Supplemental Appendix

0.0 Contents

1.0 Deviations from Preregistration

The following deviations to our preregistration were made while conducting the present meta-analysis:

First, we pre-registered that we would report both τ and τ^2, given that these values are redundant, we opted to only report τ^2.

Second, we preregistered that we would report the weights of studies, noting when they were greater than 5%. For our main analyses with more than 60 studies, no studies reached a 5% weight or higher. For comparisons with less than 60 studies, the proportion of studies with 5% or greater weights was much larger, owing to the small sample of comparisons. Rather than list each individual weight as a number, we opted to depict them in our forest plots based on the size of the square representing the standardized mean difference in each comparison.

Third, after conducting our initial analysis, we identified one study with a standardized mean difference of 5.11. This effect size is impossibly large. For comparison, the next largest

effect size in any of our comparisons was < 1.00. Upon further investigation, we identified that the authors of this paper adjusted their means but did not specify how. When contacted, the authors reported they no longer had access to the data. As such, we opted to remove the study from our main analysis. We report the results of this comparison in section SA 2.3 (below).

2.0 Supplemental Results

2.1 Pre-Test Inclusion Analysis

For many of the included studies, the original articles included pretest-posttest designs. It is generally advised to only include posttest scores in meta-analyses because the within-subjects effects can bias the results (Cuijpers, 2017). However, in many of our studies the "pretest-posttest" design is not truly within subjects. For many of these comparisons, the researchers tested the same messages in the same location at multiple dates but cannot confirm that the same people saw the messages. Additionally, the pre-tests in many of these conditions were no-message conditions. As such, the "pre-test" is conceptually the same as a no-message control condition. As to not bias the overall results, we did not include the pre-test scores as "no-message controls" in the main analyses. Instead, we report revised comparisons between message conditions that include pretest scores as no-message control conditions in Table SA1. We also include 1 additional analysis where a behavioral request was included at pretest and compared to a behavioral request paired with a descriptive norms message at posttest.

2.2. Additional Models (Comparisons with k < 10)

The present meta-analysis was designed to test as many comparisons between different types of social norms messages (and behavioral requests) as possible. However, many of these comparisons were not possible due to insufficient observations ($k < 10$). In Table SA2, we report the output of these models. We note that these models are severely underpowered and should be

interpreted with caution. We are reporting them for the purposes of transparency and in keeping with our preregistration, but we will not interpret their output.

Table SA1

Output from Additional Pretest Scores Included.

Comparison Type	**Cohen's d**	**95% CI [Upper, Lower]**	**95% PI [Upper, Lower]**	**τ^2**	***Q***	**df**	***p***	**I^2**
Control vs. ~~Behave.~~Beh. Req.	0.23	[0.20, 0.27]	[0.00, 0.46]	0.01	644.55	75	<.001	88.36
Control vs. Descriptive	0.04	[0.02, 0.05]	[-0.11, 0.18]	0.06	730.62	116	<.001	84.12
Control vs. Injunctive	-0.01	[-0.03, 0.02]	[-0.11, 0.10]	0.00	107.90	26	<.001	75.90
Control vs. ~~Behave.~~Beh. Req. & Descriptive	0.29	[0.20, 0.38]	[-0.08, 0.66]	0.03	149.97	18	<.001	88.00

Note. For each of the above analyses, we report Cohen's *d* (the Standardized Mean Difference), the 95% Confidence and Prediction Intervals, τ^2, *Q* with its associated degrees of freedom and *p*-value, and I^2. The message conditions are coded as "Control" = control, "~~Behave.~~Beh. Req" = behavioral request, "Descriptive" = descriptive norms message, and "Injunctive" = injunctive norms message. The & symbol means that 2 or more types of messages were combined in the same intervention. A Cohen's *d* with positive values indicates the message type to the right of the "vs" in the comparison type column was the more effective intervention.

Table SA2

Output from Models with Few Studies (k <10).

