Supplementary Information Appendix

For: **Chilling Results: How Explicit Warm Glow Appeals Fail to Boost Pro-Environmental Behaviour**

Paul Lohmann[*], Elisabeth Gsottbauer, Sander van der Linden, Andreas Kontoleon

* Corresponding author: Paul Lohmann, El-Erian Institute of Behavioural Economics and Policy, Cambridge Judge Business School, University of Cambridge, UK; p.lohmann@jbs.cam.ac.uk

Contents

1. Research Questions

We address the following four research questions: (1) Do intrinsically or extrinsically motivated message frames influence the anticipated and experienced emotions of engaging in PEB, relative to a Call-to-Action condition that receives only basic information on climate change? (2) Do warm glow, cold-prickle and social norm messages increase PEB relative to the baseline condition? (3) Do emotions mediate the relationship between treatment messages and PEB? (4) Does the effect of treatment messages on PEB persist over time?

Our first research question explores whether there is a main effect of the treatment interventions on pro-environmental behaviour relative to the baseline condition. To that end, we examine the relative effectiveness of different message frames in encouraging PEB. We hypothesised that all treatment messages would increase PEB relative to the baseline condition and that the warm glow message, increasing the salience of benefits-to-self, would perform best.

Our second research question asks whether intrinsically motivated messages increasing the salience of warm glow and cold-prickle emotions, or a descriptive social norm message influence anticipated and experienced emotions, relative to a Call-to-Action condition which received only basic information on climate change. We hypothesised that warm glow messaging, which highlights the positive emotional reward of PEB, would be positively associated with positive affect and negatively with negative affect. We assumed the opposite to be the case for cold-prickle messaging, which increased the salience of negative emotions (e.g., guilt and shame) from failing to act pro-environmentally. We did not expect to find a relationship between the social norm message and emotions.

The third research question examines whether anticipated and experienced emotions mediate the relationship between the treatment messages and pro-environmental behaviour. We hypothesised that anticipated positive emotions would partially mediate the relationship between the warm glow message and PEB. Similarly, we expected the cold prickle message to impact PEB via negative emotions.

Our fourth and final research question explores the persistency of treatment effects and thus contributes to an emerging literature which has largely highlighted the impermanence of behavioural interventions (Allcott and Rogers, 2014; Brandon et al., 2017; Gravert and Olsson, 2021). Specifically, we hypothesised that PEB would decrease in the second experimental wave (T2), however, to a lesser extent in the warm glow group. Moreover, we expected positive emotions to be higher in the warm glow group at T2.

In addition to our pre-registered hypotheses, we conduct additional exploratory analysis. First, we explore the relationship between value orientation (biospheric and altruistic) and PEB, and test to what extent this relationship is mediated by anticipated warm glow. Moreover, we investigate whether warm glow experiences experienced after engaging in pro-environmental behaviour mediates future pro-environmental behaviour (Brosch, 2021). The exploratory analysis allows us to examine whether pro-environmental behaviour and experienced emotions can form a positive self-reinforcing feedback loop with each other over time.

2. Attrition and balance checks

To reduce attrition, participants were given a bonus payment of £1 for completing all survey waves and received personalised reminders prior to each experimental survey.[1] As a result, attrition was low across waves. Of the 2980 participants randomly assigned to a treatment condition, 282 (9.46%) did not complete Wave 2 (main experimental survey) and 101 (3.74% of the remaining participants) did not complete Wave 3 (follow up survey). Our stratification and re-randomisation procedures were successful in achieving balance on socio-demographic and baseline ability. See Appendix Table A2. Columns (1) – (4) display the sample means for the baseline group and each treatment condition. Columns (5) to (6) display the differences in means between each treatment condition and the baseline (BL) group. We find that randomisation was successful and that after removing attritors, participants baseline characteristics in the four conditions are not statistically distinguishable from each other. We find that only biospheric values were slightly lower in the warm-glow group than in the baseline condition, significant at a 10% level.

