**Supplementary Materials for**:

Deviancy Aversion and Social Norms

Anton Gollwitzer1,2 Cameron Martel3

Anna Heinecke4 John A. Bargh5

1BI Norwegian Business School

2Max Plank Institute for Human Development

3Massachusets Institute of Technology

4AMEOS Klinikum für Forensische Psychiatrie und Psychotherapie Neustadt

5Yale University

\*To whom correspondence may be addressed: Anton Gollwitzer, BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway.

Email: [anton.gollwitzer@gmail.com](mailto:anton.gollwitzer@gmail.com)

**Data Availability**

The datasets, analysis files, and all verbatim materials of the presented studies are available at the following URL: **https://osf.io/tzqwy/?view\_only=6503ff2c468e4bb691fd4ef184feb531**

**Study 1**

**Supplemental Methods**

**Control Measures.** Additional control variables included need for closure (Webster & Kruglanski, 1994), disgust propensity and sensitivity (van Overveld et al., 2006), political orientation, and socially desirable responding (Haghighat, 2007). See Verbatim Materials for exact items.

**Attention Check.** We indirectly assessed participants’ attention via the following item: “People vary in the amount they pay attention to these kinds of surveys. Some take them seriously and read each question, whereas others go very quickly and barely read the questions at all. If you have read this question carefully, please write the word yes in the blank box below labeled other. There is no need for you to respond to the scale below.” Participants who failed to write “yes” in response were excluded.

**Supplemental Results**

**Raw Correlations.** The raw correlations of Study 1 can be found below (Table S1).

**Table S1**

*Output of Linear Regressions (Raw Links). Deviancy Aversion Predicted Social Norm Indicators Across Varying Measures.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Dependent Variables | | | | | |
| Study 1 (*N =* 243) | Negative Affect: Fictional  Social Norm Violations | Negative Affect:  Non-Fictional Social Norm Violations | Negative Affect: Littering | Self-Reported Behavior: Following Social Norms | Self-Reported Behavior: Breaking Social Norms | Social Norm Espousal |
| **Deviancy Aversion** | ***B* = .601**  ***p* < .001** | ***B* = .437**  ***p* < .001** | ***B* = .299**  ***p* < .001** | ***B* = .414**  ***p* < .001** | ***B* = -.260**  ***p* < .001** | ***B* = .222**  ***p* < .001** |

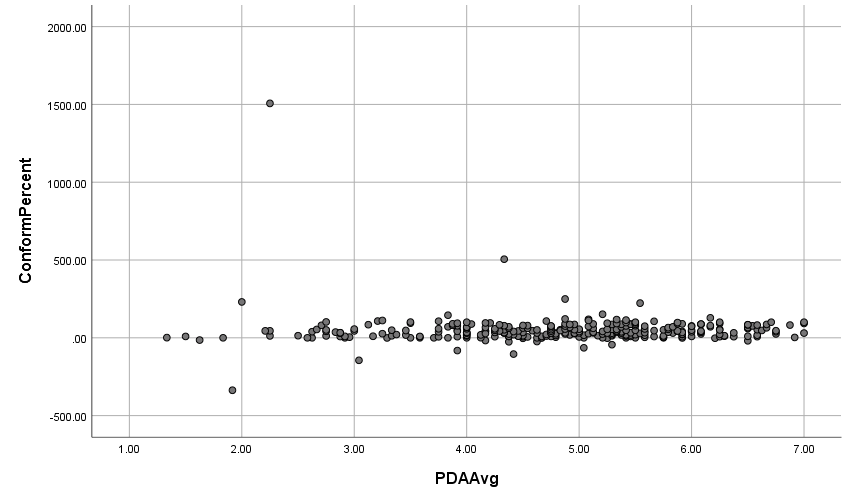
*Note*. *B* = Unstandardized coefficient.

