of Applied Social Psychology	Psychology	11	0.599	Scientific reports	Social Sciences	5	0.283
Journal of Behavioral and Experimental Economics	Economics	17	0.035	Social Influence	Psychology	5	-0.142
Journal of Consumer Research	Business	4	-0.008	Social Psychological and Personality Science	Psychology	2	-0.869
Journal of Economic Behavior & Organization	Economics	2	0.010	Social Psychology	Psychology	4	0.060
Journal of Economic Behavior and Organization	Economics	13	0.106	Social Science Quarterly	Social Sciences	6	0.352
Journal of Environmental Psychology	Psychology	67	0.276	Social science research	Social Sciences	3	0.020
Journal of Experimental Social Psychology	Psychology	16	0.155	Southern Economic Journal	Economics	6	0.743
Journal of Interpersonal Violence	Psychology	10	0.115	The Journal of Positive Psychology	Psychology	6	1.056
Journal of Marketing	Business	4	0.582	The Journal of Social Psychology	Psychology	2	0.024
Journal of Marketing Research	Business	3	0.062	The University of Chicago Law Review	Others	3	0.152
Journal of Personality and Social Psychology	Psychology	26	0.488	Theory and Decision	Economics	12	0.150

Note. Classification of journals into disciplines along with the number of treatment effects (K) and the (non-weighted) average effect size ($\widehat{\text{SMD}}$).

variable, for which the subcategories *Experimenter (general)* and *Experimenter (specific)* were treated as functionally equivalent.

The first round of cross-validation covered four subjective dimensions: the class of moral reminder, its source, the elicitation method, and the identification of distinct studies within each paper. For each dimension and paper, it was recorded whether the AI’s classification matched the human coding, and the specific label returned by the AI was documented. In this initial round, a convergence rate of approximately 70% was observed for the classification of reminder tools (see descriptive statistics in Table E).

In cases of divergence on the classification of moral reminder types, a second validation round was conducted. The prompt was revised to include additional definitions and clarifications concerning previously ambiguous subcategories. This refinement resulted in an increased convergence rate of approximately 74%.

All remaining discrepancies were subsequently reviewed jointly by human coders. These cases were reassessed in light of both the AI output and the updated classification criteria. This iterative process ultimately led to a final convergence rate of approximately 76%, indicating a high level of reliability and replicability in the classification scheme. The remaining 24% of disagreements correspond to instances where the AI returned a different label than the one assigned by human coders. In these cases, the human classification was retained, reflecting the fact that the AI’s output was judged to be inconsistent with the predefined taxonomic criteria.

Table E:

	<i>N</i> (non-missing)	Convergences (#)	Convergence (%)
<i>Study breakdown</i>	549	529	96.36
<i>Source</i>	554	280	50.54
<i>Elicitation 1</i>	547	390	71.21
<i>Elicitation 2</i>	549	335	61.02
<i>Application</i>	545	460	84.40
<i>Target</i>	544	370	68.02
<i>Field</i>	557	480	86.18
<i>Reminder tool</i>	561	425	75.76

Note. The first column counts the number of studies (out of 564) for which the AI returned an answer. The second and third column provide descriptive statistics on studies with a perfect match between the initial and AI classification.

A.5.2 Final prompt provided to the AI

The following is the full structured prompt submitted to the Mistral AI model during the final round of the cross-validation process. It was designed to enforce consistent extraction across papers and to strictly constrain responses within the defined taxonomic framework.

PRIMARY TASK Scientific Paper Analysis: Moral Reminder Experiments

Analyze the provided scientific paper and extract specific information about moral reminder experiments using **only** the taxonomic definitions provided below.

CRITICAL REQUIREMENTS Multi-Study Analysis Protocol

MANDATORY: This paper may contain multiple studies, experiments, treatments, or subgroups. You **MUST:**

1. Identify **ALL** separate studies/experiments in the paper.
2. For each study: Identify all different treatments, conditions, or subgroups.
3. Provide separate complete analysis for **EACH** individual study **AND** for **EACH** treatment / condition / subgroup within that study.
4. Label each analysis clearly with study number/name and treatment/condition/subgroup.

Taxonomic Framework (Use These Definitions Only)

1. Reminder Classes (Content categories)

- **Consequences** – Provides information on benefits (resp. costs) of prosocial (resp. antisocial) action.

- Example: Hafner (2019) “The carbon dioxide (CO₂) emissions from the standard gas boiler are approximately 210g per kilowatt-hour. Whilst the CO₂ emissions from the energy efficient heat pump are approximately 140g per kilowatt-hour. This means if you select the energy efficient system your CO₂ emissions will be reduced by one third.”
- **Priming** includes:
 - **Framing** – Frames the decision environment to trigger pro-social behavior.
 - * Example: Rege (2002) “The 10 subjects are referred to as a ‘community’. The box is referred to as ‘the community box’. The money in the box is said to belong to the community. The group envelope is marked ‘the community’, whereas the personal envelope is marked ‘mine’. Finally, the strategy choice is presented as follows: each person can choose either to be a free-rider by taking money from the community box, or to be a cooperator by not taking money from the community box.”
 - **Memory (Priming with memories)** – Primes moral feelings by triggering relevant memories.
 - * Example: Bruegger (2019) “Participants were asked to recall an environmentally friendly behavior carried out during the past week.”
 - **Priming (Priming with associations)** – Primes moral feelings using associations/imagination.
 - * Examples:
 - Khan (2006) “In the first decision, we asked them to imagine that they received an income tax rebate of \$500 and were thinking of donating \$100 of this money to a charity organization. We provided participants with a short description of the two organizations.”
 - Niesten (2017) “Participants were shown five words encompassing moral traits.”
- **Descriptive norm** – Information on how most people behave.
 - Example: Brudermann (2014) “The majority of our readers pay for their weekend newspaper.”
 - Types:
 - * *General*: Statistical information concerns entire population
 - * *Specific*: Statistical information concerns subgroup, particularly prior experiment participants
- **Injunctive norm** – Information on what other people think should be done. Includes experimenter statements where principle source isn’t stated but implicitly widely accepted.
 - Example: “Most participants feel that giving the other participants the positive consequences task...”
- **Commitment** – Take prior actions that change the cost of taking the socially virtuous action. Oath is a subcategory.
- **Oath** – Engagement through swearing and oath.
 - Example: Jacquemet et al. (2019) “[I] swear upon my honor that, during the whole experiment, I will tell the truth and always provide honest answers”

2. Source (Who Emits the Reminder)

- **Experimenter** – Explicit expectations stated by the experimenter
- **External authority** – From external authority such as university or institution;
 - Example: “This survey falls under MIT’s honor system”
- **Religious authority** – From religious figure or literature;
 - Example: The Ten Commandments
- **Other participants** – From other participants or peers;
 - Example: “In 2015, 82% of all guests in room [room number] reused their towels”
- **Personal** – From the participant themselves;
 - Example: “Please describe an act where you helped someone else...”

3. Additional Characteristics

- **Selfish** – Equals one if reminder aims at fostering selfishness or lack of pro-sociality; Equals zero if it reinforces pro-social stance

4. Target Measure of Pro-sociality

- **Behavior** includes:
 - *Actual*: Measure based on actual behavior (may/may not be incentivized)
 - * Example: donation to charity, contribution to public good
 - *Self-reported*: Participants asked about own past behavior, no truth verification possible
 - * Example: “How many times did you help people in need over the last month?”
- **Intention** – Participants asked about intended future behavior
 - Example: “Do you think you will help the next person in need you will meet after the end of the experiment?”
- **Judgment** includes:
 - *Antecedent*: Personal view about right thing to do or whether others’ actions are right (source of moral behavior)
 - * Example: “What is your level of agreement with the statement ‘Everyone should contribute the maximum level to the public good’”
 - *Implications*: Self-reflective statement about own morality (outcome from self-observation of behavior)
 - * Example: “On a scale from very moral to very immoral, how moral was your decision during the experiment?”