Comparison Type	Cohen's d	95% CI [Upper, Lower]	95% PI [Upper, Lower]	τ^2	*Q*	df	*p*	I^2
Control vs. ~~Behave.~~Beh. Req. & Injunctive	1.14	[0.98, 1.31]	-	0.00	0.17	1	.69	0.00
Control vs. Descriptive & Injunctive	0.03	[-0.01, 0.07]	[-0.10, 0.16]	0.00	16.35	4	.00	75.53
~~Behave.~~Beh. Req. vs. Injunctive	-0.01	[-0.12, 0.11]	-	0.00	0.09	3	.99	0.00
~~Behave.~~Beh. Req. vs. ~~Behave.~~Beh. Req. & Injunctive	0.58	[0.43, 0.73]	[0.31, 085]	0.06	5.59	5	.35	10.51
~~Behave.~~Beh. Req. vs. ~~Behave.~~Beh. Req. & Descriptive & Injunctive	0.29	[0.18, 0.40]	-	0.00	1.69	2	.43	0.00
Descriptive vs. Descriptive & Injunctive	0.01	[-0.01, 0.02]	[-0.07, 0.09]	0.00	23.19	7	.002	69.82
Injunctive vs. Descriptive & Injunctive	0.05	[0.01, 0.08]	[-0.06, 0.15]	0.00	9.36	4	.05	57.26
Injunctive vs. ~~Behave.~~Beh. Req./Descriptive	-0.14	[-0.31, 0.02]	[-0.67, 0.38]	0.09	4.09	3	.25	26.65

Note. For each of the above analyses, we report Cohen's *d* (the Standardized Mean Difference), the 95% Confidence and Prediction Intervals, τ^2, *Q* with its associated degrees of freedom and *p*-value, and I^2. The message conditions are coded as "Control" = control, "~~Behave.~~Beh. Req" = behavioral request, "Descriptive" = descriptive norms message, and "Injunctive" = injunctive norms message. The & symbol means that 2 or more types of messages were combined in the same intervention. A Cohen's *d* with positive values indicates the message type to the right of the "vs" in the comparison type column was the more effective intervention. In some instances, the 95% PI could not be calculated due to the low sample size or low I^2

2.3 Model with Excluded Study (d = 5.11)

When we included the one comparison with an excessively high Cohen's *d*, we found almost no difference in the effectiveness of behavioral requests compared to control conditions. We found that behavioral requests had a generally small, positive effect on behavior relative to control conditions (d = 0.24, 95% CI[0.19, 0.29]). While this effect was larger than the one reported in our main analysis, so too were the prediction intervals (95% PI[-0.07, 0.54]). We observed a higher proportion of the variation between studies that could be accounted by heterogeneity (rather than sampling error; I^2 = 93.14%, $Q(72)$ = 1049.23, $p < .001$, τ^2 = 0.02). Overall, the inclusion of this study did not affect the interpretation of our results.

The same publication that reported an extremely high Cohen's *d* for the comparison between behavioral requests and control conditions also included 5 other types of comparisons. Of these comparisons, 3 are only examined within that study: control v. behavioral requests combined with descriptive and injunctive norms messages (d = -5.83), descriptive norms messages combined with injunctive norms messages v. behavioral requests (d = -0.86), and descriptive norms messages combined with injunctive norms messages v. behavioral requests combined with descriptive and injunctive norms messages (d = -1.23).

We re-conducted 2 other analyses to include comparisons from the excluded paper. First, we compared the effects of a control condition with the combined descriptive and injunctive norms message. When adding the previously excluded study (d = 4.31), we found that the overall effect size increased from d = 0.03 to d = 0.48 (95% CI[0.30, 0.67], 95% PI[-0.19, 1.16], I^2 = 99.02%, $Q(5)$ = 509.69, $p < .001$, τ^2 = 0.05). Second, we compared the effects of behavioral requests to the messages combining behavioral requests, descriptive norms messages, and injunctive norms messages. When adding the excluded effect (d = 0.31), the estimated

standardized mean difference did not change (d = 0.29, 95% CI[0.20, 0.39], I^2 = 0.00%, $Q(3) = 1.72$, $p = .63$, $\tau^2 = 0.00$).

Overall, while we cannot determine why the one study reports consistently high values, its exclusion did not change our interpretation of the results.

2.4 Meta Regression

The central aim of the present meta-analysis is to determine which type(s) of messages are most effective at eliciting behavior change. While recategorizing messages, we not only identified behavioral requests as a separate type of message, but found several other categories of text that may influence the effectiveness of a persuasive message. To test whether each type of text (behavioral request, descriptive norms message, injunctive norms message, instructional language, justifications, financial incentives, contextual information, irrelevant information, and polite language) each effect the likelihood of behavior change, we conducted a meta-regression.