**Study 2**

**Deviations From Pre-Registration**

There were several deviations from the pre-registration of Study 2. These are listed below:

1. Political orientation was listed as a control variable in the pre-registration. Unfortunately, due to experimenter, we failed to assess political orientation. The results, however, are unlikely to have changed given that deviancy aversion predicted norm adherence when controlling for political orientation in the other reported studies.
2. We forgot to list aversion towards unbroken patterns as a control variable on the pre-registration. The results remained consistent when not controlling for aversion towards unbroken patterns, *p* = .015. We chose to include the control to be consistent with the other reported studies.
3. We did not pre-register that we would exclude extreme outliers on the conformity task. This was necessary, however, given that four participants responded in such a manner (e.g., the Eiffel Tower is “600000000” feet tall) that their conformity values skewed the distribution to an extreme extent (see Figure S1).



*Figure S1.* Graphical representation of the extreme outliers in Study 2.

1. We also planned to examine conformity assessed via the absolute difference between participants’ responses after receiving feedback and the average of this feedback (others’ responses). When calculating these difference scores, however, we found almost no reliability between the three estimation tasks. That is, the measure exhibited little to no internal reliability, *ω* = .39. As such, these analyses were abandoned.

**Supplemental Methods**

**Negative Affect: Non-Fictional Norm Violations.** We also assessed participants’ negative affect in response to non-fictional norm violations. The measure of Study 1 was adapted to include positive response-items: Two response-items per vignette: “Negative,” and “Positive.” 1 = *Strongly Disagree* to 7 = *Strongly Agree*. The positive item was reverse-coded.

**Supplemental Results**

**Raw Correlation.** The raw link of Study 2 can be found below (Table S2).

**Table S2**

*Output of Linear Regression (Raw Link). Deviancy Aversion Predicted Greater Conformity.*

|  |  |
| --- | --- |
| Study 2 (*N* = 306) | Conformity  (Accuracy-Oriented Estimation Task) |
| **Deviancy Aversion** | ***B* = 5.12, *β* = .171, *p* = .004** |

*Note*. *B* = Unstandardized coefficient. *β* = Standardized coefficient.

**Conformity: Higher Versus Lower Feedback.** As pre-registered, we examined whether our findings remained consistent across higher and lower feedback (others’ estimations being higher or lower than the ground truth). We conducted a repeated-measures GLM with estimation task (height of the Eiffel tower, weight of a cow, number of marbles) as a within-participants variable, Feedback (high versus low feedback—depending on the specific participant and estimation task; categorical variable) as a within-participants predictor, and participants’ deviancy aversion, aversion to unbroken patterns, novelty aversion as predictors. Though Feedback did predict conformity in this model—in that high feedback predicted greater conformity, *B* = 7.05, *p* = .027, accounting for this variance did not change the link between deviancy aversion and conformity, *B* = 4.82, *p* = .008. Moreover, indicating that our findings are consistent across high and low feedback, an interaction between Feedback (high versus low) and deviancy aversion predicting conformity was not observed, *p* = .989.

**Negative Affect: Non-Fictional Norm Violations**. Replicating Study 1, deviancy aversion predicted greater negative affect towards non-fictional norm violations, *p* < .001 (Table S3; Figure S2). Notably, the observed effect-size was similar to Study 1 despite that Study 2’s measures included reverse-coded items, Study 1: *β* = .38 versus Study 2: *β* = .42 (see Table 1 and Table S3). Furthermore, when separately examining the “negative” and “positive” norm response-items, deviancy aversion predicted increased negative affect towards norm violations, *β* = .318, *p* < .001, and decreased positive affect towards norm violations, *β* = -.393, *p* < .001, as expected. Finally, neither aversion towards unbroken pattern nor novelty aversion—which entailed identical response items as deviancy aversion—predicted participants’ responses to norm violations, *p*s > .058 (Table S3).

**Table S3**

*Output of Multivariate Linear Regressions. Deviancy Aversion Predicted Greater Negative Affect Towards Non-Fictional Norm Violations.*

|  |  |
| --- | --- |
|  | Negative Affect in Response to  Non-Fictional Norm Violations |
| *Descriptive Statistics* | *M* = 5.03, *SD* = 1.05, *ω* = .87 |
| *Variance Explained* | *R*2 = .46 |
| **Deviancy Aversion** | ***B* = .376, *β* = .417, *p* < .001** |
| Aversion to Unbroken Patterns | *B* = .010, *β* = .011, *p* = .840 |
| Novelty Aversion | *B* = .099, *β* = .102, *p* = .059 |

*Note*. *B* = Unstandardized coefficient. *β* = Standardized coefficient.