5. Elicitation Method

Elicitation1:

- **Scale** – Outcome measured on scale (typically Likert scale, sometimes score like IAT)
- **Experimental games** – Lab-specific elicitation methods
- **Field behavior** – Any daily-life outcome (excluding lab-in-the-field experiments)

Elicitation2 (if Experimental games):

- **Public Good Game (PGG)** – Participants contribute resources to joint project
- **Common Dilemma** – Participants take resources from joint project
- **Prisoner's Dilemma (PD)**
- **Tax Evasion Game (TEG)** – Participants freely report income to be taxed
- **Dictator Game (DG)** – Participants share resources with passive recipients
- **Trust Game (TG)** – Participants send resources that may be returned
- **Performance Reporting (PR)** – Participants self-report performance on task
- **Others** – Specify if not in above categories

6. Field of Application

- **Decontextualized** – Target behavior not specific to policy relevant domain
- **Tax** – Compliance with taxation from public authorities
- **Finance** – Debt repayment in banking sector, voluntary newspaper payment
- **Resources** – Target relates to saving natural resources (water, energy, recycling, etc.)
- **Environment** – Pro-environmental behavior other than natural resource saving
- **Health** – Organ/blood donation
- **Politics** – Voting behavior and choice of institutions
- **Societal cohesion** – Individual support for values/principles/activities promoting societal cohesion

Required Output Format

STUDY 1: [Study Name/Description]

- **Treatment/Condition A:** [Name]
 - Sample size: [N]
 - Page reference: [page numbers]
 - Reminder Classes: [Category with direct quote]

- Source: [Category]
- Selfish: [Yes/No]
- Target Measure: [Target1]
- Elicitation Method: [Elicitation1]
- Field of Application: [Application]

- **Treatment/Condition B:** [Name]
(*Same format as above*)

STUDY 2: [Study Name/Description]

- **Treatment/Condition A:** [Name]
(*Same format as above*)

[Continue for all studies and treatments]

Quality Requirements

Essential Elements:

- Include **direct quotes** of **ALL** moral reminders used in each treatment/condition
- Reference specific sections or **page numbers** where information was found
- If information cannot be determined: state "**Not specified in the paper**"
- If study has no moral reminder (e.g., control group): state "**No moral reminder - Control condition**"
- Pay attention to variations in reminder content, source, or method between different treatments within same study
- Use **exact terminology and categories** from the taxonomic framework above

B Methods

B.1 Measure of Effect sizes

For each treatment effect, $k = 1, \dots, K$, the data provides the average outcome in the treatment (\hat{y}_{kT}) and control (\hat{y}_{kC}) groups, their standard deviations ($\hat{\sigma}_{kC}^2, \hat{\sigma}_{kT}^2$) and the number of observations in each group ($N_k = N_{kT} + N_{kC}$). We compute effect sizes using the Standardized Mean Deviation (SMD) provided by Hedge's g , which corrects the small sample bias in the Cohen's d (see, e.g., Schwarzer, Carpenter, and Rücker, 2015):

$$\hat{\theta}_k = \left(1 - \frac{3}{4N_k}\right) \frac{\hat{y}_{kT} - \hat{y}_{kC}}{\sqrt{[(N_{kT} - 1)\hat{\sigma}_{kT}^2 + (N_{kC} - 1)\hat{\sigma}_{kC}^2] / (N_k - 2)}}$$

whose standard deviation is given by:

$$\hat{\sigma}_k^2 = \frac{N_k}{N_{kT} \cdot N_{kC}} + \frac{\hat{\theta}_k^2}{2(N_k - 3.94)}$$

All heterogeneity analyses are run using a random effect model, which allows for unobserved heterogeneity across studies. For any set of p variables describing observed heterogeneity, \mathbf{x}_k , and as many unknown parameters to be estimated, \mathbf{b} , the meta-analytic regression model is specified as:

$$\hat{\theta}_k = \theta_0 + \mathbf{x}_k \mathbf{b} + u_k + \varepsilon_k, \forall k, \quad \text{with } \varepsilon \approx \mathcal{N}(0, \hat{\sigma}_k^2), u_k \approx \mathcal{N}(0, \tau^2)$$

where $\hat{\tau}$ is the between study variance in the treatment-effect specific true effect, $\theta_k = \bar{\theta} + u_k$. Formally speaking, this model is a linear model with heteroskedastic standard errors due to the treatment-effect specific variance in the estimate. It is estimated using a GLS estimator with weights $\tilde{w}_k = 1/(\hat{\sigma}_k^2 + \hat{\tau}^2)$.

The proportion of true variance in the effect is measured by:

$$I^2 = \frac{\hat{\tau}^2}{\hat{\sigma}_k^2 + \hat{\tau}^2}$$

B.2 Normalization of the outcome variable

Given the heterogeneity of outcome measures across studies (e.g., monetary amounts, Likert scores, probabilities), the take-up analysis provided in Figure 2 requires to normalize the outcome variable across studies. To that end, we restrict to studies whose outcome is defined on a bounded scale, and rely on a Min-Max normalization of the outcome (\hat{y}_{kG}) in both the control ($G = C$) and the treatment group ($G = T$):

$$\hat{y}_{kG}^{norm} = \frac{\hat{y}_{kG}^{norm} - y_{\min}}{y_{\max} - y_{\min}}$$

C Additional evidence

C.1 Descriptive statistics

Table F: Descriptive statistics on studies heterogeneity

	Priming	Consequences	Descriptive norm	Injunctive norm	Overall
	21.0%	14.5%	32.5%	32.0%	100%
Discipline (%)					
Business	9.2	3.3	5.2	0.0	4.1
Economics	14.9	26.7	27.4	49.6	31.8
Psychology	70.1	40.0	62.2	37.6	52.8
Social Sciences	1.1	15.0	3.0	8.3	6.0
Others	4.6	15.0	2.2	4.5	5.3
Country (% , missing otherwise)					
Asia	2.3	6.7	0.0	8.3	4.1
Europe	31.0	18.3	41.5	20.3	29.2
North-America & Australia	64.4	56.7	54.1	66.9	60.7
Rest of the world	1.1	0.0	0.0	0.8	0.5
Missing	1.1	18.3	4.4	3.8	5.5
Year of publication (%)					
1982	0.0	0.0	0.0	1.5	0.5
1990's	0.0	13.3	8.9	13.1	9.0
2000-2005	10.3	3.3	1.5	0.8	3.4
2006-2010	16.1	6.7	5.2	2.3	6.8
2011-2015	35.6	28.3	43.0	50.0	41.5
2016-2020	37.9	48.3	41.5	32.3	38.8
Measure of the outcome					
Experimental game	60.9	13.3	32.6	44.4	39.5
Field task	11.5	53.3	45.2	37.6	36.9
Scale	27.6	33.3	22.2	18.0	23.6
Measure of the treatment effect					
Control	34.5	60.0	57.0	63.9	54.9
Placebo	65.5	40.0	43.0	36.1	45.1

Note. Descriptive statistics on the sample's distribution over the variables in row (all variables are defined in the SI, Sections A.3 and A.4).

Table G: Distribution of the sample over methodological moderators

	Priming	Consequences	Descriptive norm	Injunctive norm	Overall
	21.0%	14.5%	32.5%	32.0%	100%
Elicitation method					
Field experiment	13.8	51.7	43.0	37.6	36.4
Class experiment	5.7	0.0	10.4	3.8	5.8
Online experiment	20.7	38.3	13.3	17.3	19.8
Laboratory experiment	59.8	10.0	33.3	41.4	38.1
Incentives					
Incentivised	50.6	21.7	27.4	47.4	37.8
Target measure of pro-sociality					
Intentions	20.7	0.0	11.1	1.5	8.4
Behavior	70.1	66.7	72.6	80.5	73.7
Judgment	9.2	33.3	16.3	18.0	17.8

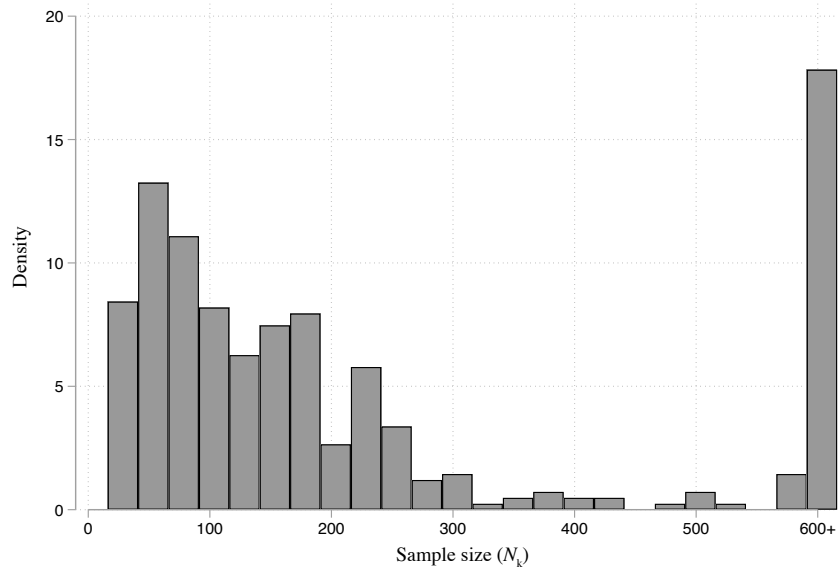
Note. Descriptive statistics on the sample's distribution over the variables in row (all variables are defined in the SI, Sections A.3 and A.4).

