Online Appendix

The Psychology of Online Political Hostility: A Comprehensive, Cross-National Test of the Mismatch Hypothesis

by Alexander Bor and Michael Bang Petersen

Table of Contents

Appendix A. Study 1-4 Materials

First, we provide full question wording on the key variables of our original surveys.

Online hostility battery

[In S2 & S3] Now please think about the past 30 days, specifically. How often did the following happen to you in political discussions that occur on the Internet, including social media and comments sections?

1. I post or share political content or comments, which get flagged or deleted for violating the site's guidelines.
2. I get banned or blocked from a website for posting or sharing political content or comments, which violate the site's guidelines.
3. I post or share political content or comments, which I later regretted or felt ashamed of.
4. I post or share political content or comments, which could be taken as offensive or aggressive.
5. I post or share political content or comments, which could be taken as a threat or harassment.
6. *[In S2 & S3] I had a difficult time tempering my emotions.*

[S4] Please think about the past 30 days, specifically. How often did the following happen to you in **text-based** political discussions that occur **on the Internet, such as on social media or in comments sections**?

1. I made fun of my political opponents.
2. I cursed a lot in a political discussion.
3. I made a comment in a political discussion that others may find hurtful.
4. In a conversation with like-minded people, we made insensitive remarks about our opponents
5. I participated in a political discussion that was aggressive, uncivil or hostile.
6. In the heat of a political discussion, I got out of control.
7. I humiliated or bullied others for their political views or actions.
8. I threatened or harassed my political opponents.

A note on the decisions behind our hostility items. We specify the 30 days window to minimize recall bias. We include flagging and banning as indicators of behaviour that the provider considers anti-social or harmful. (As noted in main text), we included an item on "regret" influenced by the accidental nature of mismatch induced hostility. We included the phrase "which could be taken as" to reduce social desirability bias, offering an easy rationalization for hostile behaviour ("I didn't mean it, but maybe they took it as…").

Offline hostility battery

[In S2 & S3] Now please think about the past 30 days, specifically. How often do the following happen to you in political discussions that occur face-to-face?

1. I make political comments, which I later regret or feel ashamed of.
2. I make political comments, which could be taken as offensive or aggressive.
3. I make political comments, which could be taken as a threat or harassment.
4. *[In S2 & S3] I had a difficult time tempering my emotions.*

[S4] Please think about the past 30 days, specifically. How often did the following happen to you in political discussions that occur in a conversation, **where you could hear and see the other person?**

[Items identical to S4 online above]

Talking about politics

[*Offline*] How often do you talk about politics or public affairs face-to-face, with the following?
[*Online*] How often do you talk about politics or public affairs on the Internet, with the following?

1. Family and friends
2. Co-workers and acquaintances
3. Strangers
4. People who agree with me

5. People who disagree with me

Perception batteries

[*Offline*] How often do you experience the following in face-to-face political discussions?
[*Online*] How often do you experience the following in political discussions that occur on the Internet?

1. My discussion partner does not allow me to drop the discussion, even though I would like to
2. I ruminate over the content of a discussion after it is over
3. I spend a lot of energy during a discussion to ruminate over the arguments that are being made or that I should make
4. People involve me in a discussion that I do not feel like having
5. I continue a discussion even though I do not enjoy it
6. A discussion is quickly over
7. When we cannot agree, we just switch to talking about something else
8. When my discussion partner says something that might be offensive, he or she apologizes

Witnessing hostility
[*Offline*] How often do the following happen to you in political discussions that occur face-to-face?
[*Online*] How often do the following happen to you in political discussions that occur on the Internet?

1. I am offended or insulted by someone.
2. I witness that someone I know is offended or insulted by someone.
3. I witness that someone I do not know is offended or insulted by someone.

Difficulties in Emotion Regulation Scale

How much do you disagree or agree with the following statements?

1. When I'm upset, I feel ashamed at myself for feeling that way.
2. When I'm upset, I become angry with myself for feeling that way.
3. When I'm upset, I have difficulty focusing on other things.
4. When I'm upset, I have difficulty getting work done.
5. When I'm upset I lose control over my behaviors.
6. I experience my emotions as overwhelming and out of control.
7. I pay attention to how I feel.
8. When I am upset, I take time to figure out what I'm really feeling.
9. When I'm upset, I believe there is nothing I can do to make myself feel better.
10. When I'm upset, I believe that wallowing in it is all I can do.
11. I have difficulty making sense out of my feelings.
12. I have no idea how I am feeling.

Status Driven Risk Taking Scale

How much do you disagree or agree with the following statements?

1. I would rather live as an average person in a safe place than live as a rich and powerful person in a dangerous place.
2. I would enjoy being a famous and powerful person, even if it meant a high risk of assassination.
3. If the pay was really high, I would be willing to work with extremely explosive materials.
4. I would risk my life for a good chance of finding a huge amount of buried treasure.
5. If I could become rich and famous by winning a major competition, I would put my life on the line to win it.
6. I would like to live in a country where people who take huge risks have the chance to gain superior social status.
7. No matter how good the pay of perks; I would not want to be a spy who takes very dangerous assignments.
8. I would rather live a secure life as an ordinary person than risk everything to be at the top.
9. For a very high-status job, I would be willing to live in a place that had an extremely high crime rate.

10. I would take a very high-status job even if I had to live in a place where there are many deadly diseases.
11. I would volunteer for a risky medical experiment if it paid me enough to retire early.
12. Being an organized crime boss would be far too dangerous for me.
13. To become a billionaire, I would be willing to trade 10 years from my life expectancy.
14. I would not go to a warzone even if the business opportunities were extremely profitable.