*Figure S2.* Partial residual effect plot. Deviancy aversion predicted greater negative affect towards non-fictional norm violations. Error bands: *CI*s (using geom\_ribbon in R).

**Study 3**

**Supplemental Methods**

**COVID-19 Knowledge Quiz.** We also included a 15-item multiple choice quiz on facts about COVID-19 (Gollwitzer et al., 2021). Participants were asked a series of questions such as “The virus stays on plastic surfaces for a longer time than cardboard surfaces” (1 = *True*, 2 = *False*), and “Which virus has killed more people? (1 = *SARS*, 2 = *H5N1 (“Bird Flu”),* 3 = *Covid-19*. Correct answers were given one point and incorrect answers zero points; participants’ final score was the sum of their points. If deviancy aversion predicts social distancing, it may also predict knowing more information about COVID-19.

**Supplemental Results**

**Raw Correlation.** As pre-registered, we also examined the raw correlation between deviancy aversion and social distancing. Consistent with the findings reported in the main text, deviancy aversion predicted greater social distancing, *r*(282) = .21, *p* < .001.

**COVID-19 Knowledge Quiz.** In line with the social distancing results, deviancy aversion also predicted greater knowledge about COVID-19 as assessed by our COVID-19 quiz, *p* < .001.[[1]](#footnote-1) And again, none of the other included variables positively predicted COVID-19 knowledge. Arguing against demand bias, the COVID-19 quiz was a performance rather than self-report measure. Additionally, and as was the cased for distancing norms, aversion to unbroken patterns and novelty aversion (which were both measured similarly to deviancy aversion) did not predict COVID-19 knowledge, *p*s > .216 (Table S4).

**Table S4**

*Output of Linear Regressions. Deviancy Aversion Predicted Greater Knowledge About COVID-19.*

|  |  |
| --- | --- |
|  | COVID-19 Knowledge Performance |
| *Descriptive Statistics* | *M* = 11.13, *SD* = 2.32, *ω* = .70 |
| *Variance Explained* | *R*2 = .34 |
| **Deviancy Aversion** | ***B* = .386, *β* = .254, *p* < .001** |
| Aversion to Unbroken Patterns | *B* = .125, *β* = .075, *p* = .217 |
| Novelty Aversion | *B* = .077, *β* = .057, *p* = .371 |
| Intolerance of Ambiguity | *B* = -.421, *β* = -.148, *p* = .019 |
| Political Orientation | *B* = -.148, *β* = -.155, *p* = .009 |

*Note*. *B* = Unstandardized coefficient. *β* = Standardized coefficient.

**Study 4**

**Supplemental Methods**

**Identifying cheating participants.** Cheaters were identified via IP address. An IP address is a 32-bit number that uniquely identifies a network interface on a machine. Specifically, IP addresses are typically written in decimal digits, formatted as four 8-bit fields separated by periods We pre-registered that IP addresses that had the same numbers for the first three 8-bit fields would be classified as the same individual.

**Supplemental Results**

**Raw Correlation.** The raw links of Study 4 can be found below (Table S5).

**Table S5**

*Logistic and Linear Regressions (Raw Links).*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Study 4 (*N* = 1,118) | Cheating  (Average) | Cheating  (First Response) |  | Number of  Times Cheated  (Average) | Number of  Times Cheated  (First Response) |
| Deviancy Aversion | *B* = -.35, *z* = -4.76, *p* < .001 | *B* = -.46, *z* = -6.46, *p* < .001 |  | *B* = -.16, *p* < .001 | *B* = -.22, *p* < .001 |

*Note*. *B* = Unstandardized coefficient.