Table H: Distribution of the sample over implementation moderators

	Priming	Consequences	Descriptive norm	Injunctive norm	Overall
	21.0%	14.5%	32.5%	32.0%	100%
Source of the reminder					
Experimenter	55.2	50.0	59.3	71.4	61.0
External authority	9.2	50.0	37.0	13.5	25.5
Other participants	1.1	0.0	3.7	3.8	2.7
Personal	32.2	0.0	0.0	7.5	9.2
Religious authority	2.3	0.0	0.0	3.8	1.7
Application (%)					
Decontextualized	48.3	3.3	5.9	39.1	25.1
Social	26.4	11.7	7.4	13.5	14.0
Environment	14.9	45.0	54.1	10.5	30.6
Ressources	3.4	18.3	14.8	9.0	11.1
Tax	6.9	18.3	8.9	9.0	9.9
Finance	0.0	3.3	8.9	18.8	9.4
Observers (%)					
Observer	12.6	1.7	0.0	27.8	11.8

Note. Descriptive statistics on the sample's distribution over the variables in row (all variables are defined in the SI, Sections A.3 and A.4).

Figure B: Density of sample sizes among included effect sizes

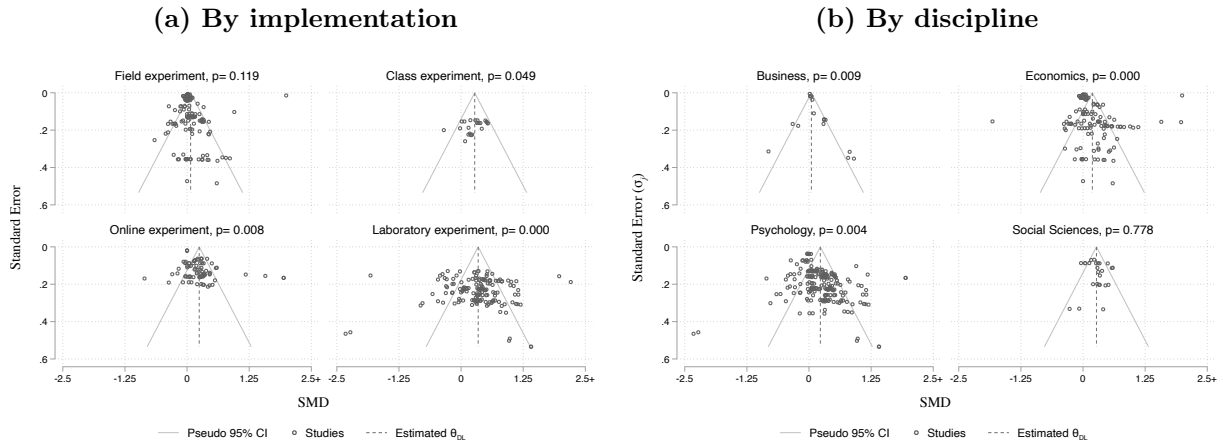


Note. Empirical distribution of the sample size.

C.2 Distribution of effect sizes

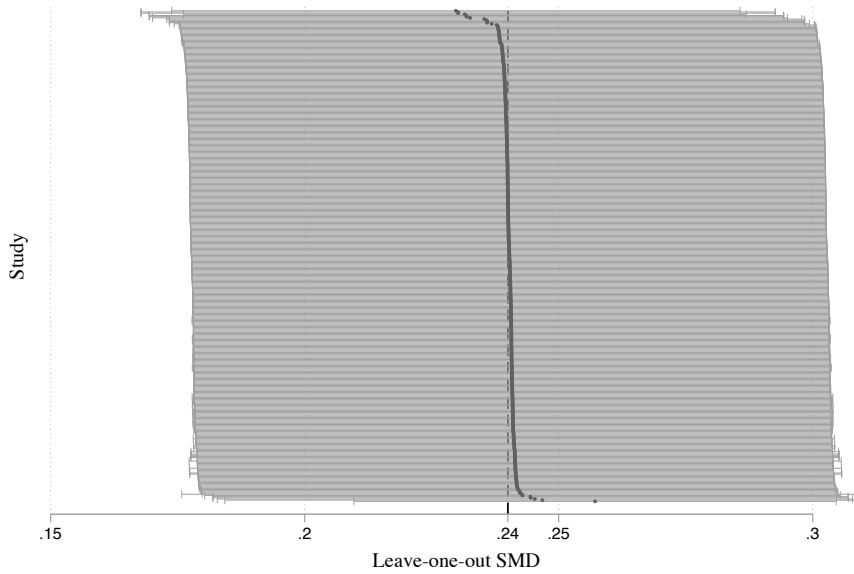
Figure C displays the funnel plots and Egger statistics of small sample bias on the subsample of studies split according to the implementation of the experiment (Panel a) and the discipline (Panel b). Figures D and E display the outcomes from leave-one-out exercises at the effect size and study levels.

Figure C: Funnel plots, by implementation and discipline



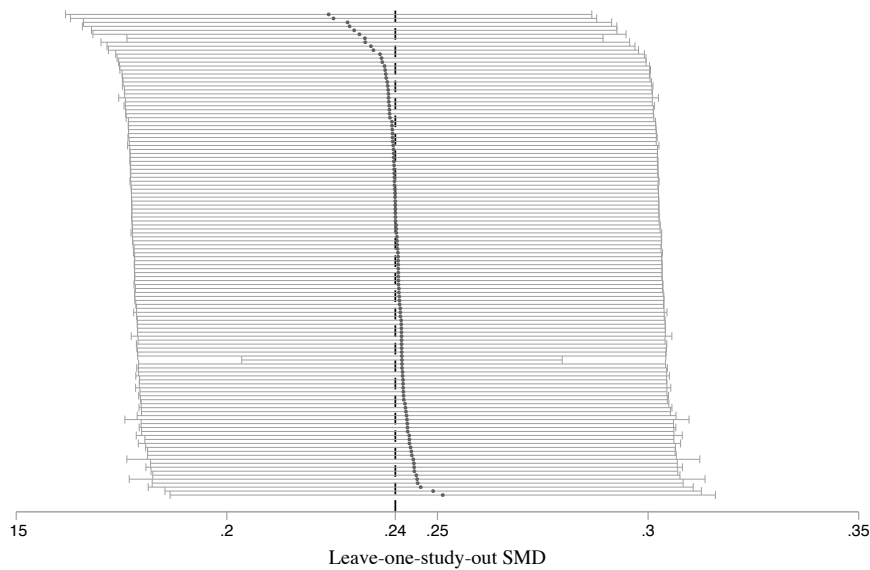
Note. Funnel plots and p-values of the Egger statistics of small sample bias (i.e., statistical significance of the slope of the regression of effect sizes on standard errors Egger, Smith, Schneider, and Minder, 1997) on the subsample of studies split according to the implementation of the experiment (Panel a) and the discipline (Panel b).

Figure D: Leave-one-out meta-analysis



Note. Each row of the leave-one-out forest plot omits a study sequentially. Estimates corresponds to the average effect size and 95% confidence interval estimated on $K - 1$ studies. Hedge's g values range from 0.23 to 0.26. All confidence intervals contain the overall effect size.

Figure E: Leave-one-study-out meta-analysis



Note. Each row of the leave-one-out forest plot omits a study sequentially. Estimates corresponds to the average effect size and 95% confidence interval estimated on nearly all treatment effects, excluding effects from a given study. Hedge's g values range from 0.21 to 0.25. All confidence intervals contain the overall effect size.

C.3 Additional regression results

Table I: Multivariate analysis (Table 5): Full results

	(1)	(2)	(3)	(4)
Classes of moral reminder (Ref. Priming)				
Consequences	0.075 (0.529)	0.164 (0.214)	0.325** (0.027)	0.436** (0.017)
Descriptive norm	-0.055 (0.572)	-0.053 (0.609)	0.057 (0.606)	0.148 (0.299)
Injunctive norm	-0.084 (0.395)	-0.072 (0.518)	-0.040 (0.734)	-0.026 (0.849)
Elicitation method (Ref. Field experiment)				
Class experiment			0.017 (0.933)	-0.065 (0.770)
Online experiment			0.031 (0.817)	-0.106 (0.493)
Laboratory experiment			0.257** (0.014)	0.050 (0.698)
Incentivized			0.294*** (0.003)	0.319*** (0.005)
Target measure of pro-sociality (Ref. Intentions)				
Behavior			0.189 (0.187)	0.092 (0.562)
Judgment			0.297* (0.077)	0.193 (0.321)
Source of the reminder (Ref. Experimenter)				
External authority				0.034 (0.753)
Other participants				0.064 (0.792)
Personal				-0.141 (0.369)
Religious authority				0.301 (0.355)
Application (Ref. Decontextualized)				
Social				0.163 (0.306)
Environment				-0.175 (0.274)
Ressources				-0.242 (0.175)
Tax				-0.381** (0.037)
Finance				-0.363* (0.090)
Observer				0.302* (0.085)
Observations	415	415	415	415

Note. Random effects meta-regression (see the SI, Section B.1, for a formal presentation of the model) of the average treatment effect. All models include controls (omitted from the table; the results are available from the authors upon request) for the country where the study took place and the discipline to which the journal is affiliated. *Legend:* * $p < .10$, ** $p < .05$, *** $p < .01$.