Table A1: Balance Checks

Variable	(1) Call to Action	(2) Warm Glow	(3) Cold Prickle	(4) Social Norm	(5) WG vs. CTA	(6) CP vs. CTA	(7) SN vs. CTA
Female (%)	0.61	0.61	0.62	0.61	-0.00	0.02	0.00
	(0.49)	(0.49)	(0.48)	(0.49)	(0.03)	(0.03)	(0.03)
Age (Years)	37.23	37.00	36.56	36.20	-0.23	-0.67	-1.03
	(14.06)	(13.86)	(13.81)	(13.50)	(0.76)	(0.76)	(0.75)
Income	3.70	3.76	3.77	3.72	0.05	0.07	0.01
	(1.60)	(1.65)	(1.65)	(1.66)	(0.09)	(0.09)	(0.09)
Life Satisfaction	6.08	6.20	6.06	6.11	0.12	-0.02	0.04
	(2.12)	(2.03)	(2.11)	(2.02)	(0.11)	(0.12)	(0.11)
Education Level	3.42	3.43	3.48	3.47	0.01	0.06	0.05
	(0.75)	(0.75)	(0.72)	(0.71)	(0.04)	(0.04)	(0.04)
Chariable Behaviour	2.02	2.04	2.03	2.04	0.02	0.01	0.02
	(0.76)	(0.78)	(0.77)	(0.79)	(0.04)	(0.04)	(0.04)
Life Worthwile	6.31	6.44	6.26	6.36	0.13	-0.05	0.05
	(2.27)	(2.18)	(2.29)	(2.31)	(0.12)	(0.12)	(0.13)
Donation Behaviour	2.44	2.42	2.45	2.49	-0.02	0.01	0.04
	(1.26)	(1.23)	(1.30)	(1.33)	(0.07)	(0.07)	(0.07)
Altrusitic Values (scale)	5.44	5.38	5.47	5.43	-0.06	0.03	-0.01
	(1.41)	(1.42)	(1.43)	(1.44)	(0.08)	(0.08)	(0.08)
Biospheric Values (scale)	5.20	5.04	5.18	5.18	-0.16*	-0.02	-0.02
	(1.62)	(1.65)	(1.71)	(1.69)	(0.09)	(0.09)	(0.09)
Baseline Ability	2.34	2.33	2.35	2.37	-0.01	0.01	0.03
	(1.05)	(1.05)	(1.05)	(1.05)	(0.06)	(0.06)	(0.06)
Perceived Task Difficulty	3.34	3.42	3.44	3.36	0.08	0.09	0.02
	(1.08)	(1.09)	(1.09)	(1.10)	(0.06)	(0.06)	(0.06)
Perceived Task Enjoyability	3.05	3.06	3.07	3.05	0.02	0.02	0.01
	(1.20)	(1.19)	(1.18)	(1.22)	(0.06)	(0.06)	(0.07)
Observations	673	687	668	670	1,360	1,341	1,343

Note: This table presents balance checks of sample characteristics between the four treatment conditions. Columns (1) to (4) display the sample mean for each group, respecitvely. Columns (5) to (7) display the differences in the means of each treatment group compared to the 'Call to Action' condition, which was pre-registered as the comparison group. Significance stars on columns (5) to (7) indicate whether differences in means are statistically significant based on p-values obtained from two-sample t-tests.

---

[1] As the original four-wave design was reduced to three waves, the bonus payment was adjusted to £0.75

3. Estimation Strategy

We estimate a series of linear regressions to explore the effect of each treatment message on donation behaviour. We estimate both cross-sectional and longitudinal models. For our baseline specification, we restrict the sample to observations from the main experimental survey (Wave 2) during which the treatment messages were administered. For the longitudinal analysis, we estimate a repeated measures linear mixed-effect model with a random effect for each individual. The statistical method for fitting the model is residual maximum likelihood. The baseline specification for the cross-sectional analysis is presented below:

$$Y_i = \alpha_1 + \beta_1 Warm_i + \beta_2 Cold_i + \beta_3 Norm_i + \gamma X_i + \varepsilon \qquad (1)$$

where Y_i represents our outcome measure of pro-environmental behaviour (donation, quantity, quality, voluntary participation) of individual i. $Warm_i$, $Cold_i$ and $Norm_i$ are treatment indicators equal to one if the individual i was shown the warm glow, cold-prickle or social norm treatment video (the baseline condition serves as the reference category). As pre-specified, X_i is a vector of socio-demographic variables for individual i that are found to be unbalanced across groups. We thus only control for baseline biospheric values, which is unbalanced between the baseline condition and the warm glow condition. We estimate heteroskedasticity robust (Eicker-Huber-White) standard errors throughout the analysis. It is important to note that throughout our analysis, we estimate intention-to-treat (ITT) effects of our treatment messages on donation behaviour. While we set the video to auto play and did not allow participants to skip the video (i.e., the continuation button appeared only after the video had finished playing), we are not able to guarantee that all participants actively watched the video and paid attention to its contents.

The heterogeneity analysis was conducted following equation (1b), where VAL_i refers to a binary indicator identifying individuals with 'high' biospheric/altruistic values, which enters both as a main effect and interacted with the three treatment indicators ($Warm_i$, $Cold_i$ and $Norm_i$). 'Low' biospheric/altruistic values serves as the omitted base category.

$$Y_i = \alpha + \beta_1 Warm_i + \beta_2 Cold_i + \beta_3 Norm_i + \gamma_1 VAL_i + \delta_1(Warm_i \times VAL_i) \qquad (1b)$$
$$+ \delta_2(Cold_i \times VAL_i) + \delta_3(Norm_i \times VAL_i) + \gamma X_i + \varepsilon]$$