**More Stringent IP Address Classification.** To test the robustness of the findings we exploratorily re-conducted our analyses when classifying cheaters in a more conservative manner. Specifically, we classified any IP addresses that had the same number for the first *two* 8-bit fields as the same participant. Doing so ensures that we did not accidentally treat the same participant as different participants, in turn artificially increasing power and in turn Type I error. This reduced the number of cheaters from 66 to 41. Notably, deviancy aversion still predicted reduced cheating for both the average and first taken analyses, *p*s < .027. Deviancy aversion no longer predicted the number of times a participants cheated, however, *p* = .131 (average), and *p* = .094 (first taken). Further analyses revealed that this was likely due to the inclusion of an outlier who completed the study 210 times. Indeed, removing this outlier revealed more hypothesis conform results, *p* = .056 (average), and *p* < .001 (first taken). Overall, these results align with those reported in the main text when using the less stringent IP classification.

**Study 5**

**Supplemental Methods**

**Deviancy Aversion.**We applied a motivational deviancy aversion paradigm validated by Gollwitzer and colleagues (2020). In line with active goals intruding on attitudes and judgment (e.g., Lewin, 1917; Ferguson & Porter, 2009), participants were induced with the goal to come up with negative (positive) attributes of nonsocial deviancy and positive (negative) attributes of nonsocial deviancy, and while this goal was active—before goal completion was reached—they completed several social norm measures. Specifically, participants in the high (low) deviancy aversion condition were told they would be entered into a raffle for a $20 reward if they came up with a larger number of negative (positive) attributes of deviancy and positive (negative) attributes of patternicity than 75% of other participants.

All participants thereafter completed two attention checks and their self-reported motivation to perform well on the task: “I feel motivated to come up with negative (positive) attributes about things that break a pattern, and positive (negative) attributes about things that follow a pattern.” Likert-scale: 1 = *Not at all agree* to 7 = *Strongly agree*.

Participants were then told that they would report the attributes they had come up with after answering some questions and thus, should come up with these attributes while completing these questions. Finally, participants were told, depending on high (versus low) deviancy aversion condition, to “start thinking of negative (positive) words that are associated with things that break a pattern (e.g., disruptive [exciting]), and positive (negative) words that are associated with things that follow a pattern (e.g., organized [boring]).” To keep the goal active throughout the study, this prompt was repeated (the term “start” was changed to “continue”) between each of the norm measures. See Verbatim Materials.

**Supplemental Results**

**Negative Affect Towards Norm Following.** As in Study 1, we also assessed participants negative affect towards fictional norm following. As expected, participants in the high (versus low) deviancy aversion condition exhibited lower negative affect towards such norm following, *B* = -0.46, *β* = -0.15, *p* = .002. Given that the measure exhibited low reliability and high skew in a number of the reported studies, however, we report these results here in the Supplements.

**Robustness Checks.** The robustness tests are described in greater detail than in the main text below:

First, none of the observed effects were moderated by participants’ tendency to engage in anthropomorphism; adding the two-way interaction between Deviancy Aversion and anthropomorphism to each of the six GLM models did not reveal significant interactions, *p*s> .058. Second, as pre-registered, we added the two-way interactions between Deviancy Aversion and participants’ self-reported motivation to perform well on the deviancy word-task to each of the six main GLM models. In the six models, three interactions were significant, *p*s< .043, one was marginal, *p* = .058, and two were non-significant, *p*s> .115. Unpacking these interactions revealed that the manipulation tended to be more powerful for participants who reported being motivated on the task. Given these results, and as pre-registered, we examined whether the significant effects of Deviancy Aversion on norm adherence observed in the original GLM models remained when including these two-way interactions; importantly, they did so, *p*s < .005. These results suggest that though self-reported motivation plays a role in the effect of deviancy aversion on norm adherence, a significant portion of the impact of deviancy aversion on norm adherence occurs outside of participants’ deliberate intention or motivation.

Third, and in line with Studies S1 and S2, mediation analyses revealed that the significant effects of the manipulation on norm adherence occurred via heightened levels of deviancy aversion, *β*s> .12, 95% CI[.05 < *β* < .35] and not via participants’ novelty aversion, *β*s< .01, 95% CI[-.02 < *β* < .04]. Furthermore, mediation analyses on the two non-significant norm adherence measure—espousal and societal tightness—revealed significant indirect effects of the manipulation on these measures via participants’ degree of deviancy aversion; that is, though we did not observe direct effects of Deviancy Aversion on these measures, indirect links appear to exist, *B* = .23, 95% CI[.12, .35], and *B* = .18, 95% CI[.10, .27], respectively.