Table J: Replication of Table 5 with a multi-level meta regression

	(1)	(2)	(3)	(4)
Classes of moral reminder (Ref. Priming)				
Consequences	0.399 (0.248)	0.341 (0.324)	0.339 (0.328)	0.008 (0.763)
Descriptive norm	0.055 (0.726)	-0.003 (0.985)	0.006 (0.970)	-0.004 (0.881)
Injunctive norm	0.021 (0.843)	-0.010 (0.935)	-0.016 (0.903)	0.041 (0.751)
Elicitation method (Ref. Field experiment)				
Class experiment			0.019 (0.929)	0.022 (0.941)
Online experiment			0.093 (0.528)	0.001 (0.997)
Laboratory experiment			0.061 (0.658)	-0.029 (0.879)
Incentivized			0.063 (0.524)	0.046 (0.725)
Target measure of pro-sociality (Ref. Intentions)				
Behavior			-0.025 (0.722)	-0.056 (0.454)
Judgment			0.243*** (0.000)	0.247*** (0.000)
Source of the reminder (Ref. Experimenter)				
External authority				0.011 (0.618)
Other participants				-0.244 (0.572)
Personal				-0.034 (0.870)
Religious authority				0.025 (0.404)
Application (Ref. Decontextualized)				
Social				0.027 (0.930)
Environment				0.020 (0.944)
Ressources				-0.157 (0.621)
Tax				-0.488 (0.186)
Finance				-0.344 (0.504)
Observer				0.488* (0.086)
Observations	415	415	415	415

Note. Multi-level meta regression of the average treatment effect nested at the paper and the study level. All models include controls (omitted from the table; the results are available from the authors upon request) for the country where the study took place and the discipline to which the journal is affiliated. *Legend:* * $p < .10$, ** $p < .05$, *** $p < .01$.

Table K: Predictors of a (significantly) negative effect (Table 6): Full results

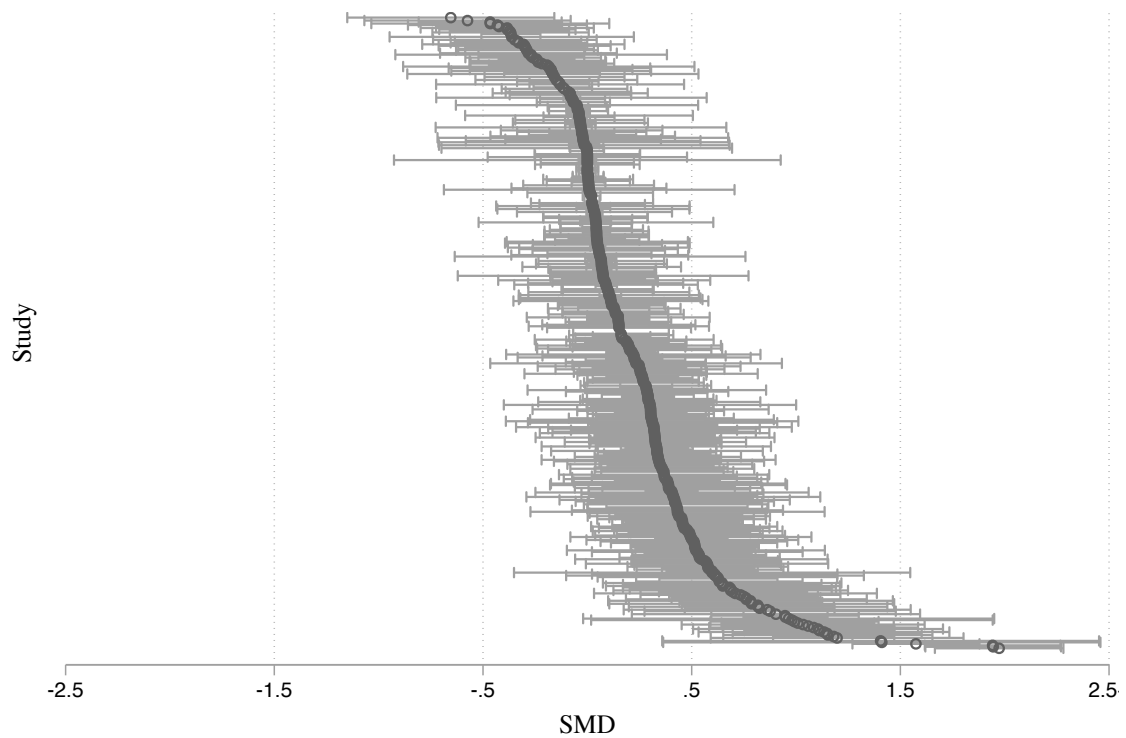
	(1)	(2)	(3)	(4)
Classes of moral reminder (Ref. Priming)				
Consequences	-0.037 (0.203)	-0.025 (0.437)	-0.038 (0.221)	-0.025 (0.595)
Descriptive norm	-0.002 (0.959)	0.002 (0.952)	-0.004 (0.901)	0.001 (0.975)
Injunctive norm	0.030 (0.371)	0.042 (0.245)	0.038 (0.216)	0.058* (0.090)
Elicitation method (Ref. Field experiment)				
Class experiment			-0.068* (0.085)	-0.030 (0.520)
Online experiment			-0.040 (0.242)	0.001 (0.979)
Laboratory experiment			0.017 (0.650)	0.069 (0.179)
Incentivized			-0.047* (0.087)	-0.051 (0.131)
Target measure of pro-sociality (Ref. Intentions)				
Behavior			-0.089* (0.082)	-0.084 (0.119)
Judgment			-0.028 (0.657)	-0.040 (0.528)
Source of the reminder (Ref. Experimenter)				
External authority				0.003 (0.915)
Other participants				0.140 (0.230)
Personal				0.053 (0.232)
Religious authority				-0.074 (0.208)
Application (Ref. Decontextualized)				
Social				0.004 (0.932)
Environment				0.007 (0.870)
Ressources				0.037 (0.494)
Tax				0.134** (0.047)
Finance				0.070 (0.341)
Observer				-0.079* (0.052)
Observations	415	415	415	415

Note. Linear regressions of the dummy variable indicating whether a study identifies en significantly negative effect. Robust standard in parenthesis. All models include controls (ommitted from the table; the results are available from the authors upon request) for the country where the study took place and the discipline to which the journal is affiliated. *Legend:* * $p < .10$, ** $p < .05$, *** $p < .01$.

D Robustness checks

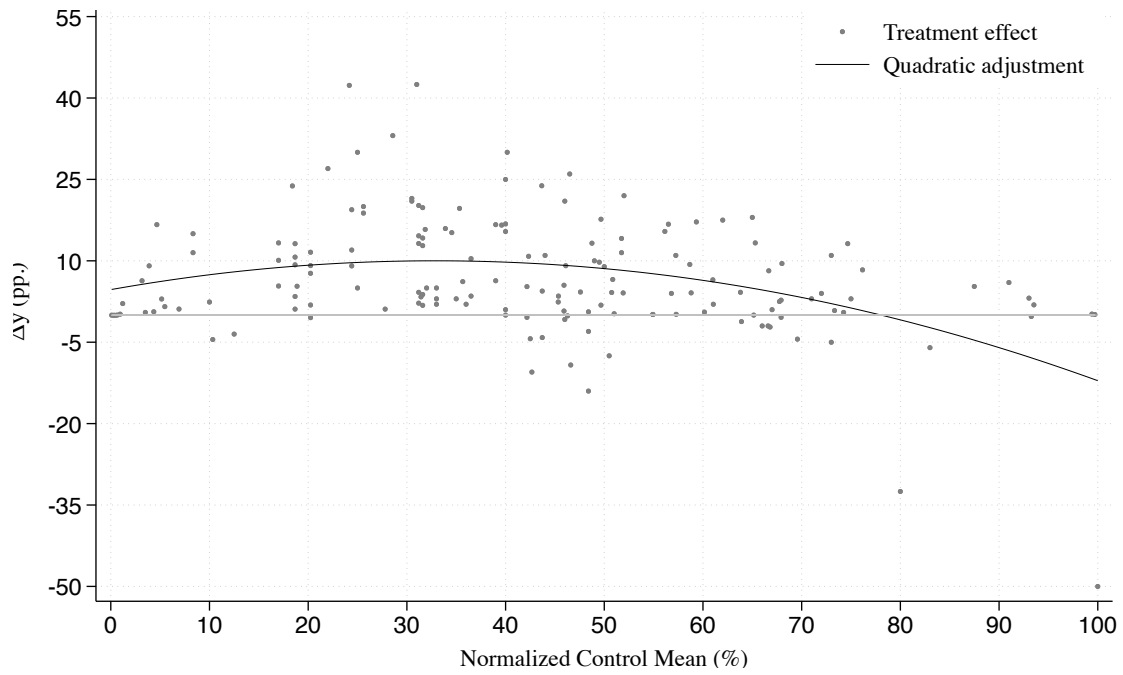
Appendix D shows that the main results hold when excluding outliers in the top and bottom 2% of the distribution. Effect sizes included in the trimmed sample range from -0.65 to 1.97 of a standard deviation and are plotted in figure F. We find an average $\widehat{\text{SMD}}$ of $0.16 \in [0.14, 0.17]$, $p < .001$, as large positive outliers reflect the publication bias towards positive results discussed in the main analysis.