For each comparison included in our main analyses, we identified whether the treatment and comparison messages included each of the above types of text (coded as 0 or 1). Thus, for each of the 9 types of text, we had 4 possible categories (2 levels for the treatment message and 2 levels for the control message[1]). For instance, a message comparison could compare a message that includes (or does not include) a behavioral request with another messages that includes (or does not include) a behavioral request. For each type of message text, we created 3 sets of dummy codes. Message comparisons wherein neither message contained a given type of text (e.g., neither message contains a behavioral request) were used as the referent group.

[1] note that messages were arbitrarily assigned to be “treatment” and “control” messages for the purpose of drawing comparisons within the Comprehensive Meta-Analysis program

Using the above-described process, we created a model with 27 parameters (1 intercept, 3 parameters for dummy codes testing the effects of behavioral requests, 3 parameters for dummy codes testing the effects of descriptive norms messages, 2 parameters for dummy codes testing the effects of injunctive norms messages, 3 parameters for dummy codes testing the effects of instructional language, 3 parameters for dummy codes testing the effects of justifications, 3 parameters for dummy codes testing the effects of financial incentives, 3 parameters for dummy codes testing the effects of contextual information, 3 parameters for dummy codes testing the effects of irrelevant texts, and 3 parameters for dummy codes testing the effects of polite language). Note that with 294 studies included in the meta-regression, we are underpowered to conduct the meta-regression. Thus, we interpret the following results with caution.

We observed the comparisons between messages that included (versus did not include) behavioral requests had larger effects than comparisons in which neither message included a behavioral request, $d = 0.14$, 95% CI[0.09, 0.19], $p < .001$ (for treatment messages with and control messages without behavioral requests) and $d = -0.08$, 95% CI[-0.14, -0.02], $p = .01$ (for treatment messages without and control messages with behavioral requests). These results suggest that above and beyond the effects of all other types of text, behavioral requests do increase the likelihood that an individual will adhere to the target behavior.

Nearly all message comparisons testing the effects of descriptive norms messages compared either two messages without descriptive norms messages or a treatment messages with one and a control message without one, but no difference was observed between these two types of comparisons ($d = 0.02$, 95% CI[-0.03, 0.07], $p = .39$). We did, however, observe that treatment messages without descriptive norms messages were less effective at changing behavior than control messages with descriptive norms messages ($d = -0.06$, 95% CI[-0.10, -0.01],

p = .02). Only one study compared two messages with descriptive norms messages, and while we do observe a standardized mean difference that is significantly different from comparisons where neither message contains a descriptive norms message, this effect is unlikely to replicate (*d* = -0.66, 95% CI[-1.16, -0.15], *p* = .01). These results suggest that descriptive norms have a small effect on the likelihood that an individual adheres to a desired behavior, even when controlling for the effects of all other types of message text.

We observed only one significant effect relevant to the inclusion of instructional language. Control messages that included instructional (how-to) information were less effective than treatment messages that did not include instructional information., *d* = 0.09, 95% CI[0.00, 0.18], *p* = .04. We note, however, that this effect is just below the traditional threshold for significance and may not replicate. Further, this model already controls for the effects of behavioral requests, which often heavily imply how to do a behavior (e.g., "please recycle by putting aluminum cans in the bin" has both a behavioral request and instructional language, but the instructional language is largely redundant with the behavioral request). Thus, it may be the case that instructional language is effective, but not when controlling for the effects of behavioral requests.

Unsurprisingly, providing people with financial incentives increased their likelihood to engage in a behavior, even when controlling for the effects of other types of message text. Both treatment messages with incentives (compared to controls without incentives; *d* = 0.54, 95% CI[0.08, 1.01], *p* = .02) and controls with incentives (compared to treatments without incentives; *d* = -0.06, 95% CI[-0.11, -0.00], *p* = .03) were more likely to change people's behavior.

We only observed one significant effect for the inclusion of contextual information: treatment messages containing contextual information were less effective at changing behavior

than control messages that did not include contextual information (d = -0.09, 95% CI[-0.14, 0.04], p = .001).

We observed that the inclusion of polite language (please or thank you) increased the likelihood that an individual would comply with the desired behavior, above and beyond the effects of any other type of message text. Treatment messages with polite language were more effective than control messages without it (d = 0.17, 95% CI[0.10, 0.24], p < .001). Further, treatment messages without polite language were less effective than controls with it (d = -0.20, 95% CI[-0.34, -0.06], p = .005). Thus, we observe that not only is polite language effective, but it has among the largest effects predicting that individuals will adhere to a desired behavior.