**Study S1: Causal Effects of Deviancy Aversion on Social Norms**

Study S1 examined whether deviancy aversion has a causal effect on social norms. Specifically, we tested whether generating and reflecting on the negative versus positive aspects of low-level deviancy heightens norm adherence on several of the norm measures included in Study 1 (e.g., choosing to follow norms). We also assessed participants’ aversion towards negative but not deviant stimuli to control for the possibility that our manipulation simply induced negative affect rather than specifically deviancy aversion.

**Participants**. A supplemental study nearly identical to Study S1 (which replicated the results of Study S1 but did not include controls so is not reported here; *f* = .29), revealed we needed 127 participants for 90% power. We aimed to recruit 180 participants, and ended with 181 participants (MTurk; 98 Female; *M*age= 33.70, *SD*age= 10.61). Twenty-eight responses were excluded for attention failures, and six for completing the survey twice. With this final sample (*N* = 147), we had 90% power to observe an effect-size of *f* = .27. The study was not pre-registered. See Verbatim Materials [here](https://osf.io/tzqwy/?view_only=6503ff2c468e4bb691fd4ef184feb531).

**Attrition.** In order to reduce participant attrition, as recommended by Zhou and Fishbach (2016), participants were asked to confirm that they did not mind open-ended questions before beginning the experiment. Participants read and responded to the following attrition prompt: “This is an anonymous survey consisting of multiple questions. A few questions are open-ended questions where you need to type a few sentences or a short paragraph or two. If a sizable number of people quit a survey halfway, however, the data quality of that survey is compromised. Please make sure you do not mind open-ended questions before taking this survey and please do your best to complete the *entire*survey. Your input is important to us and we very much appreciate your participation!” Participants then read and responded to the prompt “If you have read the above paragraph and choose to complete the survey please write ‘I agree’ in the textbox below. Thank you.” Attrition rates did not greatly differ depending on condition (high deviancy aversion condition: *n* = 69; low deviancy aversion condition: *n* = 76), *p* = .618**.**

**Deviancy Aversion.** Participants were randomly assigned to the high (versus low) deviancy aversion condition. Depending on condition, they read: “Think of things/objects that break a pattern, are out of line, and create disorder. What are three negative (positive) attributes of such things/objects?…For example, what is something negative (positive) about a few blueberries in a bowl of many strawberries, or what is something negative (positive) about the images below?” The two images depicted two broken geometric patterns from Study 1. Participants thereafter were told to reflect in vivid detail on how negatively (positively) they feel towards nonsocial deviancy. See Verbatim Materials.

**Norm Adherence.** We included three measures: Negative affect towards fictional norm violations, negative affect towards non-fictional norm violations, and self-reported norm behavior. These measures were as in Study 1 except for the inclusion of some reverse-coded items (see Verbatim Materials).

**Control Variables.** Negativity aversion—participants’ aversion towards negative but not deviant stimuli—was assessed via aversion towards scenes of poor weather—poor weather, though negative, is not generally considered deviant (Gollwitzer et al., 2017; Gollwitzer & Clark, 2018). Participants read: “At the moment, how uncomfortable do the following scenes make you feel?” and responded to 4 images of poor weather and 4 images of good weather. See Verbatim Materials.

**Manipulation Check.** Participants’ responses towards two broken and two unbroken patterns of geometric shapes functioned as the manipulation check: “How ‘positive’ is the above image?” and “How ‘negative’ is the above image?” 1 = *Not at all* to 7 = *Very*. Positive item reverse-coded.

**Attention Checks.** Participants completed an attention check regarding the manipulation, a face-valid attention check, and the attention check of Study 1 (see Verbatim Materials).

**Procedure.** After the manipulation, participants completed the norm measures (randomized order). Participants then completed the negativity aversion measure, and thereafter the deviancy manipulation check. Finally, participants completed attention checks, a social desirability measure (see Study 1), and demographics.

**Results**

A GLM with Deviancy Aversion (high versus low) as a categorical predictor indicated that the manipulation successfully altered participants’ deviancy aversion, *ω* = .93, *p* < .001, ηp2 = .196 (Table S6).