Figure F: Replication of Figure 1 on trimmed data



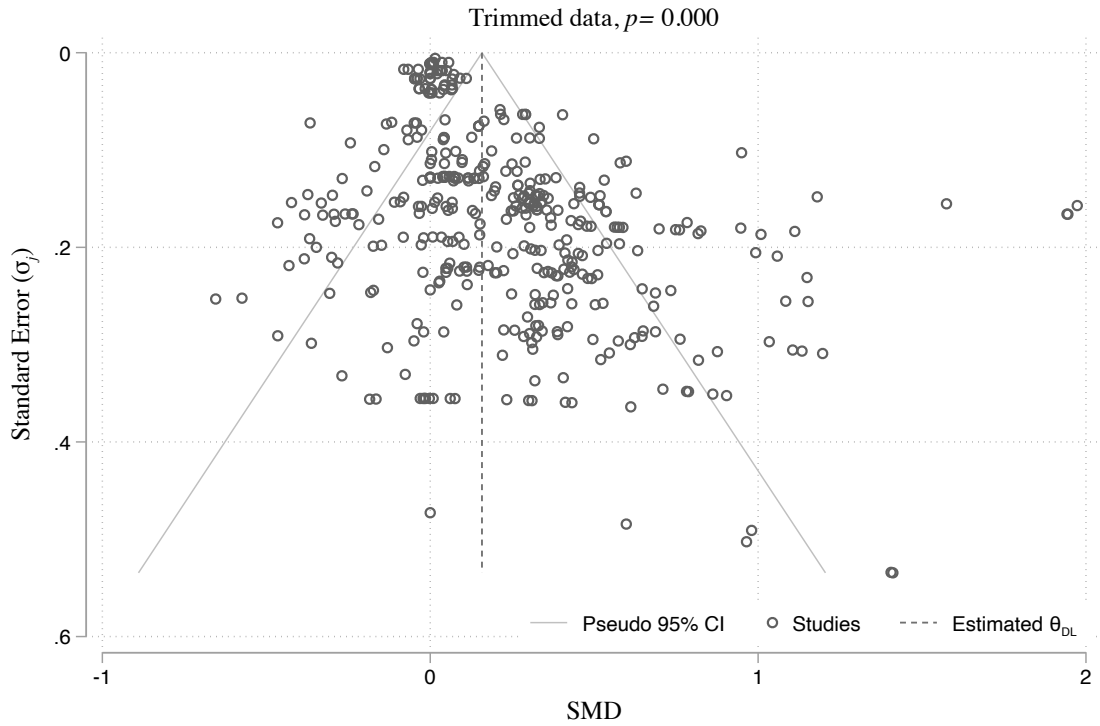
Note. Forest plot of effect sizes in the trimmed sample excluding the top and bottom 2% of effect sizes ($K = 399$) with their corresponding 95% confidence intervals. Extracted Hedge's g values ranged from -0.67 to 1.97 standard deviations. $\widehat{\text{SMD}} = 0.16$, 95% CI: $[0.14, 0.17]$, $p < .001$.

Figure G: Replication of Figure 2 on trimmed data



Note. Plot of the treatment effects on the treated groups in the trimmed data relative to the control group (difference in percentage points), as a function of the average pro-sociality in the control group ($N = 166$, see the Appendix, Section B.2). The trimmed sample excludes the top and bottom 2% of effect sizes. The control mean is normalized between 0 and 100%, where 100% is the most prosocial outcome feasible. The maximum is not systematically available, either because it was not reported in the paper or because the outcome had no limit, e.g., in charity donation or energy saving tasks.

Figure H: Replication of Figure 3 on trimmed data



Note. The funnel plot shows the average treatment effect and standard error of each study to test for small-study (or publication) bias, along with the p-value of the Egger statistic for small sample bias (i.e., statistical significance of the slope of the regression of effect sizes on standard errors Egger, Smith, Schneider, and Minder, 1997). The dotted lines show the pseudo 95% confidence interval of the overall treatment effect ($\widehat{sMD} = 0.16$) as the standard error varies. Under no small-study bias, we expect to see studies distributed symmetrically around the overall treatment effect, and closer to the \widehat{sMD} when precision is high. The trimmed sample excludes the top and bottom 2% of effect sizes.

Table L: Replication of Table 2 on trimmed data

Study	K		Hedges's g with 95% CI	p-value
Elicitation method				
Field experiment	147		0.02 [0.01, 0.03]	0.002
Class experiment	24		0.27 [0.19, 0.36]	0.000
Online experiment	81		0.26 [0.20, 0.31]	0.000
Laboratory experiment	147		0.38 [0.31, 0.45]	0.000
Test of group differences: $Q_b(3) = 181.13$, $p = 0.00$				
Incentives				
	248		0.09 [0.07, 0.11]	0.000
	151		0.30 [0.25, 0.35]	0.000
Test of group differences: $Q_b(1) = 61.88$, $p = 0.00$				
Target measure of pro-sociality				
Intentions	34		0.04 [-0.00, 0.09]	0.066
Behavior	293		0.12 [0.11, 0.14]	0.000
Judgment	72		0.36 [0.27, 0.45]	0.000
Test of group differences: $Q_b(2) = 37.47$, $p = 0.00$				
Overall			0.16 [0.14, 0.17]	0.000
		0 .5		

Note. Effect sizes (\widehat{SMD}) estimated on trimmed data excluding the top and bottom 2% of effect sizes across the categories listed in each row. The second column on the left corresponds to the number of effect sizes in each category. The columns on the right provide the standardised mean difference, the 95% confidence interval and the p-value at the 5% significance level.

Table M: Replication of Table 3 on trimmed data

Study	K		Hedges's g with 95% CI	p-value
Source of the reminder				
Experimenter	244		0.18 [0.16, 0.21]	0.000
External authority	101		0.09 [0.06, 0.12]	0.000
Other participants	11		0.14 [-0.10, 0.38]	0.264
Personal	36		0.22 [0.15, 0.29]	0.000
Religious authority	7		0.11 [0.02, 0.20]	0.014
Test of group differences: $Q_b(4) = 32.21, p = 0.00$				
Applications				
Decontextualized	100		0.34 [0.28, 0.40]	0.000
Social	57		0.33 [0.25, 0.40]	0.000
Environment	123		0.14 [0.10, 0.18]	0.000
Ressources	44		0.04 [0.01, 0.06]	0.008
Tax	39		0.04 [0.01, 0.07]	0.008
Finance	36		0.05 [0.00, 0.10]	0.036
Test of group differences: $Q_b(5) = 141.79, p = 0.00$				
Observers				
No observer	351		0.12 [0.11, 0.14]	0.000
Observer	48		0.51 [0.43, 0.59]	0.000
Test of group differences: $Q_b(1) = 83.22, p = 0.00$				
Overall			0.16 [0.14, 0.17]	0.000

Note. Effect sizes (\widehat{SMD}) estimated on trimmed data excluding the top and bottom 2% of effect sizes across the categories listed in each row. The second column on the left corresponds to the number of effect sizes in each category. The columns on the right provide the standardised mean difference, the 95% confidence interval and the p-value at the 5% significance level.