Trait aggression
How much do you disagree or agree with the following statements?

1. I have trouble controlling my temper.
2. I am an even-tempered person.
3. Sometimes I fly off the handle for no good reason.
4. Given enough provocation, I may hit another person.
5. There are people who pushed me so far that we came to blows.
6. If I have to resort to violence to protect my rights, I will.
7. I tell my friends openly when I disagree with them.
8. When people annoy me, I may tell them what I think of them.
9. My friends say that I'm somewhat argumentative.
10. Other people always seem to get the breaks.
11. I sometimes feel that people are laughing at me behind my back.
12. When people are especially nice, I wonder what they want.

Notes:

The Hostility, Talking about politics, Perception and Witnessing hostility batteries were answered on the following 7-point scale: 1. Never 2. Only occasionally 3. A few times a month 4. A few times a week 5. Most days 6. Every day 7.Several times a day.

The Status-driven risk taking, Difficulties in emotion regulation and Trait aggression batteries were asked on a standard 7-point agree-disagree scale.

Appendix B. Some descriptive insights for Studies 1-3

B1. Sample demographics

Table B1. Sample demographics

Studies	S1	S2 & S8	S3 & S8	S4*	S5*	S6	S7
Country	USA	Denmark	USA	USA	USA	USA	USA
Provider	YouGov	YouGov	Lucid	YouGov	Yougov	MTurk	MTurk
Age	44.2 (12.91)	41.17 (13.49)	45.08 (16.56)	42.63 (17.02)	46.00 (17.64)	37.02 (11.39)	38.95 (12.08)
Female	0.51 (0.50)	0.48 (0.50)	0.51 (0.50)	0.49 (0.50)	0.52 (0.50)	0.60 (0.49)	0.58 (0.49)
White	0.70 (0.46)		0.72 (0.45)	0.65 (0.48)	0.66 (0.47)		
Higher educated	0.41 (0.49)	0.39 (0.49)	0.44 (0.50)	0.44 (0.50)	0.41 (0.49)	0.65 (0.48)	0.67 (0.47)
Party ID: Democrat	0.46 (0.50)		0.48 (0.50)	0.55 (0.50)	0.51 (0.50)	0.45 (0.50)	0.48 (0.50)
Party ID: Republican	0.33 (0.47)		0.37 (0.48)	0.27 (0.45)	0.29 (0.46)	0.24 (0.43)	0.24 (0.43)
Party ID: Blue block		0.38 (0.49)					

Party ID: Red block		0.51 (0.50)					
N	1515	1041	998	770	1317	1923	1640

* Note that Study 4 is a subset of Study 5.

B2. Reliability of indices

Table B2 below reports alpha reliability statistics for all indices in our study. In general, the reliability of most indices is high or satisfactory. The sole exception is the perceptions of conflicts indices in Study 3. A post-hoc factor analysis indicates that a two factor solution fits the data well, wherein the items tapping into perceived severity of conflicts form the first dimension, and reverse coded items tapping into resolution of conflicts form the second dimension. Following our pre-registration plan and ensuring consistency across S2 and S3, we create indices by averaging across all items.

Table B2. Raw alpha reliability statistics for the indices

	Study1 USA	Study2 Denmark	Study3 Lucid	Study4 USA	Study6 MTurk	Study7 MTurk
Aggression				0.843		0.762
Percep. of conflicts - offline		0.618	0.255			
Percep. of conflicts - online		0.671	0.242			
Difficulties in emotion regulation	0.859	0.921			0.955	0.951
Hostility - offline	0.924	0.938	0.904	0.937		
Hostility - online	0.948	0.957	0.947	0.935	0.931	
Status-driven risk taking	0.878	0.916		0.881	0.910	0.916
Talking about politics - offline	0.880	0.871		0.851		
Talking about politics - online	0.929	0.915		0.869		
Negative tone - offline	0.909	0.867				
Negative tone - online	0.943	0.831				

B3. Descriptive statistics

The tables below report means, standard deviations and simple bivariate correlations for the main variables in our analysis in each of our samples.

Table B3a. Descriptives S1 USA

	num	mean	sd	1	2	3	4	5	6
Hostile Offline	1	0.07	0.17	-					
Hostile Online	2	0.07	0.17	0.89	-				
Talk Offline	3	0.25	0.20	0.48	0.47	-			
Talk Online	4	0.18	0.21	0.54	0.57	0.69	-		
SDRT	5	0.27	0.20	0.39	0.41	0.17	0.26	-	
Trait Aggression	6	0.38	0.19	0.42	0.41	0.18	0.24	0.41	-

Table B3b. Descriptives -- S2 Denmark

	num	mean	sd	1	2	3	4	5	6	7	8
Hostile Offline	1	0.07	0.16	-							
Hostile Online	2	0.06	0.16	0.87	-						
Talk Offline	3	0.26	0.16	0.58	0.53	-					
Talk Online	4	0.19	0.19	0.57	0.6	0.63	-				
Perception Offline	5	0.35	0.19	0.39	0.36	0.41	0.36	-			
Perception Online	6	0.31	0.20	0.38	0.37	0.35	0.52	0.55	-		
SDRT	7	0.23	0.19	0.56	0.53	0.34	0.39	0.3	0.35	-	
Trait Aggression	8	0.33	0.17	0.51	0.46	0.33	0.37	0.31	0.32	0.51	-

Table B3c. Descriptives -- S3 USA

	num	mean	sd	1	2	3	4
Hostile Offline	1	0.16	0.21	-			
Hostile Online	2	0.13	0.21	0.8	-		
Perception Offline	3	0.42	0.11	0.48	0.46	-	
Perception Online	4	0.42	0.11	0.46	0.5	0.68	-