We observed no significant effects for injunctive norms messages, justification statements, or irrelevant information, when controlling for any other type of message. The full output of the meta-regression analysis is provided in Table SA3.

Table SA3

Output of Meta-Regression Model Testing all Types of Message Text

Message Text	Control	Treatment	*d*	95% CI	*z*	*p*	Overall *Q*(df), *p* (For Message)
Behavioral Request	0	1	0.14	[0.09, 0.19]	5.24	.000	*Q*(3) = 43.50,
	1	0	-0.08	[-0.14, -0.02]	-2.63	.009	*p* < .001
	1	1	0.05	[-0.01, 0.11]	1.71	.087	
Descriptive Norms Message	0	1	0.02	[-0.03, 0.07]	0.87	.385	*Q*(3) = 12.88,
	1	0	-0.05	[-0.10, -0.01]	-2.27	.023	*p* = .005
	1	1	-0.66	[-1.16, -0.15]	-2.55	.011	
Injunctive Norms Message	0	1	0.00	[-0.06, 0.06]	-0.01	.993	*Q*(2) = 0.07,
	1	0	-0.03	[-0.25, 0.19]	-0.26	.794	*p* = .966
Instructional Language	0	1	-0.05	[-0.11, 0.01]	-1.54	.123	*Q*(3) = 7.84,
	1	0	0.09	[0.00, 0.18]	2.03	.043	*p* = .049
	1	1	-0.02	[-0.06, 0.01]	-1.37	.171	
Justification	0	1	0.05	[0.00, 0.11]	1.95	.051	*Q*(3) = 7.86,
	1	0	-0.03	[-0.07, 0.00]	-1.72	.086	*p* = .049
	1	1	-0.01	[-0.06, 0.04]	-0.5	.619	
Financial Incentive	0	1	0.54	[0.08, 1.01]	2.28	.023	*Q*(3) = 9.86,
	1	0	-0.06	[-0.11, 0.00]	-2.13	.033	*p* = .020
	1	1	0.00	[-0.07, 0.07]	0.03	.973	
Contextual Information	0	1	-0.09	[-0.14, -0.03]	-3.26	.001	*Q*(3) = 11.97,
	1	0	0.03	[-0.10, 0.16]	0.49	.623	*p* = .008
	1	1	-0.02	[-0.08, 0.04]	-0.75	.454	
Irrelevant Information	0	1	0.45	[0.04, 0.87]	2.16	.031	*Q*(3) = 7.53,
	1	0	0.07	[-0.02, 0.16]	1.45	.147	*p* = .057
	1	1	-0.04	[-0.13, 0.05]	-0.96	.337	
Polite Language	0	1	0.17	[0.10, 0.24]	4.69	.00	*Q*(3) = 33.94,
	1	0	-0.20	[-0.34, -0.06]	-2.82	.005	*p* < .001
	1	1	0.05	[0.00, 0.10]	1.84	.066	

Note. For each message text type, we report whether the control and treatment messages did (1) or did not (0) include that type of message text, the *d* value estimating the standardized mean difference for that message comparison type (along with the 95% Confidence Interval). We then report the *z* and *p*-values indicating whether that message comparison was significantly more likely to elicit behavior change than a message comparison whether neither message contained that type of message text. Finally, we report the overall *Q*-value, degrees of freedom, and *p*-value for that type of message comparison. Not depicted is the estimated intercept for the meta regression (*d* = 0.04, 95% CI[-0.02, 0.10], *p* = .15).

3.0 Categorization Analysis

One of the central themes of the present meta-analysis was the need to clarify the categorization of the different messages used to test the effects of social norms. After we completed the meta-analysis, we were curious as to the proportion of messages that we categorized similarly to the original authors. Thus, for each unique message, we compared our categorization with that of the message's original authors. In general, we found that while we often concurred with the original authors categorization for descriptive norms messages, 40% included behavioral requests. Further, 15% of messages identified as controls also included behavioral requests. We had the most disagreement regarding the categorization of injunctive norms messages. We found that a majority (57%) of injunctive norms messages included behavioral requests, but only a minority (37%) included injunctive norms messages. This brief examination of the categorization of messages, reveals the extent to which researchers have not used consistent labeling across studies when testing the effects of social norms messages. Further results are depicted in Table SA4.