Extending Gollwitzer et al. (2017), two GLMs with Deviancy Aversion as a predictor and the norm violation measures as the outcome variables, respectively, found that increasing (versus reducing) deviancy aversion causally heightened participants’ negative affect in response to fictional, *p* < .001, and non-fictional norm violations, *p* = .003 (Table S6; Figure S3). Additionally, inducing high (versus low) deviancy aversion also incited increased norm following, *p* = .010, η2 = .045, and decreased norm breaking, *p* = .002, η2 = .089 (Table S6; Figure S3). Social desirability did not moderate any of these findings, *p*s > .215.

Robustness analyses supported our results. First, our findings were not driven by general condition differences in negative mood or attitudes. For one, controlling for participants’ negativity aversion (calculated by subtracting their aversion towards good weather, *ω* = .93, from their aversion towards bad weather, *ω* = .89) in our GLM models did not change our findings, *p*s< .007. For another, separate GLM models found no effect of Deviancy Aversion on negativity aversion, *p* = .118, and found a greater effect of Deviancy Aversion on the deviancy aversion manipulation check than on negativity aversion, *p* < .001. Finally, mediation analyses using the SPSS PROCESS macro (Hayes, 2012; default settings applied; Model 4) found that the effect of Deviancy Aversion on norm responding occurred via heightened levels of the deviancy aversion manipulation check (indirect effects: fictional norms, *B* = .51, 95% CI[.23, .87]; non-fictional norms, *B* = .47, 95% CI[.28, .70]) and not by altering participants’ degrees of negativity aversion (fictional norms, *B* = .00, 95% CI[-.06, .10]; non-fictional norms, *B* = .02, 95% CI[-.02, .10]).

Second, participants may have anthropomorphized or perceived the geometric shapes as agentic. Arguing against this possibility, a small minority of participants reported social examples of deviancy in response to the manipulation (11.6%, *n* = 17; e.g., it makes me feel uncomfortable when someone skips a line), and our findings remained when excluding these participants, .001 < *p*s< .023, .040 < η2 < .077. Similarly, only a small minority of participants referenced agency or intentionality, for instance, that the deviating triangle was “trying to be different” or that someone had intentionally broken the pattern (9.2%, *n* = 12); importantly, additionally excluding these participants did not change our results, .001 < *p* < .050, .033 < η2 < .084. Finally, the degree of social-content in participants’ responses (quantified via NLP methods; LIWC; Pennebaker et al., 2015) neither accounted for, *p*s < .012, nor moderated our effects, *p* = .176.[[2]](#footnote-2)

**Additional Analyses: Negative Affect Towards Social Norm Following.** As predicted, increasing (versus reducing) deviancy aversion also reduced participants’ negative affect towards social norm following in a fictional society (*α* = .48), *p* = .006, ηp2 = .052. We do not include this finding in the main text given the poor inter-item reliability of the measure. This low inter-item reliability was driven by the inclusion of a positive reverse-coded item (‘happy’). The effect of deviancy aversion separately for the negative items and this positive item was: Negative items: *F*(1, 145) = 6.87, *p* = .010, ηp2 = .045, Positive item: *F*(1, 145) = 1.22, *p* = .272, ηp2 = .008 (though not significant, the latter result was in the predicted direction). We do not find this null effect particularly meaningful, however, given that the other measures of social norm adherence exhibited good inter-item reliability and exhibited hypothesis-supporting results.

**Table S6**

*Means, SDs, and Results of the Nonsocial Deviancy Aversion Manipulation in Study S1.*