Table B3d. Descriptives -- S4 USA

	num	mean	sd	1	2	3	4	5	6
Hostile Offline	1	0.18	0.21	-					
Hostile Online	2	0.18	0.21	0.84	-				
Talk Offline	3	0.29	0.20	0.7	0.61	-			
Talk Online	4	0.28	0.22	0.6	0.69	0.68	-		
SDRT	5	0.31	0.20	0.52	0.52	0.39	0.36	-	
Trait Aggression	6	0.42	0.20	0.56	0.51	0.38	0.36	0.51	-

B4. Talking about politics online and offline

To acknowledge the potentially large differences between online and offline political discussion environments, in Figure B1 we report the observed differences with regards to discussion partners. The plot depicts the mean frequency of talking with various groups about politics online (dark bars) and offline (light bars). Results show that people talk substantively more about politics in face-to-face interactions than on the internet. The only exception from this rule is discussions with strangers which occur more online, especially in relative terms. Besides, people in general seem to prefer talking with people they agree with, but this in-group bias is somewhat larger in offline discussions compared to online discussions, at least in the United States. Overall, we argue that while this plot likely shows only the tip of the iceberg when it comes to differences between online and offline discussions, it is remarkable that we find very high parallels in self-reported hostile behavior.

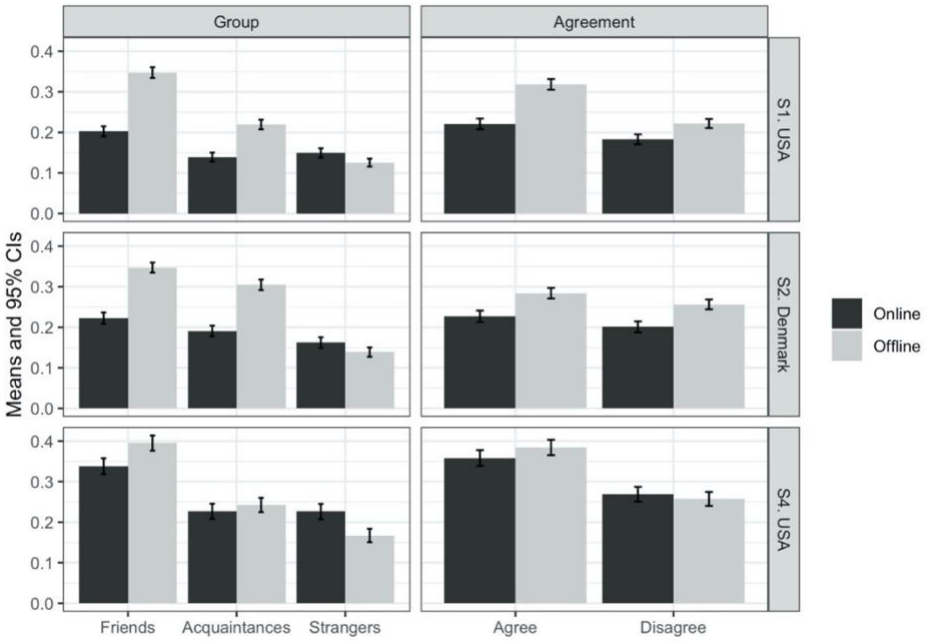


Figure B1. *Talking about politics online and offline Studies 1, 2 and 4*

Appendix C. Full model details for Studies 1-4

C1. Change hypothesis

Table C1. Regressing online hostility on offline hostility in our three studies

	Dependent variable:			
	Online hostility			
	USA	Denmark	USA	USA
	(1)	(2)	(3)	(4)
Offline hostility	0.885*** (0.012)	0.875*** (0.016)	0.803*** (0.019)	0.838*** (0.020)
Intercept	0.006*** (0.002)	0.003 (0.003)	0.008 (0.005)	0.028*** (0.005)
Observations	1,515	954	998	770
R^2	0.789	0.755	0.641	0.704
Note:				*p<0.1; **p<0.05; ***p<0.01

Table C2. Regressing self-reported hostility on SDRT and demographic covariates

	Dependent variable:					
	Online	Offline	Online	Offline	Online	Offline
	S1. United States		S2. Denmark		S4. United States	
	(1)	(2)	(3)	(4)	(5)	(6)
SDRT	0.29 (0.02)***	0.29 (0.02)***	0.42 (0.03)***	0.45 (0.03)***	0.51 (0.04)***	0.52 (0.04)***
Female	-0.02 (0.01)**	-0.03 (0.01)***	0.004 (0.01)	-0.002 (0.01)	-0.02 (0.01)	-0.01 (0.01)
Age	-0.07 (0.01)***	-0.07 (0.01)***	-0.04 (0.02)**	-0.03 (0.02)*	-0.001 (0.0005)	-0.0004 (0.0005)
White	-0.03 (0.01)***	-0.02 (0.01)**			-0.003 (0.02)	-0.02 (0.02)
Higher educated	0.005 (0.01)	0.01 (0.01)	-0.01 (0.01)	-0.02 (0.01)*	-0.03 (0.02)**	-0.03 (0.02)*
Income	-0.02 (0.02)	0.0003 (0.02)	-0.02 (0.02)	-0.02 (0.02)	0.06 (0.03)*	0.05 (0.03)*
US - Party ID	0.003 (0.01)	0.01 (0.01)			-0.03 (0.02)	-0.03 (0.02)*