In a supplementary exploratory mediation analysis (Hayes, 2018) we examine the relationship between biospheric values, warm glow and pro-environmental behaviour. Specifically, we first explore whether the effect of biospheric and altruistic values on PEB is mediated by anticipated warm glow. Second, we examine whether experienced warm glow mediates the relationship between past and future pro-environmental effort. We follow a

causal mediation method based on Imai et al. (2010a, 2010b)[2]. In its general form, the causal mediation analysis is based on the following two models:

$$MV_i = \alpha_2 + \beta_2 IV_i + \delta_2 X_i + \varepsilon_{i2} \tag{2}$$

$$Y_i = \alpha_3 + \beta_3 IV_i + \gamma MV_i + \delta_3 X_i + \varepsilon_{i3} \tag{3}$$

Where MV_i is the mediating variable and IV_i is the independent variable; Y_i is the outcome variable and X_i is a vector of control variables (excluding the mediator). In the first step (equation 2), we estimate the direct effect of the independent variable on the mediating variable (β_2). In the second step (equation 3), we estimate the effect of both the independent (β_3) and the mediating variable (γ) on the outcome variable Y_i. Following Hicks and Tingeley (2011), we compute the Average Causal Mediated Effect (ACME) by taking the product of the coefficient on the independent variable (IV_i) in equation (2) with the coefficient on the mediating variable (γ) in equation (3). The ACME is calculated by "simulating predicted values of the mediator or outcome variable, which we do not observe, and then calculating the appropriate quantities of interest" (Hicks and Tingley, 2011, p.4). Specifically, standard errors and confidence intervals for the ACME are estimated based on the quasi-Bayesian Monte Carlo approximation of King et al. (2000).

4. Manipulation Checks

First, we assess how the treatment videos were perceived. To do so, we asked participants to rate the general sentiment of the video on a five-point Likert scale ranging from "extremely negative" to "extremely positive", immediately after viewing the video. Figure A1 shows the distribution of responses for each treatment condition in percentages.

A visual assessment of the response distributions indicate that the video sentiment was largely perceived as intended. Nearly all participants (90%) perceived the positively framed warm glow message to be either somewhat positive or extremely positive, while the negatively framed cold prickle message was perceived to be negative by most participants (60%). Although both the Call-to-Action condition (basic information message) and the social norm message did not explicitly address emotions in relation to PEB, they were perceived to portray a positive rather than a negative sentiment. The majority of participants indicated that they portrayed a somewhat positive or extremely positive sentiment. A simple ordered probit regression of perceived video sentiment on treatment indicators as in equation (1) confirms the visual assessment and shows that, on average, videos were perceived as intended (see Appendix Table A2).

---

[2] The mediation analysis is implemented in Stata using the user-contributed command "medeff" (Hicks and Tingley, 2011).
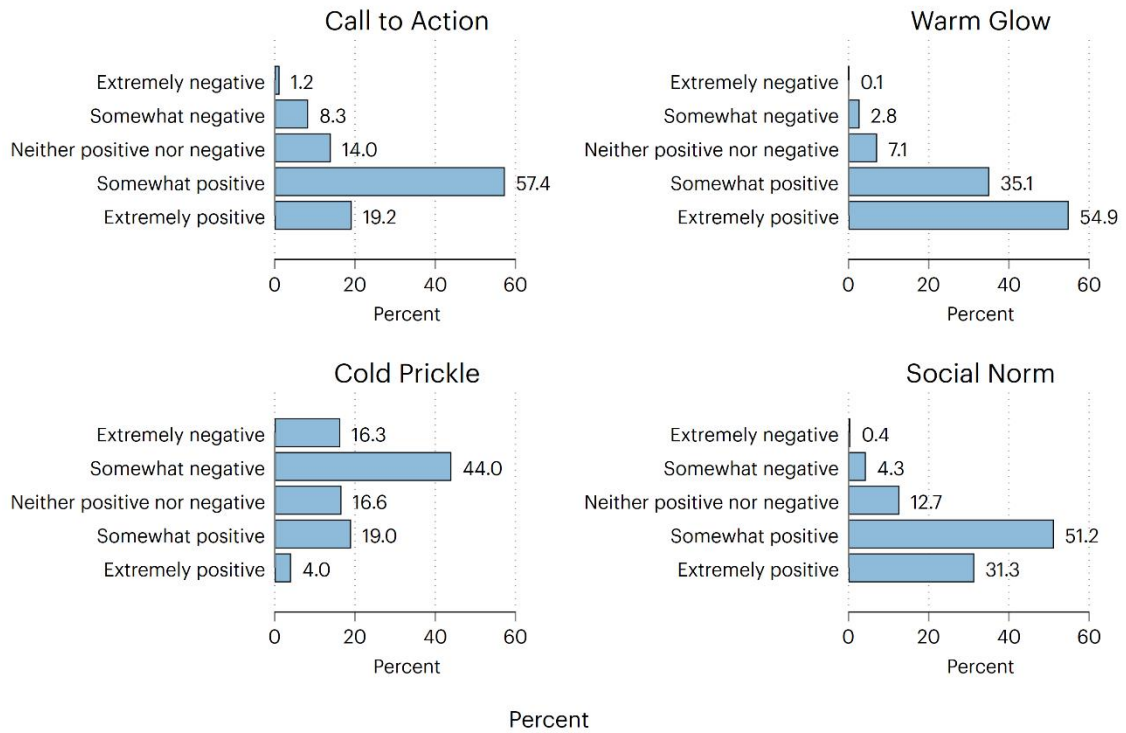
Call to Action

