

Appendix 1

Pretest of control threat and morality threat manipulation

Prior to the research presented in this article, we conducted a binational pilot study aimed at pretesting manipulation materials. The pilot study questionnaire was translated from English to Polish and German. Following the translation the research team (some of whom are fluent in all three languages) discussed the Polish and German versions in order to ensure their equivalence. The pilot study utilized a between-subjects factorial design with one experimental factor (individual threat) with three levels (morality threat vs. control threat vs. baseline condition). The manipulation materials read as follows:

Morality threat	Control threat	Baseline
People's life is guided by the social rules that define what is good and bad, and what one should or shouldn't do.	People most of the time feel that they have a lot of influence over what happens in their lives.	People like to have positive feelings in everyday life.
However, sometimes, people severely violate these rules, which usually results in negative feelings of guilt and shame.	However, sometimes people recognize that this is not true at all, which usually results in negative feelings of helplessness and anxiety.	However, sometimes people are in a negative mood without any reason, which results in negative feelings of displeasure and uneasiness.
Now, please, spend a moment thinking about two situations from your own life when you broke one or more of these important rules and later felt bad about it.	Now, please, spend a moment thinking about two situations from your own life when you felt that you did not have any influence over some aspect of your life.	Now, please, spend a moment thinking about two situations from your own life when you were in a negative mood without any obvious reason.
Please describe these two events below.	Please describe these two events below.	Please describe these two events below.

Dependent variables

Semantic Differential. Directly after the manipulation participants were asked to indicate how they felt in situations that they described. Their answers were assessed on a 7-point semantic differential scale containing 12 bipolar choices. The different feelings that the participants were asked to evaluate pertained to morality (3 items, e.g., *dishonest-honest*) and control (4 items, e.g., *uncertain-certain*) dimension; there were also 2 items pertaining to judgments of competence (e.g., *competent-incompetent*), 2 items pertaining to warmth (e.g., *friendly-unfriendly*) and one additional filler item (*sad-happy*).

Evaluation by an Unknown Observer. Following the semantic differential scale and a filler task (PANAS; Watson & Clark, 1999), the participants were asked to imagine that they were being evaluated by a stranger who does not know anything about them and who would form their opinion about them based solely on the descriptions of the two events that the participants had provided earlier. The participants were then presented with a list of 22 descriptive items pertaining to control (e.g., *with no influence*), morality (e.g., *dishonest*), competence (e.g., *resourceful*), and warmth (e.g., *friendly*), there were also 3 filler items: *happy*, *depressed*, and *sad*. The participants indicated the extent to which the observer would judge them to possess each of these traits on a scale from 1 = *not at all* to 7 = *very much*.

Results in Germany

Participants. One hundred fifty-five German participants volunteered to participate in the study. The data was collected in University of Leipzig (Germany) facilities, and therefore the sample consisted mostly of University of Leipzig students (92%). Among the participants 59% were female, 40% male and 1% indicated “other” as their gender. Participants’ age ranged from 18 to 42, $M = 23.45$, $SD = 3.75$.

Results. Participants' responses on the semantic differential scale were factor-analyzed using a Principal Component Analysis with an Oblimin rotation. The solution returned 3 factors that explained 28.19%, 22.01%, and 10.10% of the variance. Three items, *dishonest*, *guilty* and *foolish*, loaded equally strongly on more than one dimension and therefore were disregarded. All morality and warmth items loaded on the first factor. The second factor was loaded by control items only. The third factor was loaded by one control item (*uncertain*) and one competence item (*incompetent*). All items loading onto the same factor were averaged to create composite scores. As a result, we obtained three subscales pertaining to morality and warmth (*friendly*, *unfair*, *nice*; $\alpha = .74$); control (*agentic*, *powerless* - *reversed*, *I had a lot of control*; $\alpha = .73$); control and competence (*competent*, *certain*; $\alpha = .46$). Additionally, since we were primarily interested in the effects of the experimental manipulation on the morality and control dimension, we recreated the first factor using the morality item only (*unfair*). A higher score on a given subscale indicated a higher self-evaluation on its respective dimension.