DK - Party ID:Other			0.04 (0.02)***	0.02 (0.02)		
DK - Party ID:Red block			0.01 (0.01)	0.003 (0.01)		
Constant	0.07 (0.02)***	0.06 (0.02)***	-0.02 (0.02)	-0.01 (0.02)	0.07 (0.03)**	0.07 (0.03)**
Observations	1,315	1,315	840	833	683	683
Adjusted R^2	0.19	0.18	0.31	0.31	0.28	0.28
Note:						*p<0.1; **p<0.05; ***p<0.01

C2. Selection hypothesis

Table C3. Talking about politics regressed on status drive and covariates

	Dependent variable:					
	Online	Offline	Online	Offline	Online	Offline
	S1.United States		S2.Denmark		S4.United States	
	(1)	(2)	(3)	(4)	(5)	(6)
SDRT	0.23***	0.13***	0.36***	0.25***	0.38***	0.35***
	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)
Female	-0.04***	-0.05***	0.03**	0.03**	-0.03**	-0.04**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Age	-0.07***	-0.05***	0.04**	0.03*	-0.0003	-0.0004
	(0.02)	(0.02)	(0.02)	(0.02)	(0.001)	(0.0005)
White	0.01	0.01			-0.001	-0.01
	(0.01)	(0.01)			(0.02)	(0.02)
Higher educated	0.02	0.02**	-0.03**	-0.01	-0.01	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Income	0.06**	0.15***	-0.02	0.001	0.13***	0.16***
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
US - Party ID	-0.02	-0.02			-0.01	0.002
	(0.02)	(0.01)			(0.02)	(0.02)
DK - Party ID:Other			0.08***	0.05***		
			(0.02)	(0.02)		
DK - Party ID:Red block			0.02	0.02		
			(0.01)	(0.01)		
Constant	0.16***	0.21***	0.08***	0.17***	0.16***	0.18***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Observations	1,315	1,315	865	865	683	683
Adjusted R^2	0.08	0.09	0.14	0.10	0.17	0.19
Note:						*p<0.1; **p<0.05; ***p<0.01

C3. Asymmetries in witnessing hostility

Table C4. Likelihood of witnessing hostility online and offline against various parties

	Dependent variable:	
	Witnessing hostility	
	USA	Denmark
	(1)	(2)
Online environment	0.06 (0.01)***	0.21 (0.01)***
Target: Friends	-0.01 (0.01)**	-0.03 (0.004)***
Target: Self	-0.08 (0.01)***	-0.06 (0.01)***
Online × Friends	-0.03 (0.01)***	-0.14 (0.01)***
Online × Self	-0.04 (0.01)***	-0.17 (0.01)***
Constant	0.28 (0.01)***	0.17 (0.01)***
Observations	5,988	6,264
Adjusted R^2	0.03	0.16
Note:		*p<0.1; **p<0.05; ***p<0.01

Appendix D. Robustness checks and additional analyses for Studies 1-4

D1. Replicating the hostility gap with positive tone ratings

Table D1. T-tests between positive online and offline tone ratings

name	Offline – M(SD)	Online – M(SD)	Difference	T	DF	p-value
Tone - US	0.49 (0.27)	0.35 (0.28)	0.14	22.25	1,514	0
Tone - Denmark	0.60 (0.19)	0.36 (0.20)	0.24	29.10	981	0

D2. Levels of offline and online hostility with paired t-tests

The mismatch theory predicts that online environments may trigger hostility among otherwise peaceful individuals. We believe that the best instrument to test this hypothesis is a regression, which provides an estimate for both the strength of the linear association between the two variables, and the mean level of online hostility among individuals never hostile offline. However, these regressions have difficulties quantifying the overall (im)balance between the two environments and rely on naïve null hypotheses. As an alternative method for comparing levels of online and offline hostility, here we report estimates based on paired t-tests relying both on traditional null hypotheses (zero mean difference), and an alternative null hypothesis specifying the smallest effect size of interest in favor of the change hypothesis based on a sensitivity power analysis. Table D2 displays the results in all three samples. It demonstrates that in all cases, there is slightly more hostility offline than online. The highly significant equivalence tests show that it would be very unlikely to observe this data, if the change hypothesis was true (ts < -4.2, ps < 0.001).

Table D2. Paired t-tests on self-reported hostility offline and online

	M.On	SD.On	M.Off	SD.Off	diff	Ttest. t	Ttest. p	TOST. bound	TOST. t	TOST. p	N
S1. USA	0.07	0.17	0.07	0.17	-0.003	-1.31	0.19	0.01	-4.23	0.0000	1,515
S2. DK	0.07	0.16	0.07	0.16	-0.01	-2.40	0.02	0.01	-5.32	0.0000	954
S3. USA	0.13	0.21	0.16	0.21	-0.02	-5.45	0.0000	0.01	-8.37	0	998
S4. USA	0.18	0.21	0.18	0.21	-0.001	-0.24	0.81	0.01	-3.16	0.001	770

Note: On = Online, Off = Offline, Ttest = standard t-test of zero mean difference, TOST = Equivalence test of difference below 0.01).

D3. Question-order experiments (Studies 3-4)

A potential limitation of our approach estimating the relationship between online and offline hostility is that it may suffer from common source bias: people may report similar levels of online and offline hostility to remain consistent throughout the survey. To rule out this alternative explanation, in Study 3 and 4 we randomized whether participants first saw the block of question pertaining to the online or the offline environment. Moreover, in Study 3 we embedded ~5 min distractor task between the two blocks to reset short term memory. As the most stringent test of our hypothesis, we compare the self-reported hostility levels from the first block of the survey, relying on an independent sample t-test.