Table N: Replication of Table 4 on trimmed data

Study	K		Hedges's g with 95% CI	p-value
Priming	84		0.31 [0.24, 0.38]	0.000
Consequences	54		0.04 [0.02, 0.06]	0.000
Descriptive norm	131		0.13 [0.10, 0.16]	0.000
Injunctive norm	130		0.21 [0.17, 0.25]	0.000
Test of group differences: $Q_b(3) = 104.98, p = 0.00$				
Overall			0.16 [0.14, 0.17]	0.000

0 .5

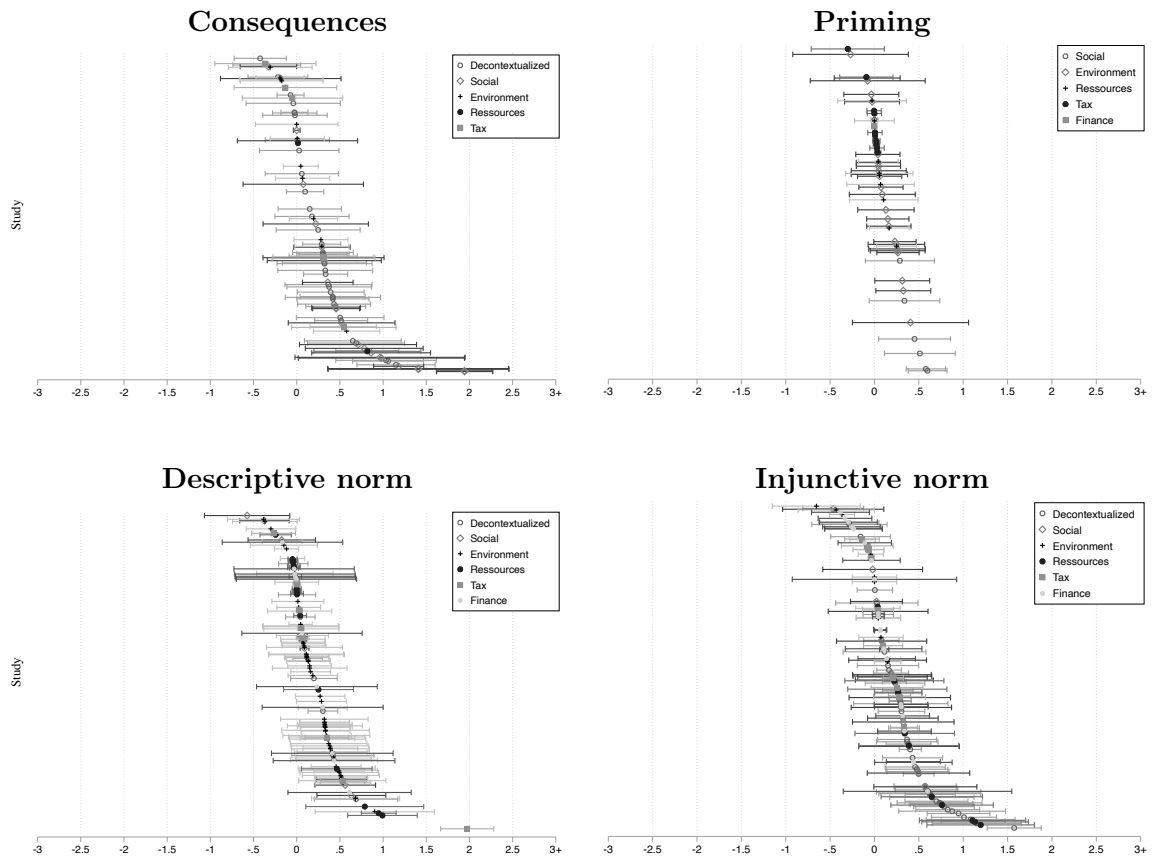
Note. Effect sizes (\widehat{SMD}) estimated on trimmed data excluding the top and bottom 2% of effect sizes across the categories listed in each row (see Table 1 for formal definitions). The second column on the left corresponds to the number of effect sizes in each category. The columns on the right provide the standardised mean difference, the 95% confidence interval and the p-value at the 5% significance level.

Table O: Replication of Table 5 on trimmed data

	(1)	(2)	(3)	(4)
Consequences	-0.189*** (0.000)	-0.178*** (0.000)	-0.098*** (0.006)	-0.116*** (0.006)
Descriptive norm	-0.141*** (0.000)	-0.160*** (0.000)	-0.062** (0.035)	-0.069* (0.050)
Injunctive norm	-0.085*** (0.004)	-0.093*** (0.004)	-0.055* (0.087)	-0.099*** (0.007)
Observer				0.240*** (0.000)
Observations	399	399	399	399

Note. Meta regression of the average treatment effect estimated on trimmed data excluding the top and bottom 2% of effect sizes (see the SI, Section B.1, for a formal presentation of the model). All models include controls (omitted from the table; the results are available from the authors upon request) for the country where the study took place and the discipline to which the journal is affiliated. *Legend:* * $p < .10$, ** $p < .05$, *** $p < .01$.

Figure I: Replication of Figure 4 on trimmed data



Note. Forest plot of effect sizes in the trimmed sample by tools (refer to Table N for the average effect sizes), distinguishing the applications. The trimmed sample excludes the top and bottom 2% of effect sizes.

Table P: Replication of Table 6 on trimmed data

	(1)	(2)	(3)	(4)
Consequences	-0.030 (0.156)	0.001 (0.979)	-0.032 (0.226)	-0.033 (0.453)
Descriptive norm	0.003 (0.903)	0.006 (0.825)	-0.020 (0.579)	-0.019 (0.654)
Injunctive norm	0.076* (0.095)	0.104** (0.047)	0.087** (0.023)	0.108** (0.012)
Observer				-0.066 (0.199)
Observations	399	399	399	399

Note. Linear regressions of the dummy variable indicating whether a study identifies a significantly negative effect. The trimmed sample excludes the top and bottom 2% of effect sizes. Robust standard in parenthesis. All models include controls (omitted from the table; the results are available from the authors upon request) for the country where the study took place and the discipline to which the journal is affiliated. (full results available in the SI, Table K). *Legend:* * $p < .10$, ** $p < .05$, *** $p < .01$.

E Included studies

The references have been updated according to the current status of the paper. Publications dates can be posterior to 2020 for articles that were included as forthcoming publications at the time of the systematic review of the literature.

- AKELEY SPEAR, J., AND A. N. MILLER (2012): “The Effects of Instructor Fear Appeals and Moral Appeals on Cheating-Related Attitudes and Behavior of University Students,” *Ethics & Behavior*, 22(3), 196–207.
- ALM, J., W. D. SCHULZE, C. VON BOSE, AND J. YAN (2019): “Appeals to Social Norms and Taxpayer Compliance,” *Southern Economic Journal*, 86(2), 638–666.
- ANDORFER, V. A., AND U. LIEBE (2015): “Do Information, Price, or Morals Influence Ethical Consumption? A Natural Field Experiment and Customer Survey on the Purchase of Fair Trade Coffee,” *Social Science Research*, 52, 330–350.
- AQUINO, K., D. FREEMAN, A. REED, V. K. G. LIM, AND W. FELPS (2009): “Testing a Social-Cognitive Model of Moral Behavior: The Interactive Influence of Situations and Moral Identity Centrality,” *Journal of Personality and Social Psychology*, 97(1), 123–141.
- ARIEL, B. (2012): “Deterrence and Moral Persuasion Effects on Corporate Tax Compliance: Findings from a Randomized Control Trial,” *Criminology*, 50(1), 27–69.
- BICCHIERI, C., E. DIMANT, AND S. SONDEREGGER (2023): “It’s Not a Lie If You Believe the Norm Does Not Apply: Conditional Norm-Following and Belief Distortion,” *Games and Economic Behavior*, 138, 321–354.
- BILANCINI, E., L. BONCINELLI, V. CAPRARO, T. CELADIN, AND R. DI PAOLO (2020): ““Do the Right Thing” for Whom? An Experiment on Ingroup Favouritism, Group Assorting and Moral Suasion,” *Judgment and Decision Making*, 15, 182–192.
- BLANKEN, I., N. VAN DE VEN, M. ZEELLENBERG, AND M. H. MEIJERS (2014): “Three Attempts to Replicate the Moral Licensing Effect,” *Social Psychology*, 45(3), 232–238.
- BLAUFUS, K., J. HUNDSDOERFER, M. JACOB, AND M. SÜNWOLDT (2016): “Does Legality Matter? The Case of Tax Avoidance and Evasion,” *Journal of Economic Behavior & Organization*, 127, 182–206.
- BLUMENTHAL, M., C. CHRISTIAN, AND J. SLEMROD (2001): “Do Normative Appeals Affect Tax Compliance? Evidence from a Controlled Experiment in Minnesota,” *National Tax Journal*, 54(1), 125–138.
- BOLSEN, T. (2013): “A Light Bulb Goes On: Norms, Rhetoric, and Actions for the Public Good,” *Political Behavior*, 35(1), 1–20.
- BOTT, K. M., A. W. CAPPELEN, E. Ø. SØRENSEN, AND B. TUNGODDEN (2020): “You’ve Got Mail: A Randomized Field Experiment on Tax Evasion,” *Management Science*, 66(7), 2801–2819.
- BRUDERMANN, T., G. BARTEL, T. FENZL, AND S. SEEBAUER (2015): “Eyes on Social Norms: A Field Study on an Honor System for Newspaper Sale,” *Theory and Decision*, 79(2), 285–306.
- BRÜGGER, A., AND B. HÖCHLI (2019): “The Role of Attitude Strength in Behavioral Spillover: Attitude Matters—But Not Necessarily as a Moderator,” *Frontiers in Psychology*, 10, 1018.