We used the exact same procedure to determine the factorial structure of the scale measuring evaluation by a stranger. All 19 items (i.e., all items except for the filler items) were entered into a Principal Component Analysis with an Oblimin rotation. The solution returned 4 factors that explained 28.93%, 19.08%, 7.49% and 6.03% of variance. Two items, *competent* and *efficient*, loaded equally strongly on more than one dimension and therefore were disregarded. All morality and warmth items loaded on the first factor. The second and third factor pertained to low and high control dimensions respectively. The fourth factor comprised one control item (*uncertain*) and one competence item (*smart*), however since both of them loaded negatively we disregarded this factor as presenting the analysis residuals without any theoretical sense. All items loading onto the same factor were averaged to create composite scores. As a result, we obtained three subscales pertaining to morality and warmth

(*immoral-reversed, good-natured, friendly, fair, untrustworthy-reversed, loyal, sincere, dishonest-reversed, selfless*; $\alpha = .89$); low control (*helpless, weak, powerless, with no control*; $\alpha = .82$) and high control (*agentic, mighty*; $\alpha = .68$). Additionally, since we were primarily interested in the effects of the experimental manipulation on the morality and control dimension, we recreated the first factor using the morality items only (*immoral-reversed, fair, untrustworthy-reversed, loyal, sincere, dishonest-reversed, selfless*; $\alpha = .84$). A higher score on a given subscale indicated a higher self-evaluation on its respective dimension.

In order to verify whether the experimental manipulation proved effective in affecting participants' emotions and their assumption of how they might be judged by a stranger we conducted two mixed model ANOVAs. Since we were specifically interested in the effects on the morality and control dimension, only subscales pertaining to these two dimensions are reported below. They were entered as a repeated measure dependent variable with two (for the first DV: morality, control) or three (for the second DV: morality, low control, high control) levels. The experimental group was entered as a between subjects factor (see Table A in Appendix 1).

With regard to the first dependent variable, the semantic differential scale, neither the main effect of experimental group was significant, $F(2, 145) = .59, p = .557, \eta_p^2 = .008$ nor was the main effect of the evaluation dimension, $F(1, 145) = 1.02, p = .314, \eta_p^2 = .007$. As predicted, we observed a significant interaction, $F(2, 145) = 30.10, p < .001, \eta_p^2 = .293$. Participants in the control deprivation condition felt less in control than participants in the morality ($p < .001$) and baseline ($p = .004$) conditions, and the two latter conditions also differed significantly from each other ($p = .004$). Furthermore, participants who were asked to remember their immoral acts declared experiencing emotions related to morality significantly less intensely than did participants in the control condition ($p < .001$) and marginally less

intensely than those the baseline condition ($p = .057$), while the two latter conditions did not differ significantly from each other ($p = .313$)¹.

The results were similar for evaluations by an imaginary stranger although this time the main effect for experimental group was significant, $F(2, 150) = 25.18, p < .001, \eta_p^2 = .251$, and so was the main effect of the dimension, $F(2, 281) = 36.84, p < .001, \eta_p^2 = .197$. Again, these main effects were qualified by a significant interaction, $F(4, 281) = 13.72, p < .001, \eta_p^2 = .155$. Under control threat participants were significantly more prone to expect a stranger to see them as low in control and significantly less prone to expect a stranger to attribute high control to them as compared to participants in the morality deprivation condition ($p < .001$ and $p = .013$). The differences between the control deprivation and baseline condition were not significant for the low ($p = .256$) or high control dimension ($p = 1.00$). The morality deprivation and baseline groups differed significantly or marginally from each other for both subscales ($p = .032$ and $p = .069$ respectively).

In terms of stranger's judgment on the morality dimension, all groups were significantly different from one another. Specifically, the morality deprivation group believed that a stranger would judge them to be significantly less moral than both the control deprivation ($p < .001$) and the baseline group ($p < .001$). The control threat group expected a stranger to rate them as significantly more moral than the baseline group ($p = .005$)².