Sentiment	Percent
Extremely negative	1.2
Somewhat negative	8.3
Neither positive nor negative	14.0
Somewhat positive	57.4
Extremely positive	19.2

Warm Glow

Sentiment	Percent
Extremely negative	0.1
Somewhat negative	2.8
Neither positive nor negative	7.1
Somewhat positive	35.1
Extremely positive	54.9

Cold Prickle

Sentiment	Percent
Extremely negative	16.3
Somewhat negative	44.0
Neither positive nor negative	16.6
Somewhat positive	19.0
Extremely positive	4.0

Social Norm

Sentiment	Percent
Extremely negative	0.4
Somewhat negative	4.3
Neither positive nor negative	12.7
Somewhat positive	51.2
Extremely positive	31.3

Percent

Figure A1: Self-reported Perception of Video Sentiment by Treatment Group

Table A2: Video Sentiment Check

	(1) Video Sentiment	
Call to Action	0.000	(.)
Warm Glow	0.814***	(0.062)
Cold Prickle	-1.356***	(0.066)
Social Norm	0.301***	(0.056)
Observations	2,698	

Note: Table presents estimates of an ordered probit regression. The dependent variable is a categorical (ordinal) variable capturing perceived video sentiment (measured on a 5-point scale ranging from "extremely negative" to "extremely positive"). The independent variables are indicators for each treatment condition (the baseline condition being the omitted category). Intercept cut-offs omitted from output. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, we explore whether the treatment videos were successful in manipulating anticipated positive and negative affect, which was assessed after participants reported the perceived sentiment. For our analysis we constructed measures of positive and negative affect based on 10 individual emotion items (see Section 2.5). Figure A2 depicts the average anticipated positive and negative affect scores for each treatment condition with data from Wave 2 only. Table A4 presents the corresponding OLS estimates of equation (1) including biospheric values score as a control variable.
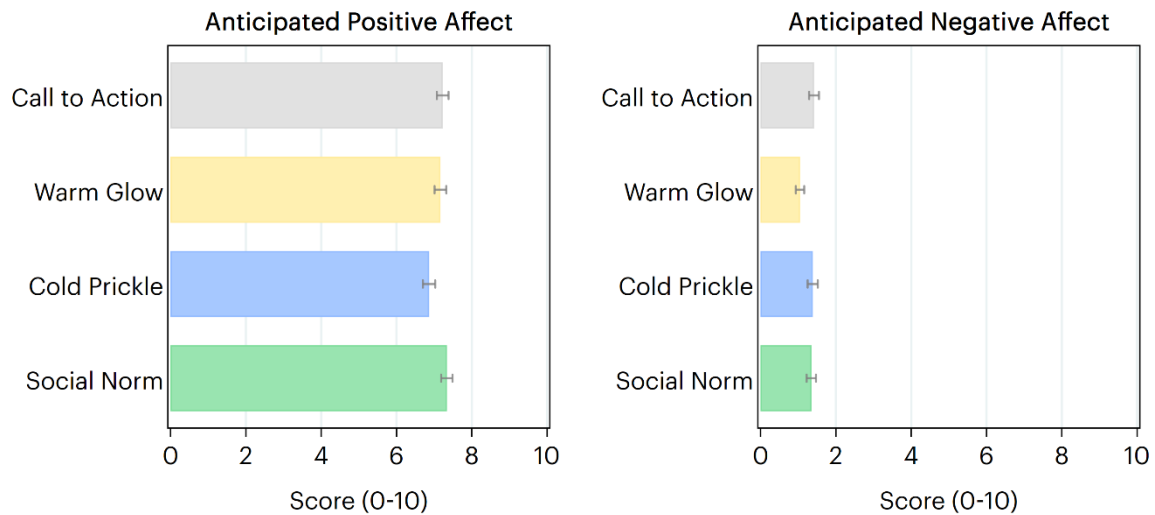

Figure A2: Anticipated Positive and Negative Affect Scores in Wave 2 by Treatment Condition. Score range 0 – 10. N= 2,698.

In line with our hypotheses, we make the following three observations: (1) positive and negative affect are not significantly different in the social norm group, compared to the baseline group, (2) anticipated positive affect was significantly lower in the cold prickle group and (3) anticipated negative affect was significantly lower in the warm glow condition relative to the baseline condition, with both estimates being statistically significant at the 1% level. However, we do not find the hypothesised positive correlation between warm glow messaging and positive affect, nor a positive correlation between cold prickle messaging and negative affect. Although, as previously discussed, participants had an accurate perception of the sentiment portrayed by the videos this analysis suggest that highlighting the emotional consequences of PEB only partially succeeded in experimentally manipulating anticipated affect.