- BURSZTYN, L., S. FIORIN, D. GOTTLIEB, AND M. KANZ (2019): “Moral Incentives in Credit Card Debt Repayment: Evidence from a Field Experiment,” *Journal of Political Economy*, 127(4), 1641–1683.
- CAPPELEN, A. W., T. HALVORSEN, E. Ø. SØRENSEN, AND B. TUNGODDEN (2017): “Face-Saving or Fair-Minded: What Motivates Moral Behavior?,” *Journal of the European Economic Association*, 15(3), 540–557.
- CAPRARO, V., G. JAGFELD, R. KLEIN, M. MUL, AND I. V. DE POL (2019): “Increasing Altruistic and Cooperative Behaviour with Simple Moral Nudges,” *Scientific Reports*, 9(1), 11880.
- CASTRO, L., AND C. SCARTASCINI (2015): “Tax Compliance and Enforcement in the Pampas Evidence from a Field Experiment,” *Journal of Economic Behavior & Organization*, 116, 65–82.
- CHEN, X.-P., M. M. PILLUTLA, AND XIN YAO (2009): “Unintended Consequences of Cooperation Inducing and Maintaining Mechanisms in Public Goods Dilemmas: Sanctions and Moral Appeals,” *Group Processes & Intergroup Relations*, 12(2), 241–255.
- CLOT, S., G. GROLLEAU, AND L. IBANEZ (2014): “Smug Alert! Exploring Self-Licensing Behavior in a Cheating Game,” *Economics Letters*, 123(2), 191–194.
- COFFMAN, L. C., C. R. FEATHERSTONE, AND J. B. KESSLER (2017): “Can Social Information Affect What Job You Choose and Keep?,” *American Economic Journal: Applied Economics*, 9(1), 96–117.
- CYAN, M. R., A. M. KOUMPIAS, AND J. MARTINEZ-VAZQUEZ (2017): “The Effects of Mass Media Campaigns on Individual Attitudes towards Tax Compliance; Quasi-Experimental Evidence from Survey Data in Pakistan,” *Journal of Behavioral and Experimental Economics*, 70, 10–22.
- CZAJKOWSKI, M., K. ZAGÓRSKA, AND N. HANLEY (2019): “Social Norm Nudging and Preferences for Household Recycling,” *Resource and Energy Economics*, 58, 101110.
- D’ADDA, G., V. CAPRARO, AND M. TAVONI (2017): “Push, Don’t Nudge: Behavioral Spillovers and Policy Instruments,” *Economics Letters*, 154, 92–95.
- DAL BÓ, E., AND P. DAL BÓ (2014): ““Do the Right Thing:” The Effects of Moral Suasion on Cooperation,” *Journal of Public Economics*, 117, 28–38.
- DAMGAARD, M. T., AND C. GRAVERT (2018): “The Hidden Costs of Nudging: Experimental Evidence from Reminders in Fundraising,” *Journal of Public Economics*, 157, 15–26.
- DANESE, G., AND L. MITTONE (2018): “Trust and Trustworthiness in Organizations: The Role of Monitoring and Moral Suasion,” *Managerial and Decision Economics*, 39(1), 46–55.
- DEAN, A. J., K. S. FIELDING, AND K. A. WILSON (2019): “Building Community Support for Coastal Management — What Types of Messages Are Most Effective?,” *Environmental Science & Policy*, 92, 161–169.
- DEMARQUE, C., L. CHARALAMBIDES, D. J. HILTON, AND L. WAROQUIER (2015): “Nudging Sustainable Consumption: The Use of Descriptive Norms to Promote a Minority Behavior in a Realistic Online Shopping Environment,” *Journal of Environmental Psychology*, 43, 166–174.
- DIMANT, E., G. A. VAN KLEEF, AND S. SHALVI (2020): “Requiem for a Nudge: Framing Effects in Nudging Honesty,” *Journal of Economic Behavior & Organization*, 172, 247–266.

- DØSKELAND, T., AND L. J. T. PEDERSEN (2016): “Investing with Brain or Heart? A Field Experiment on Responsible Investment,” *Management Science*, 62(6), 1632–1644.
- FANGHELLA, V., G. D’ADDA, AND M. TAVONI (2019): “On the Use of Nudges to Affect Spillovers in Environmental Behaviors,” *Frontiers in Psychology*, 10, 61.
- FELLNER, G., R. SAUSGRUBER, AND C. TRAXLER (2013): “Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information,” *Journal of the European Economic Association*, 11(3), 634–660.
- FERRARO, P. J., J. J. MIRANDA, AND M. K. PRICE (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.
- FERRARO, P. J., AND M. K. PRICE (2013): “Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment,” *Review of Economics and Statistics*, 95(1), 64–73.
- FRITSCHKE, I., E. JONAS, D. N. KAYSER, AND N. KORANYI (2010): “Existential Threat and Compliance with Pro-Environmental Norms,” *Journal of Environmental Psychology*, 30(1), 67–79.
- GINO, F., AND S. D. DESAI (2012): “Memory Lane and Morality: How Childhood Memories Promote Prosocial Behavior,” *Journal of Personality and Social Psychology*, 102(4), 743–758.
- GRANT, A., AND J. DUTTON (2012): “Beneficiary or Benefactor: Are People More Prosocial When They Reflect on Receiving or Giving?,” *Psychological Science*, 23(9), 1033–1039.
- HASSELDINE, J., P. HITE, S. JAMES, AND M. TOUMI (2007): “Persuasive Communications: Tax Compliance Enforcement Strategies for Sole Proprietors*,” *Contemporary Accounting Research*, 24(1), 171–194.
- HURLSTONE, M. J., S. LEWANDOWSKY, B. R. NEWELL, AND B. SEWELL (2014): “The Effect of Framing and Normative Messages in Building Support for Climate Policies,” *PLoS ONE*, 9(12), e114335.
- HURST, R., T. TIDWELL, AND D. HAWKINS (2017): “Down the Rathole? Public Support for US Foreign Aid,” *International Studies Quarterly*, 61(2), 442–454.
- ITO, K., T. IDA, AND M. TANAKA (2018): “Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand,” *American Economic Journal: Economic Policy*, 10(1), 240–267.
- JAEGER, C. M., AND P. W. SCHULTZ (2017): “Coupling Social Norms and Commitments: Testing the Underdetected Nature of Social Influence,” *Journal of Environmental Psychology*, 51, 199–208.
- KIDWELL, B., A. FARMER, AND D. M. HARDESTY (2013): “Getting Liberals and Conservatives to Go Green: Political Ideology and Congruent Appeals,” *Journal of Consumer Research*, 40(2), 350–367.
- KLEINLOGEL, E. P., J. DIETZ, AND J. ANTONAKIS (2018): “Lucky, Competent, or Just a Cheat? Interactive Effects of Honesty-Humility and Moral Cues on Cheating Behavior,” *Personality and Social Psychology Bulletin*, 44(2), 158–172.
- KONOW, J. (2019): “Can Ethics Instruction Make Economics Students More Pro-Social?,” *Journal of Economic Behavior & Organization*, 166, 724–734.
- KOUCHAKI, M. (2011): “Vicarious Moral Licensing: The Influence of Others’ Past Moral Actions on Moral Behavior,” *Journal of Personality and Social Psychology*, 101(4), 702–715.