Table SA4

Original Label of Messages and Contents of Those Messages as Recategorized in the Present Meta-Analysis

Original Label	***N***	**No Message**	**~~Behave.~~Beh. Req.**	**Descriptive**	**Injunctive**	**Polite Lang.**	**Instruct. Lang.**	**Justification**	**Financial Incitive**	**Contextual Info.**	**Irrelevant Info**
Control	59	33 (56%)	9 (15%)	1 (2%)	-	8 (14%)	16 (27%)	8 (14%)	4 (7%)	3 (5%)	5 (8%)
~~Behave.~~Beh. Req.	4	-	4 (100%)	-	-	3 (75%)	-	1 (25%)	-	2 (50%)	-
Descriptive	58	-	23 (40%)	58 (100%)	-	9 (16%)	26 (45%)	16 (28%)	2 (3%)	12 (21%)	2 (3%)
Injunctive	35	-	20 (57%)	4 (11%)	13 (37%)	12 (34%)	6 (17%)	13 (37%)	4 (11%)	9 (26%)	-
Descriptive & Injunctive	3	-	-	3 (100%)	2 (67%)	1 (33%)	3 (100%)	1 (33%)	-	1 (33%)	-
Non-Standardized Label	80	-	40 (50%)	15 (19%)	5 (6%)	27 (34%)	37 (46%)	44 (55%)	7 (9%)	19 (24%)	6 (8%)

Note. For each of the messages included in the meta-analysis, we identify the number that were identified as a control message, behavioral request, descriptive norms message, injunctive norms message, and combination of descriptive and injunctive norms message, or had another (non-standardized) label. We then identify the number (and percentage) of those messages that included no message, a behavioral request, a descriptive norms message, an injunctive norms message, polite language, instructional language, a justification for completing the behavior, a financial incentive for completing the behavior, additional contextual information, or irrelevant information.

4.0 Additional References (Studies Included Only in Analysis 2.1)

Arpan, L. M., Barooah, P., & Subramany, R. (2015). The role of values, moral norms, and descriptive norms in building occupant responses to an energy-efficiency pilot program and to framing of related messages. *Applied Environmental Education & Communication, 14*, 23-32.

Bonan, J., Cattaneo, C., d'Adda, G., & Tavoni, M. (2020). The interaction of descriptive and injunctive social norms in promoting energy conservation. *Nature Energy, 5*, 900-909.

Castro, J. F., Velásquez, D., Beltrán, A., & Yamada, G. (2022). The direct and indirect effects of messages on tax compliance: Experimental evidence from Peru. *Journal of Economic Behavior & Organization, 203*, 483-518.

Chongwilaikasaem, S., & Ayaragarnchanakul, E. (2023). Electricity saving behavior in student dorms: What message framing works?. *Journal of Behavioral and Experimental Economics, 107*, 102088.

Eyink, J. R., Motz, B. A., Heltzel, G., & Liddell, T. M. (2020). Self-regulated studying behavior, and the social norms that influence it. *Journal of Applied Social Psychology, 50*, 10-21.

Huitink, M., Poelman, M. P., van Den Eynde, E., Seidell, J. C., & Dijkstra, S. C. (2020). Social norm nudges in shopping trolleys to promote vegetable purchases: A quasi-experimental study in a supermarket in a deprived urban area in the Netherlands. *Appetite, 151*, 104655.

Mertens, S. N., & Schultz, P. W. (2020). Referent Group Specificity: Optimizing Normative Feedback to Increase Residential Recycling. *Journal of Environmental Psychology, 73*, 101541.

Neighbors, C., O'Connor, R. M., Lewis, M. A., Chawla, N., Lee, C. M., & Fossos, N. (2008). The relative impact of injunctive norms on college student drinking: the role of reference group. *Psychology of Addictive Behaviors, 22,* 576.

Reid, A. E., & Aiken, L. S. (2013). Correcting injunctive norm misperceptions motivates behavior change: a randomized controlled sun protection intervention. *Health Psychology, 32,* 551.

Salmivaara, L., & Lankoski, L. (2021). Promoting sustainable consumer behaviour through the activation of injunctive social norms: a field experiment in 19 workplace restaurants. *Organization & Environment, 34,* 361-386.