|  |  |  |  |
| --- | --- | --- | --- |
|  | High Nonsocial Deviancy Aversion | Low Nonsocial Deviancy Aversion | Significance Test |
| **Study S1** | *n* = 72 | *n* = 75 |  |
| Manipulation Check | *M*, *SD* | *M*, *SD* |  |
| Deviancy Aversion | 2.42, 1.71 | 0.87, 1.46 | *F*(1, 145) = 35.28, *p* < .001, η2 = .196 |
| Dependent Variables |  |  |  |
| Negative Affect Towards  Fictional Norm Breaking (*ω* = .79) | 4.17, 1.64 | 3.15, 1.42 | *F*(1, 145) = 16.32, *p* < .001, η2 = .101 |
| Negative Affect Towards  Non-Fictional Norm Breaking (*ω* = .90) | 5.21, 0.98 | 4.68, 1.18 | *F*(1, 145) = 8.93, *p* = .003, η2 = .058 |
| Self-Reported  Norm Following Behavior (*α* = .68) | 4.99, 1.64 | 4.30, 1.56 | *F*(1, 145) = 6.86, *p* = .010, η2 = .045 |
| Self-Reported  Norm Breaking Behavior (*α* = .78) | 2.64, 1.50 | 3.60, 1.59 | *F*(1, 145) = 14.13, *p* < .001, η2 = .089 |



*Figure S3.* The effect of Deviancy Aversion on A) negative affect towards fictional norm violations, B) negative affect towards non-fictional norm violations, C) self-reported norm breaking, and C) self-reported norm following. Error bars: +- 1 *SE*.

**Study S2: Directly Accounting for Negativity Aversion**

Building on Study S1, Study S2 examined whether inducing deviancy aversion heightens norm adherence compared to a closely matched negativity aversion control condition. We examined whether inducing aversion towards deviancy heightens norm adherence as compared to inducing aversion towards negative but not deviant stimuli (aversion towards scenes of poor weather).

**Method**

**Participants**. The power-analysis was based on the smallest effect in Study S1 (*f* = .23). We needed 146 participants to have 80% power. We aimed to recruit 200 participants and recruited 196 participants (MTurk; 119 Female; *M*age= 38.64, *SD*age= 12.62). Fifteen responses were excluded for attention failures or for completing the study twice. With our final sample (*N* = 181), we had 90% power to observe an effect-size of *f* = .24. The study was not pre-registered. Verbatim Materials [here](https://osf.io/tzqwy/?view_only=6503ff2c468e4bb691fd4ef184feb531).

**Aversion.** The deviancy aversion condition was the high deviancy aversion condition of Study S1. The negativity aversion condition was identical except participants reflected on the negative aspects of poor weather instead of deviancy: “Think of cloudy, grey, and overcast weather…” See Verbatim Materials.

**Norm Adherence.** We included the negative affect towards fictional norm violations measure of Study S1.

**Manipulation Checks.** The deviancy aversion manipulation check was as in Study S1. Aversion towards poor weather (over good weather) served as a further manipulation check (see Verbatim Materials).

**Attention Checks.** Attention checks were as in Study S1 except adapted to Study S2’s manipulation.

**Procedure.** Participants completed the manipulation and then the norm adherence measure. They then completed the manipulation checks, attention checks, and demographics.

**Results**

The manipulation was successful. Deviancy aversion (versus negativity aversion) induced higher deviancy aversion, *p* < .001. Additionally, negativity aversion (versus deviancy aversion) induced higher negativity aversion, *p* = .003 (Table S7).

Replicating Study S1, GLMs found that inducing deviancy aversion (versus negativity aversion) heightens participants’ negative affect towards fictional norm violations, *p* < .001, η2 = .081 (Table S7; Figure S4). Additionally, as in Study S1, and as expected, mediation analyses revealed that our effects occurred via heightened deviancy aversion, *B* = .29, 95% CI[.08, .55] and not via negativity aversion, *B* = -.01, 95% CI[-.13, .08].

Table S7

*Means, SDs, and Results of the Aversion Manipulation in Study S2.*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Deviancy Aversion | Negativity Aversion | Significance Test |
| **Study S2** | *n* = 75 | *n* = 106 |  |
| Manipulation Check | *M*, *SD* | *M*, *SD* |  |
| Deviancy Aversion | 2.83, 1.67 | 1.23, 1.73 | *F*(1, 179) = 39.18, *p* < .001, η2 = .180 |
| Negativity Aversion | 0.54, 0.62 | 0.88, 0.85 | *F*(1, 179) = 8.85, *p* = .003, η2 = .047 |
| Dependent Variables |  |  |  |
| Negative Affect Towards Fictional Norm Breaking | 4.32, 1.55 | 3.50, 1.24 | *F*(1, 179) = 15.77, *p* < .001, η2 = .081 |



*Figure S4.* The effect of Deviancy Aversion on negative affect towards fictional norm violations in Study S2. Error bars: +- 1 *SE*.