In Study 3, we find that people self-report slightly but statistically significantly *less* hostility online than offline (mean difference of m = 0.048, t(995) = 3.636, p < 0.001). Given that we find a significant effect opposite to the mismatch hypothesis' prediction, our pre-registered an equivalence test with the boundaries of Cohen's d = +/-.21 finds that the observed effect is not equivalent to zero, t(995) = -0.316, p = 0.624. In Study 4 in turn, we find that people report very similar levels of hostility online and offline although both a standard null-hypothesis test (mean difference of m = 0.019, t(754) = 1.285, p = .20), and an equivalence test (t(754) = 1.63, p = 0.05) yield non-significant results. Overall, these results reassure us that our findigns are not an artefact of common source bias.

D4. The difference in the levels of status-driven risk taking between the average person in an online and an offline political discussion

As an attempt to put the magnitude of the (reverse) selection effect into perspective, we calculate the chance of encountering someone with above median levels of status-driven risk taking about politics online and offline. We use the talking about politics indices as weights to estimate the characteristics of the latent subpopulations of "online talkers" and "offline talkers". Respondents reporting higher values on either of the two index get a larger weight because they participate in a larger share of discussions. Meanwhile, those who admittedly never participate in these discussions, should not be included in the calculation. As expected – given the general interest of high SDRT individuals in politics –the average discussion partner is more likely to have high SDRT both offline and online. (US Offline: M = 0.52, US Online: M = 0.59; DK Offline: M = 0.56, DK Online: M = 0.62). The difference between talking to a high SDRT person offline and online is thus 7 and 6 percentage points in the US and Denmark, respectively (*p*s < 0.01). Note, that as we cannot replicate this selection effect in Study 4, we find no meaningful asymmetry in the level of status drive of offline (0.61) and online (0.62) discussion partners.

D5. Selection effects disaggregated by groups of discussion partners

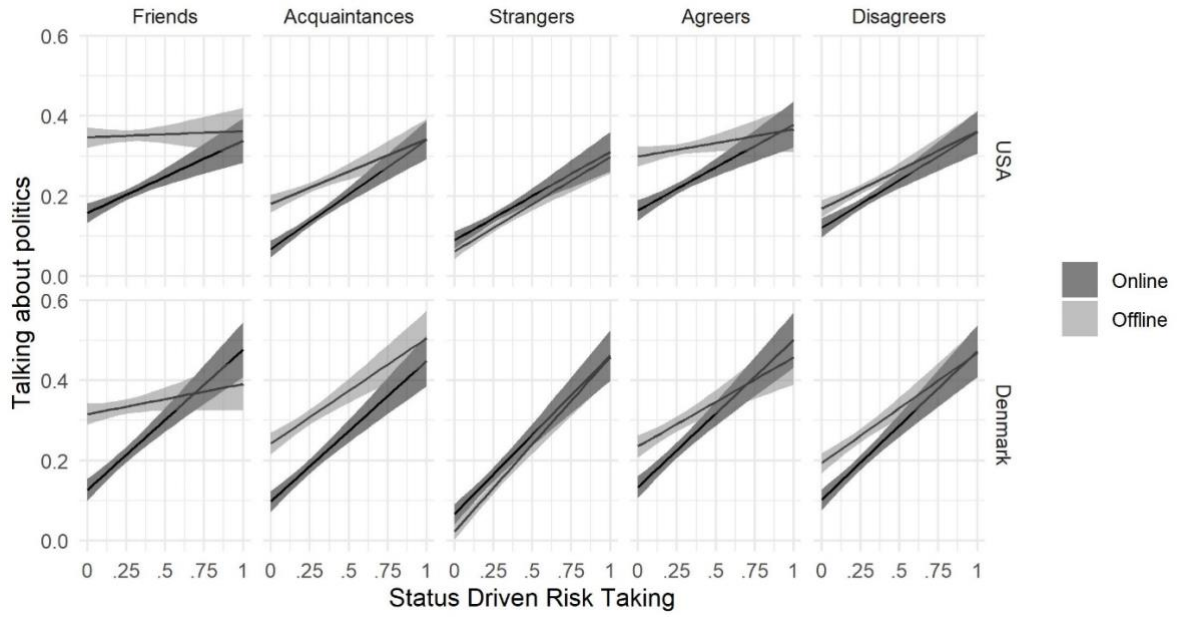


Figure D1 *Talking about politics with various groups and SDRT*

Figure 4 in the main text reports a reverse selection effect, wherein peaceful individuals select out from online political discussions. Here, we ask if any of the five groups constituting the talking about politics indices are driving this effect. Re-running our models with each of the five talking variables as DV, we find that the largest change between offline

and online talking behavior is present for the groups closest to the respondent: friends and family and those with whom the respondent agrees. As Figure D1 demonstrates, high and low SDRT respondents alike often talk about politics offline with friends. However, people low on SDRT talk as little about politics with friends online as they talk with any other group. Note, that we failed to replicate the selection finding in Study 4, thus we do not report disaggregated patterns here. They could be found at the paper's OSF repository.

D6. Trait aggression – An alternative measure of strategic hostility predispositions

Trait aggression taps into individual differences in a predisposition to engage in physical or verbal aggression, feel angry and have hostile feelings towards others. As such, it could be considered as a robustness check for our chosen measure of strategic hostility motivations – status-driven risk seeking. On the one hand, it directly taps in to aggression and it is thus conceptually closer to our key dependent variable – political hostility. On the other hand, it is not limited to political discussions or issues. Thereby, it offers another test of change hypothesis. If online political hostility was "accidental" or an unintended consequence of unique environmental factors, we would expect to find a weak relationship between trait aggression and online political hostility (or at least weaker than for offline political hostility). Similarly, under the Selection hypothesis, we would expect participants high on trait aggression to favor online (vs offline) political discussions.