Finally, we explore differences in anticipated emotions for individual emotions items which were used to construct anticipated positive and negative affect scores (Figure A3). We find that positive emotions are lower in the cold prickle group across all five items (happy, proud, hopeful, inspired and warm). Interestingly, we observe that both positive and negative emotions are generally highest in the Call-to-Action condition. With respect to negative emotions, only the warm glow message significantly lowered anticipated emotions of cold, guilt, anxiety, anger and sadness. The cold prickle message was unsuccessful in eliciting these emotions, when compared to the Call-to-Action condition.

Table A3: Direct Effect of Treatments on Anticipated Affect

	(1) Anticipated Positive Affect	(2) Anticipated Negative Affect
Warm Glow	0.031	-0.375***
	(0.098)	(0.087)
Cold Prickle	-0.352***	-0.036
	(0.101)	(0.096)
Social Norm	0.121	-0.073
	(0.097)	(0.092)
Biospheric Values	0.579***	-0.012
	(0.023)	(0.019)
Constant	4.218***	1.484***
	(0.143)	(0.115)
R^2	0.223	0.008
Observations	2698	2698

Notes: OLS estimates of equation (1). The dependent variables are the *anticipated* positive and negative affect scores (ranging from 0 to 10), respectively. *Warm Glow*, *Cold Prickle* and *Social Norm* are treatment indicators identifying individuals randomly assigned to a respective condition. The omitted category is the *Call-to-Action* group. Robust standard errors in brackets.
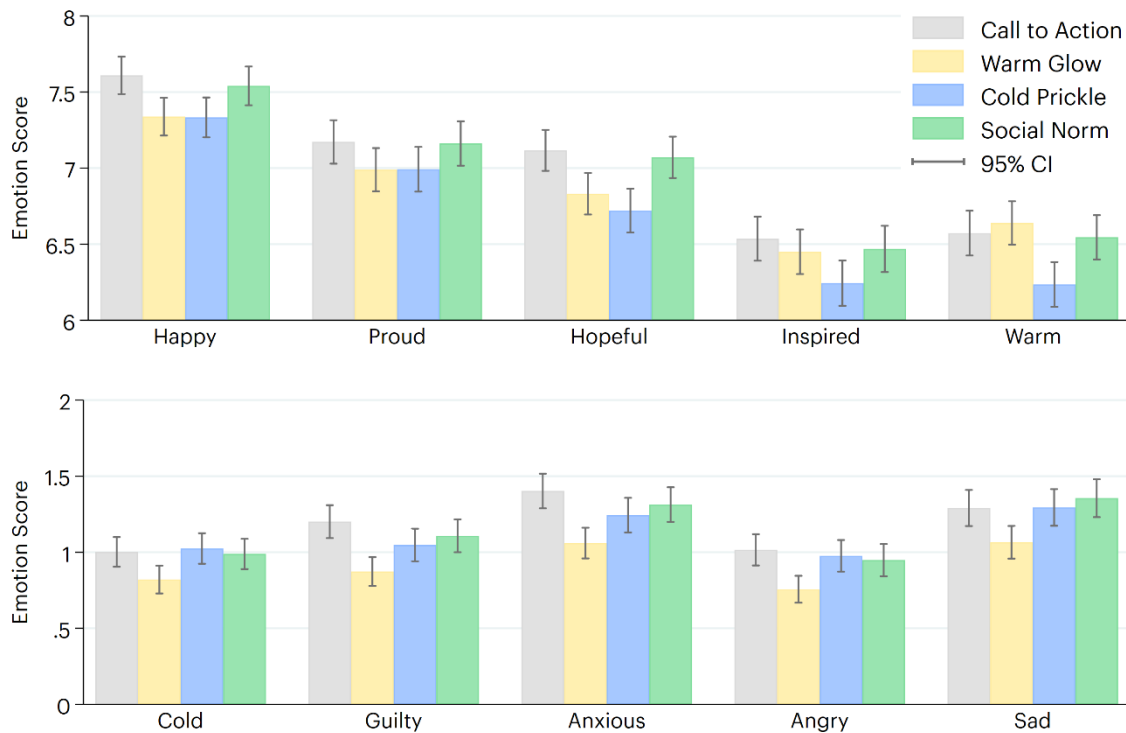* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A3: Anticipated Emotions in Wave 2 by Treatment Condition *Note*: y-axes are truncated. Score range 0 – 10. Error bars indicate 95% confidence intervals. N= 2,698.