- LAMBARAA, F., AND G. RIENER (2015): “On the Norms of Charitable Giving in Islam: Two Field Experiments in Morocco,” *Journal of Economic Behavior & Organization*, 118, 69–84.
- LEE, S., K. P. WINTERICH, AND W. T. ROSS (2014): “I’m Moral, but I Won’t Help You: The Distinct Roles of Empathy and Justice in Donations,” *Journal of Consumer Research*, 41(3), 678–696.
- LINDAUER, M., M. MAYORGA, J. D. GREENE, P. SLOVIC, D. VÄSTFJÄLL, AND P. SINGER (2020): “Comparing the Effect of Rational and Emotional Appeals on Donation Behavior,” *Judgment and Decision Making*, 15(3), 413–420.
- LINDER, N., T. LINDAHL, AND S. BORGSTRÖM (2018): “Using Behavioural Insights to Promote Food Waste Recycling in Urban Households—Evidence From a Longitudinal Field Experiment,” *Frontiers in Psychology*, 9, 352.
- MARTICHUSKI, D. K., AND P. A. BELL (1991): “Reward, Punishment, Privatization, and Moral Suasion in a Commons Dilemma,” *Journal of Applied Social Psychology*, 21(16), 1356–1369.
- MASSER, B., AND L. PHILLIPS (2003): ““What Do Other People Think?” — The Role of Prejudice and Social Norms in the Expression of Opinions against Gay Men,” *Australian Journal of Psychology*, 55(3), 184–190.
- MCGRAW, K. M., AND J. T. SCHOLZ (1991): “Appeals to Civic Virtue Versus Attention to Self-Interest: Effects on Tax Compliance,” *Law & Society Review*, 25(3), 471–498.
- MORETON, S. G., A. ARENA, M. J. HORNSEY, C. R. CRIMSTON, AND N. TILIOPOULOS (2019): “Elevating Nature: Moral Elevation Increases Feelings of Connectedness to Nature,” *Journal of Environmental Psychology*, 65, 101332.
- MORTENSEN, C. R., R. NEEL, R. B. CIALDINI, C. M. JAEGER, R. P. JACOBSON, AND M. M. RINGEL (2019): “Trending Norms: A Lever for Encouraging Behaviors Performed by the Minority,” *Social Psychological and Personality Science*, 10(2), 201–210.
- NOLAN, J. M., P. W. SCHULTZ, R. B. CIALDINI, N. J. GOLDSTEIN, AND V. GRISKEVICIUS (2008): “Normative Social Influence Is Underdetected,” *Personality and Social Psychology Bulletin*, 34(7), 913–923.
- PALOMO-VÉLEZ, G., J. M. TYBUR, AND M. VAN VUGT (2018): “Unsustainable, Unhealthy, or Disgusting? Comparing Different Persuasive Messages against Meat Consumption,” *Journal of Environmental Psychology*, 58, 63–71.
- PETH, D., O. MUSSHOF, K. FUNKE, AND N. HIRSCHAUER (2018): “Nudging Farmers to Comply With Water Protection Rules – Experimental Evidence From Germany,” *Ecological Economics*, 152, 310–321.
- PRUCKNER, G. J., AND R. SAUSGRUBER (2013): “Honesty on the Streets: A Field Study on Newspaper Purchasing,” *Journal of the European Economic Association*, 11(3), 661–679.
- RAIHANI, N. J., AND K. MCAULIFFE (2014): “Dictator Game Giving: The Importance of Descriptive versus Injunctive Norms,” *PLoS ONE*, 9(12), e113826.
- REED, A., A. KAY, S. FINNEL, K. AQUINO, AND E. LEVY (2016): “I Don’t Want the Money, I Just Want Your Time: How Moral Identity Overcomes the Aversion to Giving Time to Prosocial Causes.,” *Journal of Personality and Social Psychology*, 110(3), 435–457.

- REESE, G., K. LOEW, AND G. STEFFGEN (2014): “A Towel Less: Social Norms Enhance Pro-Environmental Behavior in Hotels,” *Journal of Social Psychology*, 154(2), 97–100.
- REESE, P. P., K. GLANZ, A. SHAH, A. MUSSELL, S. LEVSKY, L. SHUDA, J. SHULTS, AND J. B. KESSLER (2020): “A Randomized Trial of Theory-Informed Appeals for Organ Donor Registration Using Internet Advertisements,” *Kidney International Reports*, 5(12), 2238–2245.
- REGE, M., AND K. TELLE (2004): “The Impact of Social Approval and Framing on Cooperation in Public Good Situations,” *Journal of Public Economics*, 88(7-8), 1625–1644.
- SACHDEVA, S., R. ILIEV, AND D. L. MEDIN (2009): “Sinning Saints and Saintly Sinners: The Paradox of Moral Self-Regulation,” *Psychological Science*, 20(4), 523–528.
- SCHLÜTER, A., AND B. VOLLAN (2015): “Flowers and an Honour Box: Evidence on Framing Effects,” *Journal of Behavioral and Experimental Economics*, 57, 186–199.
- SCHULTZ, P. W. (1999): “Changing Behavior With Normative Feedback Interventions: A Field Experiment on Curbside Recycling,” *Basic and Applied Social Psychology*, 21(1), 25–36.
- SCHULTZ, W. P., A. M. KHAZIAN, AND A. C. ZALESKI (2008): “Using Normative Social Influence to Promote Conservation among Hotel Guests,” *Social Influence*, 3(1), 4–23.
- SCHWARTZ, R. D., AND S. ORLEANS (1967): “On Legal Sanctions,” *The University of Chicago Law Review*, 34(2), 274.
- SERVÁTKA, M. (2009): “Separating Reputation, Social Influence, and Identification Effects in a Dictator Game,” *European Economic Review*, 53(2), 197–209.
- SEVERSON, A. W., AND E. A. COLEMAN (2015): “Moral Frames and Climate Change Policy Attitudes,” *Social Science Quarterly*, 96(5), 1277–1290.
- SHANG, J., A. REED, A. SARGEANT, AND K. CARPENTER (2020): “Marketplace Donations: The Role of Moral Identity Discrepancy and Gender,” *Journal of Marketing Research*, 57(2), 375–393.
- SHARIFF, A. F., AND A. NORENZAYAN (2007): “God Is Watching You: Priming God Concepts Increases Prosocial Behavior in an Anonymous Economic Game,” *Psychological Science*, 18(9), 803–809.
- SHU, L. L., F. GINO, AND M. H. BAZERMAN (2011): “Dishonest Deed, Clear Conscience: When Cheating Leads to Moral Disengagement and Motivated Forgetting,” *Personality and Social Psychology Bulletin*, 37(3), 330–349.
- SMEESTERS, D., L. WARLOP, E. VAN AVERMAET, O. CORNEILLE, AND V. YZERBYT (2003): “Do Not Prime Hawks with Doves: The Interplay of Construct Activation and Consistency of Social Value Orientation on Cooperative Behavior,” *Journal of Personality and Social Psychology*, 84(5), 972–987.
- SPARKMAN, G., AND G. M. WALTON (2017): “Dynamic Norms Promote Sustainable Behavior, Even If It Is Counternormative,” *Psychological Science*, 28(11), 1663–1674.
- TERRIER, L., AND B. MARFAING (2015): “Using Social Norms and Commitment to Promote Pro-Environmental Behavior among Hotel Guests,” *Journal of Environmental Psychology*, 44, 10–15.
- THOMSON, A. L., J. NAKAMURA, J. T. SIEGEL, AND M. CSIKSZENTMIHALYI (2014): “Elevation and Mentoring: An Experimental Assessment of Causal Relations,” *Journal of Positive Psychology*, 9(5), 402–413.

- TINKLER, J. E., J. CLAY-WARNER, AND M. ALINOR (2018): “Communicating About Affirmative Consent: How the Threat of Punishment Affects Policy Support and Gender Stereotypes,” *Journal of Interpersonal Violence*, 33(21), 3344–3366.
- TORGLER, B. (2013): “A Field Experiment in Moral Suasion and Tax Compliance Focusing on Underdeclaration and Overdeduction,” *FinanzArchiv*, 69(4), 393–411.
- TYLER, T. R., R. ORWIN, AND L. SCHURER (1982): “Defensive Denial and High Cost Prosocial Behavior,” *Basic and Applied Social Psychology*, 3(4), 267–281.
- VAN DE VYVER, J., AND D. ABRAMS (2015): “Testing the Prosocial Effectiveness of the Prototypical Moral Emotions: Elevation Increases Benevolent Behaviors and Outrage Increases Justice Behaviors,” *Journal of Experimental Social Psychology*, 58, 23–33.
- VAN DER LINDEN, S. (2015): “Exploring Beliefs About Bottled Water and Intentions to Reduce Consumption: The Dual-Effect of Social Norm Activation and Persuasive Information,” *Environment and Behavior*, 47(5), 526–550.
- WHITE, K., AND B. SIMPSON (2013): “When Do (and Don’t) Normative Appeals Influence Sustainable Consumer Behaviors?,” *Journal of Marketing*, 77(2), 78–95.
- WOLSKO, C. (2017): “Expanding the Range of Environmental Values: Political Orientation, Moral Foundations, and the Common Ingroup,” *Journal of Environmental Psychology*, 51, 284–294.
- WOLSKO, C., H. ARICEAGA, AND J. SEIDEN (2016): “Red, White, and Blue Enough to Be Green: Effects of Moral Framing on Climate Change Attitudes and Conservation Behaviors,” *Journal of Experimental Social Psychology*, 65, 7–19.
- ZHAO, J., Z. DONG, AND R. YU (2019): “Don’t Remind Me: When Explicit and Implicit Moral Reminders Enhance Dishonesty,” *Journal of Experimental Social Psychology*, 85(103895).
- ZHU, L. L., S. L. D. RESTUBOG, K. LEAVITT, L. ZHOU, AND M. WANG (2020): “Lead the Horse to Water, but Don’t Make Him Drink: The Effects of Moral Identity Symbolization on Coworker Behavior Depend on Perceptions of Proselytization,” *Organizational Behavior and Human Decision Processes*, 156, 53–68.