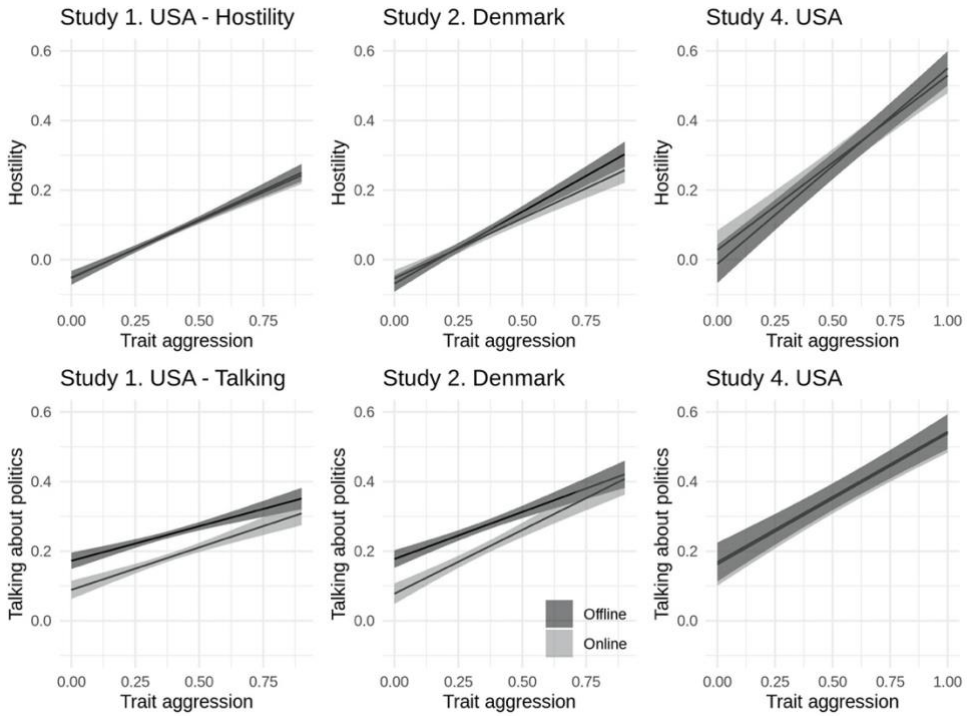


Figure D2 *Reproducing change and selection results with trait aggression*

Our findings closely replicate the results reported in the main text and thus offer no support for the mismatch hypothesis. There is a strong association between trait aggression and hostility both offline and online (see the first row in Figure D2). Meanwhile, if anything trait aggression offer even less support for the Selection hypothesis. In the US, the relationship between talking about politics and trait aggression is very similar offline and online (the lines are parallel). In Denmark, we see the previously established tendency for people low on strategic hostility motivations (here, trait aggression) to select out of online political discussions, but this difference is smaller than the difference found for status-driven risk taking.

D7. Difficulties in Emotion Regulation – A measure of "involuntary" hostility predispositions

When it comes to individual differences predictive of hostility both offline and online, our manuscript focuses on strategic considerations. But, admittedly, the mismatch theory suggests hostility could be accidental, rooted in emotion regulation difficulties caused by the quirks of online environments. Psychological research shows that difficulties in

emotion regulation vary across individuals in a relatively stable manner [1] and predict the propensity to engage in hostile behavior [2]. If online political discussions make it particularly difficult to regulate emotions, then individuals with general emotion regulation problems are especially prone to succumb to aggression in online relative to offline contexts. Moreover, we would expect that difficulties in emotion regulation is a better predictor of online hostility than status concerns.

Similarly, to the results reported for SDRT, we find essentially identical relationships between the two hostility variables and the DERS scale (S1 US Offline: $\beta = 0.326$, US Online: $\beta = 0.315$; Denmark Offline: $\beta = 0.245$, S2 Denmark Online: $\beta = 0.219$, S4 US Offline: $\beta = 0.538$, US Online: $\beta = 0.492$; all *p*s < 0.001). Difficulties in emotion regulation is an important correlate of hostility both in the United States and Denmark; yet, we find no evidence that it has a disproportionate effect online. Moreover, contrary to the change hypothesis, emotion regulation is not a more important correlate of online political hostility than status-driven risk taking. Indeed, if both traits are simultaneously included in a multiple regression analysis the two associations are about equal in strength in the United States (S1 DERS: $\beta = 0.25$; SDRT: $\beta = 0.22$; S4 DERS: $\beta = 0.34$; SDRT: $\beta = 0.37$), but status drive is a stronger correlate in Denmark (DERS: $\beta = 0.14$, SDRT: $\beta = 0.36$).

D8. Seemingly Unrelated Regression (SUR) models testing the significance of differences across online and offline environments

The mismatch hypothesis is about the adverse effects caused by the differences between the attributes of offline and online discussion environments. Accordingly, many of our analyses are contrasting relationships with the personality measures and offline and online behavior. In the manuscript, we report simple OLS regressions to succinctly summarize our findings. However, these models do not offer a formal statistical test for the contrast between the two environments. To achieve this we re-run our tests relying on seemingly unrelated regression (SUR) models implemented in a structural equation modeling (SEM) framework. Specifically, we run both the online and the offline version of the regression simultaneously and allow the covariance between the two dependent variables to be freely estimated. This enables to calculate contrast effects between the intercepts and slopes of interest in our model. These contrasts then provide an estimate whether the difference between the size of coefficient estimates is statistically significant from zero or not.