**Additional Results: Negative Affect to Fictional Norm Following.** As in Study S1, we also assessed participants’ affect in response to fictional norm *following*. Because the results varied across these studies (likely because the measure was largely ceilinged and suffered from low internal reliability) we report these results as additional results. As in Study 1, in Study S2, we found that the deviancy aversion manipulation reduced participants’ negative affect towards the social norm follower, that is, participants in the high deviancy aversion condition exhibited less negative affect towards a fictional norm follower than participants in the negativity aversion condition, *p* = .004.

**Study S3: Linking Deviancy Aversion to Negative Affect**

**Towards Fictional Norm Following**

In Study 1 we assessed negative affect towards fictional norm following. Deviancy aversion surprisingly related positively to such aversion, though this was not significant, *B* = .11, *p* = .071. Additional analyses revealed that this unexpected link was likely driven by a strong floor effect; very few participants endorsed negative affect towards the fictional norm follower (skewness: 2.13, kurtosis: 3.72; accepted limits are ±2). As such, we conducted Study S3 in which we altered the scale of the negative affect towards fictional norm following measure to include reverse items (to reduce skewed responding). We then examined whether doing so reveals the originally expected results. That is, we examined whether deviancy aversion predicts *reduced* negative affect towards fictional norm following, as expected.

**Methods**

**Participants**. The study was conducted on Prolific. We aimed to recruit 100 participants and ended with 101 participants (47 Female; *M*age= 33.73, *SD*age= 12.01). Four responses were excluded for attention failures. Verbatim Materials [here](https://osf.io/tzqwy/?view_only=6503ff2c468e4bb691fd4ef184feb531).

**Deviancy Aversion.** The deviancy aversion measure was that of Study 3.

**Negative Affect: Fictional Social Norm Following**. We assessed negative affect towards a fictional character (a Flurp) following the Flurp norm of living in a blue house: “This Flurp makes me feel…” “Uncomfortable,” “Annoyed,” “Anxious.” 1 = *Not at all agree* to 7 = *Strongly agree*. Importantly, and unlike Study 1, we also included three positively valanced items: “Happy,” “Comfortable,” “Calm.” 1 = *Not at all agree* to 7 = *Strongly agree* (reverse-coded).

**Control Variables.** We assessed aversion towards *un*broken patterns as in Study 2.

**Attention Check.** The attention check of Study 1 was included.

**Procedure.** Participants completed the three deviancy aversion measures (randomized, clustered together, including the unbroken pattern control measure) and the negative affect towards fictional norm following in random order. Participants then completed the attention check and demographics.

**Results**

As hypothesized, and contrary to Study 1, participants’ deviancy aversion predicted *reduced* negative affect towards fictional norm following (a Flurp living in a Blue house), *r*(95) = -.24, *p* = .014. This link remained when additionally controlling for participants’ aversion towards *un*broken patterns, *r*(94) = -.21, *p* = .039. These results indicate that the slightly positive link between deviancy aversion and negative affect towards fictional norm following in Study 1 was spurious or due to the high skewness of the original measure.

**Supplemental References**

Haghighat, R. (2007). The development of the brief social desirability scale (BSDS). *Europe’s Journal of Psychology*, *3*.

Van Overveld, W. J. M., De Jong, P. J., Peters, M. L., Cavanagh, K., & Davey, G. C. L. (2006). Disgust propensity and disgust sensitivity: Separate constructs that are differentially related to specific fears. *Personality and Individual Differences*, *41*, 1241-1252.

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology, 111*, 493-504.

1. Eight additional participants were excluded from the COVID-19 knowledge results as they admitted to looking up some answers to the quiz online. [↑](#footnote-ref-1)
2. The overall interaction effect is reported here. For the individual measures, a significant interaction was observed for the negative affect towards fictional norm violations measure, *p* = .023. Further analyses revealed, however, that the effect of deviancy aversion was stronger for participants whose responses included *fewer* rather than greater socialword-use. [↑](#footnote-ref-2)