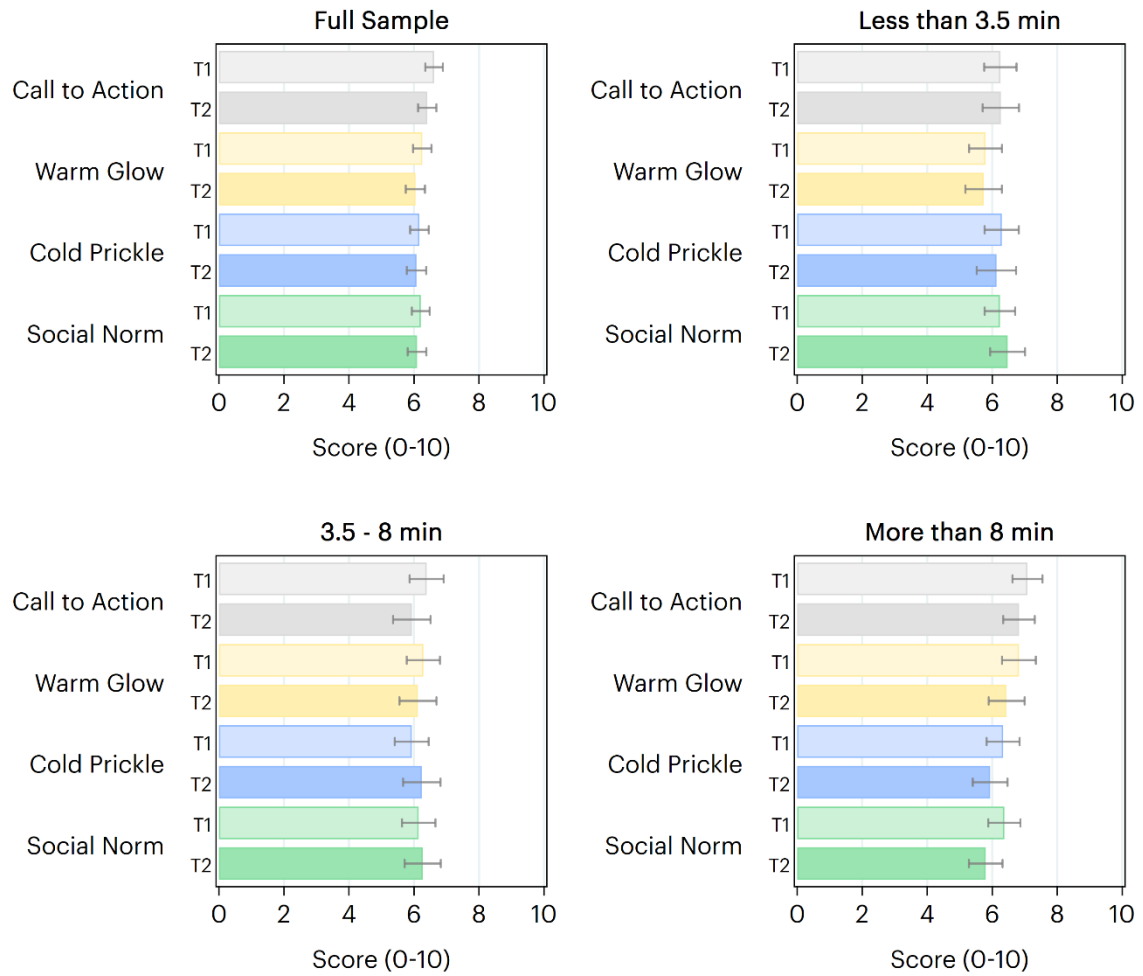## 5. Main Analysis – Additional Figures

Figure A4: Experienced positive affect score in experimental survey (T1) and follow-up (T2) by time invested in the voluntary task (at T1).

6. Main Analysis – Additional Tables

Table A4: Main analysis with additional controls

	(1) Donation (£)	(2) Voluntary Part	(3) Time Invested (Minutes)	(4) Effort Invested (Share Correct)
Warm Glow	-0.133	-0.017	-0.312	-0.008
	(0.088)	(0.027)	(0.194)	(0.013)
Cold Prickle	-0.138	-0.027	-0.289	-0.005
	(0.087)	(0.027)	(0.196)	(0.013)
Social Norm	-0.038	0.008	-0.127	0.013
	(0.090)	(0.027)	(0.198)	(0.012)
Female	0.289***	0.074***	0.612***	0.018*
	(0.063)	(0.020)	(0.139)	(0.010)
Age 35 to 49	0.069	-0.005	0.214	-0.017
	(0.076)	(0.023)	(0.165)	(0.010)
Age 50 or older	-0.244***	-0.119***	-0.397**	-0.063***
	(0.077)	(0.024)	(0.176)	(0.014)
Income	0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Degree	0.098	0.027	0.185	0.008
	(0.064)	(0.020)	(0.141)	(0.009)
Biospheric Values	0.131***	0.046***	0.317***	-0.005
	(0.017)	(0.006)	(0.037)	(0.003)
Constant	0.316***	0.192***	0.725***	0.901***
	(0.121)	(0.040)	(0.267)	(0.021)
R^2	0.036	0.041	0.037	0.031
Observations	2,696	2,696	2,696	1,246