Table D3 reports these estimates for the models regressing self-reported hostility on status-driven risk taking along with demographic covariates (age, sex, education, income, partisan identity and race in the US). Remember, our OLS models showed that status drive is positively associated with hostility and that the effects are essentially identical across the two environments. Our SUR models provides the same conclusions. Indeed the coefficient estimates are nearly identical to simple OLS models. Moreover, the substantively small contract effects and confidence intervals containing zero reaffirm that our data shows no asymmetry between the two platforms.

Table D4, in turn, reports the estimates for the models regressing talking about politics on status-driven risk taking and covariates. Recall that we find little evidence that people with high status drives are selecting into online conversations (at the expense of offline conversations). We do find, however, that more peaceful respondents are less likely to participate in online discussions. Congruently with this, the SUR models show negative and statistically significant contrast between the intercepts. This means that people with average levels of status drive (i.e. these are people lower on the scale given the right skew in the distribution) are significantly less likely to report talking about politics online. Meanwhile, the models show a significant positive contrast with regards to the association of status-drive and the DVs. This means that status-drive is a significantly stronger associate of talking about politics online than offline. This helps to overcome the handicap low status-drive people accumulate online.

[1] Kim L Gratz and Lizabeth Roemer, "Multidimensional Assessment of Emotion Regulation and Dysregulation," *Journal of Psychopathology and Behavioral Assessment* 26, no. 1 (2004): 41–54, https://doi.org/10.1023/B:JOBA.0000007455.08539.94.
[2] Thomas F. Denson, C. Nathan DeWall, and Eli J. Finkel, "Self-Control and Aggression," *Current Directions in Psychological Science*, 2012, https://doi.org/10.1177/0963721411429451.

Table D3. Seemingly unrelated regressions for online and offline hostility

Country	Parameter	Environment	B	SE	p	ci.lower	ci.upper
S1.USA	Intercept	offline	0.06	0.02	0.0003	0.03	0.10
S1.USA	Status drive	offline	0.29	0.02	0	0.24	0.33
S1.USA	Intercept	online	0.07	0.02	0.0000	0.04	0.10
S1.USA	Status drive	online	0.29	0.02	0	0.25	0.34
S1.USA	intercept	contrast	0.01	0.01	0.39	-0.01	0.02
S1.USA	Status drive	contrast	0.004	0.01	0.77	-0.02	0.03
S2.Denmark	Intercept	offline	-0.01	0.02	0.66	-0.04	0.03
S2.Denmark	Status drive	offline	0.45	0.03	0	0.39	0.51
S2.Denmark	Intercept	online	-0.02	0.02	0.23	-0.05	0.01
S2.Denmark	Status drive	online	0.42	0.03	0	0.37	0.48
S2.Denmark	intercept	contrast	-0.01	0.01	0.22	-0.03	0.01
S2.Denmark	Status drive	contrast	-0.03	0.02	0.12	-0.06	0.01
S4.USA	Intercept	offline	0.07	0.03	0.03	0.01	0.13
S4.USA	Status drive	offline	0.52	0.04	0	0.45	0.60
S4.USA	Intercept	online	0.07	0.03	0.03	0.01	0.13
S4.USA	Status drive	online	0.51	0.04	0	0.44	0.59
S4.USA	intercept	contrast	-0.003	0.02	0.90	-0.04	0.04
S4.USA	Status drive	contrast	-0.01	0.03	0.67	-0.06	0.04

Table D4. Seemingly unrelated regressions for online and offline talk about politics

Country	Parameter	Environment	B	SE	p	ci.lower	ci.upper
S1.USA	Intercept	offline	0.21	0.02	0	0.17	0.25
S1.USA	Status drive	offline	0.13	0.03	0.0000	0.08	0.19
S1.USA	Intercept	online	0.16	0.02	0	0.11	0.20
S1.USA	Status drive	online	0.23	0.03	0	0.16	0.29
S1.USA	intercept	contrast	-0.05	0.02	0.003	-0.09	-0.02
S1.USA	Status drive	contrast	0.09	0.02	0.0002	0.04	0.14
S2.Denmark	Intercept	offline	0.16	0.02	0	0.13	0.20
S2.Denmark	Status drive	offline	0.27	0.03	0	0.20	0.33
S2.Denmark	Intercept	online	0.08	0.02	0.0003	0.04	0.12
S2.Denmark	Status drive	online	0.38	0.04	0	0.31	0.46
S2.Denmark	intercept	contrast	-0.09	0.02	0.0000	-0.12	-0.05
S2.Denmark	Status drive	contrast	0.12	0.03	0.001	0.05	0.18
S4.USA	Intercept	offline	0.18	0.03	0	0.12	0.25
S4.USA	Status drive	offline	0.35	0.04	0	0.28	0.43
S4.USA	Intercept	online	0.16	0.03	0.0000	0.10	0.23
S4.USA	Status drive	online	0.38	0.04	0	0.30	0.46
S4.USA	intercept	contrast	-0.02	0.03	0.49	-0.08	0.04
S4.USA	Status drive	contrast	0.03	0.03	0.40	-0.04	0.10

D9. Hostility relative to talking by SDRT groups

An interesting dilemma regarding the roles of traits versus states is the frequency of political hostility relative to total time spent discussing politics. As an exploratory analysis, we regress online political hostility on SDRT, talking about politics online and their interactions. In Figure D3 below, the dots denote respondents, the black lines are conditional means from the linear regression, and while the red lines relax the linearity assumption in the interaction. The plot implies that while people with low (one standard deviation below mean) or average levels of SDRT show little tendency to be hostile even if they spend a lot of time talking about politics, many of those respondents high (one standard deviation above mean) on SDRT are hostile almost as often as often they talk about politics (points and loess curve aligning close to the diagonal).