¹ We repeated the analysis for the three original subscales of morality and warmth, high control, and high control and competence. The results for the morality and warmth subscale descriptively replicated results observed in the analysis using its shortened version (the morality subscale) however this time all conditions differed significantly from each other. The results of the high control and high competence dimension replicated descriptively those for the pure high control subscale but only the difference between baseline and control condition was statistically significant.

² We repeated the analysis replacing the morality subscale with the original morality and warmth subscale. The results did not differ from those reported here.

Table A

The Effects of the Individual Morality and Control Threat on Morality and Control Self-evaluation Subscale of Semantic Differential Measure and on Morality and Control Subscales of the Evaluation by a Stranger Scale

	Baseline (<i>n</i> = 51)		Morality threat (<i>n</i> = 50)		Control threat (<i>n</i> = 47)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Morality (own feelings)	3.08	1.32	2.40 _a	1.20	3.55 _b	1.76
Control (own feelings)	2.87 _a	1.02	3.65 _b	1.36	2.08 _c	1.17
	Baseline (<i>n</i> = 52)		Morality threat (<i>n</i> = 50)		Control threat (<i>n</i> = 51)	
Morality (evaluation by a stranger)	4.26 _a	1.04	3.05 _b	1.38	4.98 _c	.96
Low Control (evaluation by a stranger)	4.40 _a	1.40	3.75 _b	1.25	4.84 _a	1.20
High Control (evaluation by a stranger)	3.01	1.18	3.53 _a	1.14	2.88 _b	1.11

Note. Means with differing subscripts (in each row) are significantly different at $p < .05$ (or lower) based on Bonferroni's post hoc paired comparisons.

Results in Poland

Participants. One hundred and forty-one individuals participated in the pretest in Poland. The data was collected at the main library of the University of Warsaw. Thus, the sample consisted mainly of students. They were between 18 and 29 years of age ($M = 21.65$; $SD = 2.19$). The majority of the sample (71.6%) was female, 27.7% was male, while 1 person did not indicate their gender (0.7%).

Results. The participants' responses to the semantic differential questions were factor-analyzed using a Principal Component Analysis with an Oblimin rotation. All items pertaining to control loaded on a single factor and all items pertaining to morality loaded on a single factor. The competence items were split between the *morality* ("smart-stupid") and

control (“competent-incompetent”) factors and since this solution did not make theoretical sense, they were excluded from further analyses. Items pertaining to warmth loaded on a single factor, but as warmth was not the main focus of the study we do not describe these results below. We next computed composite scores on the two subscales of interest: the morality subscale (3 items; $\alpha = .63$) and the control subscale (4 items; $\alpha = .79$). Higher scores on these subscales indicate higher evaluations of one’s morality or control.

The exact same procedure was used to determine the factorial structure of the scale measuring evaluation by a stranger: we entered all 19 items (i.e., all items except for the filler items) into a Principal Component Analysis with an Oblimin rotation. The solution returned 3 factors that explained 27.66%, 25.62%, and 8.84% of variance. All items loading onto the same factor were averaged to create composite scores pertaining to competence and power (*competent, efficient, smart, agentic, mighty*; $\alpha = .86$), morality (*dishonest, loyal, untrustworthy, immoral, sincere, unfair, good-natured*; $\alpha = .85$; note that all these items were recoded so that higher scores reflect higher morality); and control (*uncertain, helpless, powerless, with no control, weak*; $\alpha = .89$; please note that all items were recoded so that higher scores reflect higher judgments of control); 2 items, *friendly* and *selfless*, loaded equally strongly on the competence and morality factors and were therefore excluded from subsequent analyses. As with the semantic differential scale, we were primarily interested in the effects of the experimental manipulation on the morality and control subscales and therefore only these results are reported below.

With the aim to verify whether the experimental manipulation indeed affected participants’ own feelings and their perceptions of how they might be judged by a stranger in terms of their morality and control, two mixed model ANOVAs were conducted with judgments of control and morality (either by the self or by a stranger) entered as a repeated

measure dependent variable with two levels and experimental group entered as a between subjects factor (see Table B in Appendix 1).