Notes: Omitted base categories are the *Call-to-Action* condition, male participants, participants aged under 35 and no higher education degree. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Cragg-Hurdle Model

	(1) Cragg-Hurdle Estimates		(2) Marginal Effects	
Outcome Model				
Warm Glow	-0.270	(0.225)	-0.071	(0.059)
Cold Prickle	-0.252	(0.226)	-0.066	(0.059)
Social Norm	-0.180	(0.220)	-0.048	(0.058)
Female	0.429**	(0.174)	0.243***	(0.065)
35 to 49	0.292	(0.183)	0.101	(0.074)
50 or older	0.116	(0.224)	-0.168**	(0.080)
Income	-0.000	(0.000)	-0.000	(0.000)
Degree	0.126	(0.168)	0.078	(0.064)
Biospheric Values (scale)	0.109**	(0.055)	0.134***	(0.020)
Constant	0.770*	(0.417)		
Selection Model				
Female	0.146***	(0.052)		
35 to 49	0.024	(0.059)		
50 or older	-0.220***	(0.068)		
Income	-0.000	(0.000)		
Degree	0.050	(0.052)		
Biospheric Values (scale)	0.116***	(0.016)		
Baseline Ability	0.186***	(0.025)		
Donate a few times a year	0.159**	(0.063)		
Donate about once a month (or more)	0.322***	(0.081)		
Donate about once a week (or more)	0.382***	(0.126)		
Constant	-1.351***	(0.123)		
Ln sigma	0.747***	(0.039)		
Observations	2,696		2,696	

Notes: Robust standard errors in parentheses. The selection model models the decision to continue to the voluntary part of the study. Predictors in the selection model include socio-demographic variables, baseline task ability and baseline donation behaviour. Ln sigma refers to the estimated standard deviation of the error term in the outcome model.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Direct Effect of Treatment Messages on Experienced Positive Affect

	(1) Full Sample	(2) Less than 3.5 min	(3) 3.5 - 8 min	(4) More than 8 min
Warm Glow	-0.303	-0.457	-0.098	-0.259
	(0.201)	(0.355)	(0.363)	(0.315)
Cold Prickle	-0.416**	0.041	-0.455	-0.750**
	(0.197)	(0.351)	(0.349)	(0.326)
Social Norm	-0.336*	-0.014	-0.242	-0.708**
	(0.199)	(0.346)	(0.348)	(0.341)
Biospheric Values	0.454***	0.477***	0.413***	0.437***
	(0.048)	(0.079)	(0.083)	(0.088)
Constant	4.112***	3.718***	4.142***	4.617***
	(0.286)	(0.478)	(0.477)	(0.524)
R^2	0.072	0.089	0.060	0.073
Observations	1212	411	398	403

Notes: OLS estimates of equation (1). The dependent variables is the *experienced* positive affect score (ranging from 0 to 10). Column (1) uses data from the full sample of participants who started the voluntary PEB task. Columns (2) – (4) utilize sub-groups of the data, based on the time spent on the PEB task, corresponding to three approximately equally sized groups (terciles). *Warm Glow*, *Cold Prickle* and *Social Norm* are treatment indicators identifying individuals randomly assigned to a respective condition. The omitted category is the *Call-to-Action* group. Robust standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Direct Effect of Treatment Messages on Experienced Negative Affect

	(1) Full Sample	(2) Less than 3.5 min	(3) 3.5 - 8 min	(4) More than 8 min
Warm Glow	-0.360***	-0.297	-0.248	-0.517***
	(0.112)	(0.199)	(0.211)	(0.170)
Cold Prickle	-0.154	-0.234	-0.014	-0.202
	(0.129)	(0.226)	(0.230)	(0.220)
Social Norm	-0.227**	-0.181	-0.199	-0.291
	(0.114)	(0.200)	(0.206)	(0.185)
Biospheric Values	-0.002	-0.016	-0.011	0.026
	(0.026)	(0.044)	(0.044)	(0.044)
Constant	1.033***	1.163***	0.952***	0.937***
	(0.162)	(0.270)	(0.288)	(0.276)
R^2	0.008	0.006	0.006	0.020
Observations	1212	411	398	403

Notes: OLS estimates of equation (1). The dependent variables is the *experienced* negative affect score (ranging from 0 to 10). Column (1) uses data from the full sample of participants who started the voluntary PEB task. Columns (2) – (4) utilize sub-groups of the data, based on the time spent on the PEB task, corresponding to three approximately equally sized groups (terciles). *Warm Glow*, *Cold Prickle* and *Social Norm* are treatment indicators identifying individuals randomly assigned to a respective condition. The omitted category is the *Call-to-Action* group. Robust standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Heterogeneity analysis

	(1) Biospheric Values		(2) Altruistic Values	
Warm Glow	-0.073	(0.116)	-0.003	(0.115)
Cold Prickle	-0.229**	(0.114)	-0.083	(0.116)
Social Norm	0.007	(0.119)	-0.029	(0.113)
Above Median	0.367***	(0.129)	0.412***	(0.131)
Warm Glow × Above Median Values	-0.146	(0.179)	-0.319*	(0.179)
Cold Prickle × Above Median Values	0.186	(0.177)	-0.120	(0.179)
Social Norm × Above Median Values	-0.074	(0.182)	-0.008	(0.183)
Constant	1.030***	(0.084)	1.015***	(0.080)
Observations	2,698		2,698	

Standard errors in parentheses
$^{*} p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$

7. Equivalence Tests

In equivalence testing, the researcher decides on a smallest effect size of interest (SESOI or Δ), where if the effect size is no larger than that value it can be considered negligible (List et al., 2011). As our primary outcome variable (donations) is obtained from a novel pro-environmental effort task, there exists no clear precedent on which to base the SESOI. In such a case, in which no theoretical or practical boundaries exist that indicate a meaningful effect size, a common approach is to utilise the minimum detectable effect size (MDE) obtained from ex-post power analysis as the SESOI (Lakens, 2017). For our primary outcome, we compute a MDE of £0.26 at 80% power and a 5% significance level, using the observations corresponding to our main cross-sectional analysis sample. We round up this value to £0.30 which corresponds to 12 correctly completed trials (or approximately 30 seconds extra time spent on the task) and represents a small treatment effect (Cohen's Delta = 0.18). Results from a two one-sided tests (TOST) procedure (Schuirmann, 1987) are shown in Table A8.

Table A8: Equivalence Tests

	β	s.e.	H_0^+		H_{01}^-		H_{02}^-		Rej H_0^+	Rej H_0^-	Relevance
			t	p	t1	p1	t2	p2			
Warm Glow	-0.126	0.088	-1.430	0.153	4.835	0.000	1.974	0.024	No	Yes	Equivalence
Cold Prickle	-0.122	0.088	-1.393	0.164	4.810	0.000	2.024	0.022	No	Yes	Equivalence
Social Norm	-0.022	0.090	-0.248	0.804	3.572	0.000	3.077	0.001	No	Yes	Equivalence

Note: Table presents the test for difference (H_0^+) of the treatment effects β and two one-sided tests for the equivalence of means (H_0^-) within a symmetric equivalence interval ($\Delta = 0.3$), at a 5% significance level ($\alpha = 0.05$). Rej H_0^+ and Rej H_0^- report whether the test was rejected given pre-specified values of Δ and α. Relevance reports the conclusion drawn from the combined tests for difference and equivalence.

References

Allcott, H., Rogers, T., 2014. The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. Am. Econ. Rev. 104, 3003–3037. https://doi.org/10.1257/aer.104.10.3003

Brandon, A., Ferraro, P.J., List, J.A., Metcalfe, R.D., Price, M.K., Rundhammer, F., 2017. Do the Effects of Social Nudges persist? Theory and Evidence from 38 Natural Field Experiments.

Brosch, T., 2021. Affect and emotions as drivers of climate change perception and action: a review. Curr. Opin. Behav. Sci. 42, 15–21. https://doi.org/10.1016/j.cobeha.2021.02.001

Gravert, C., Olsson, L., 2021. When nudges aren't enough : Norms , incentives and habit formation in public transport usage ☆. J. Econ. Behav. Organ. 190, 1–14. https://doi.org/10.1016/j.jebo.2021.07.012

Hicks, R., Tingley, D., 2011. Causal mediation analysis. Stata J. 11, 605–619. https://doi.org/10.1177/1536867x1201100407

Imai, K., Keele, L., Tingley, D., 2010a. A General Approach to Causal Mediation Analysis. Psychol. Methods 15, 309–334. https://doi.org/10.1037/a0020761

Imai, K., Keele, L., Yamamoto, T., 2010b. Identification, inference and sensitivity analysis for causal mediation effects. Stat. Sci. 25, 51–71. https://doi.org/10.1214/10-STS321

King, G., Tomz, M., Wittenberg, J., 2000. Making the Most of Statistical Analyses: Improving Interpretation and Presentation. Am. J. Pol. Sci. 44, 347. https://doi.org/10.2307/2669316

Lakens, D., 2017. Equivalence Tests : A Practical Primer for t Tests , Correlations , and Meta-Analyses. Soc. Psychol. Personal. Sci. 8, 355–362. https://doi.org/10.1177/1948550617697177

List, J.A., Sadoff, S., Wagner, M., 2011. So you want to run an experiment , now what ? Some simple rules of thumb for optimal experimental design. Exp. Econ. 14, 439–457. https://doi.org/10.1007/s10683-011-9275-7

Schuirmann, D.J., 1987. A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability. J. Pharmacokinet. Biopharm. 15.