With regard to the semantic differential self-evaluation scale, the main effect of experimental group was significant, $F(2, 129) = 10.59, p < .001, \eta_p^2 = .14$ and so was the main effect of the evaluation subscale, $F(1, 129) = 118.34, p < .001, \eta_p^2 = .48$. As predicted, these main effects were qualified by a significant interaction, $F(2, 129) = 16.11, p < .001, \eta_p^2 = .20$. Participants in the control deprivation condition declared that they experienced emotions associated with control less than participants in the morality threat ($p < .001$) and baseline ($p < .001$) conditions, who only marginally differed from each other ($p = .053$). Contrary to our predictions, participants who were asked to remember their immoral acts did not differ from participants in the other two conditions in terms of experiencing morality-related feelings (all p s were highly insignificant and thus the Bonferroni adjustments yielded only values of 1.000).³

The results were similar for evaluations by an imaginary stranger, although this time the main effect for experimental group was not significant $F(2, 129) = 2.35, p = .100, \eta_p^2 = .04$. The participants believed that a stranger would rate them as significantly more moral than in control, $F(1, 129) = 19.59, p < .001, \eta_p^2 = .13$. Again, these main effects were qualified by a significant interaction, $F(2, 129) = 39.44, p < .001, \eta_p^2 = .38$. Participants in the control deprivation condition reported that a stranger would judge them to have significantly less control than participants in the morality deprivation condition ($p < .001$), but the comparison

³ We did create composite scores of all items loading on the morality and control factors (including the 2 competence items) and repeated the mixed-design ANOVAs reported here for these two composite measures. The results for the control dimension (inclusive of the “competent-incompetent” competence item did not change) while the results for the morality dimension (inclusive of the “smart-stupid” competence item) actually became significant in the expected direction – i.e., participants in the morality threat condition rated their own morality significantly lower than those in the control threat and baseline conditions. However, as these latter effects were driven primarily by the “smart-stupid” item, which arguably is not as strongly related to the concept of morality as judgments of what is right and wrong, we decided not to report this result and focus on the pure morality dimension instead.

with the baseline group was not significant ($p = .099$). The morality deprivation and baseline groups did not differ from each other ($p = .246$). In terms of evaluations of morality, all groups were significantly different from one another. Specifically, the morality deprivation group believed that a stranger would judge them to be significantly less moral than both the control deprivation ($p < .001$) and the baseline group ($p < .001$). The control threat group expected a stranger to rate them as significantly more moral than the baseline group ($p = .008$).⁴

Table B

The Effects of the Individual Morality and Control Threat on Morality and Control Self-evaluation Subscale of Semantic Differential Measure and on Morality and Control Subscales of the Evaluation by a Stranger Scale

	Baseline ($n = 43$)		Morality threat ($n = 43$)		Control threat ($n = 46$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Morality (own feelings)	3.91	0.87	3.81	0.85	3.84	0.96
Control (own feelings)	2.74 _a	1.41	3.32 _a	1.06	1.79 _b	0.88
Morality (evaluation by a stranger)	4.57 _a	0.97	4.31 _b	1.14	5.26 _c	1.01
Control (evaluation by a stranger)	3.78	1.54	3.32 _a	1.26	3.14 _b	1.52

Note. Means with differing subscripts (in each row) are significantly different at $p < .05$ (or lower) based on Bonferroni's post hoc paired comparisons.

⁴ We also excluded the 3 competence items from the first factor of the evaluation by a stranger scale factor solution in order to create a composite score of high power (which is similar to high control) and entered this composite measure as a third within-subjects dependent variable in a mixed-design ANOVA but the effect of the experimental manipulation on this measure was not significant ($p > .500$).

Conclusions

Overall the pretest study results provided substantial support for the validity of the experimental manipulation and showed that indeed recalling past states of morality and control threat resulted in experiencing such threat. While the morality threat did not result in the expected shift in self-evaluations on the morality dimension in Poland, we decided to keep this manipulation as it was successful in Germany and also had the expected effect on the evaluation by a stranger measure.

In most comparisons, the baseline condition means were in between the means of the two other conditions (morality and control threat). In order to make the baseline condition even more emotionally neutral in the main study, we decided to change its content for a description of an everyday habit.