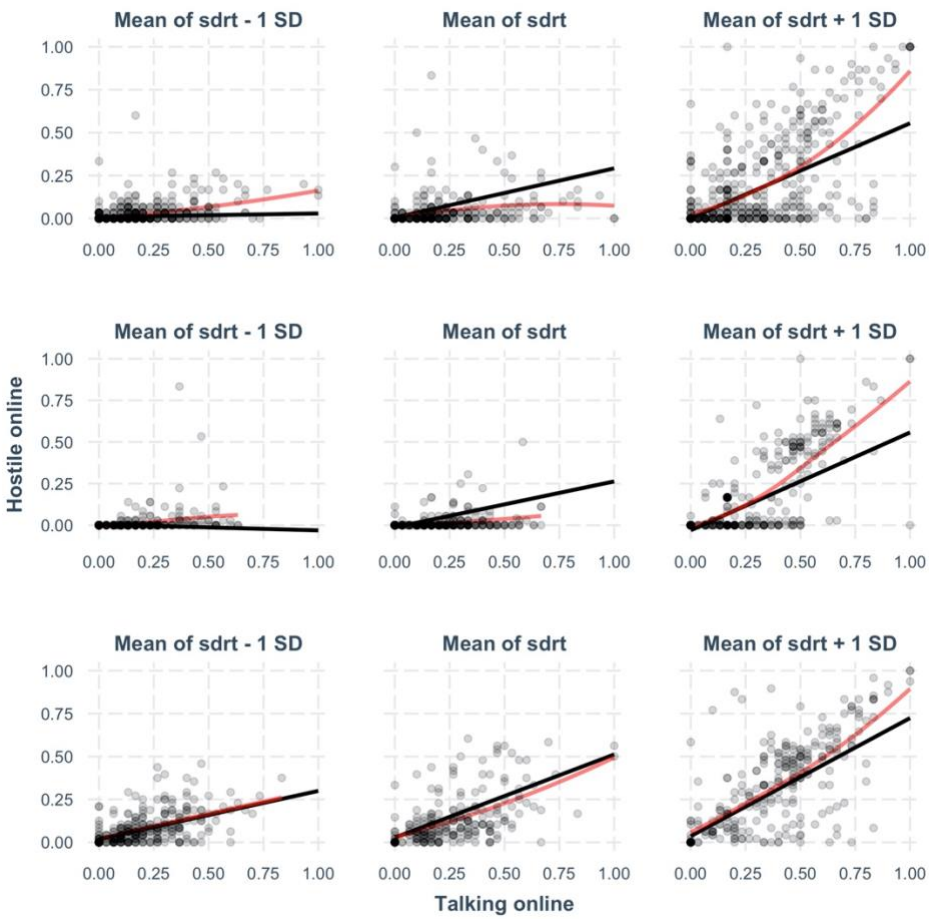


Figure D3 *Online hostility as a function of SDRT * Talking about politics online*

Appendix E. Materials and model details for Study 5 – Vignette experiment

E1. Vignettes

(Within-respondent manipulation in black. Order of messages was randomised).

Imagine that you participate in a political discussion **at a dinner party with 5 other people / public townhall meeting / private chat group with 5 other people / public internet forum**.

The discussion turns to the topic of immigration. Someone makes the following comment:

> *"Folks who think that the current levels of immigration are sustainable are stupid and irresponsible. Crime and unemployment due to immigration hurt hypocrite Democrats as much as everyone else. Wicked and sadistic immigrants harass innocent people. They should respect the culture of people living in this country more."*

The discussion turns to the topic of COVID-19 pandemic. Someone makes the following comment:

"I wish Trump put at least as much effort into fighting COVID-19 as he spent on trolling people on Twitter. Granted, he had a tough job after several years of eroding trust in science and mainstream media. Most Republicans hardly believed the coronavirus exists. Still, too many people have died due to his inaction."

The discussion turns to the topic of the siege of the US Capitol. Someone makes the following comment:

"I can't believe they let a mob of blood-thirsty lunatics siege the U.S. Capitol. These animals should be put in jail along with all the reckless politicians who abetted them. We can't tolerate such attacks at the heart of our nation."

The discussion turns to the topic of abortion rights. Someone makes the following comment:

"I'm so glad that more and more US States are passing anti-abortion laws. The right to life is the most sacred value. Too many innocent fetuses and babies have been murdered because of the Supreme Court's historical blunder legalizing abortion. I hope the new Conservative majority in the Supreme Court overturns Roe v Wade."

E2. Dependent variables (repeated after each vignette)

How offensive does this comment seem to you? (1) Not offensive at all - (7) Extremely offensive

How appropriate or inappropriate would this comment be in the given context? (1) Perfectly appropriate - (7) Extremely inappropriate

Finally, please consider how common or rare it is that such a comment is made in the given context. (1) Very common - (7) Very rare

E3. Model details

Table E1 below reports detailed regression results for the Study 5 relying on fixed-effects regressions using respondents and messages as fixed effects terms (Models 1-3). Note that our focal independent variable is the online-offline manipulation. We also included a public private manipulation to conceal the focus of the study and to create a bit of variation between vignettes. As an exploratory analysis we investigate the effects of this private-public manipulation, but find negligible effects on all three DVs. Interestingly, participants are not greatly influenced by the publicity of a message when judging its offensiveness or inappropriateness. Likewise, participants report that hostile messages are about equally rare in private and public discussions (although it remains a question whether this reflects potentially biased perceptions or personal experience). Do demonstrate that despite these small effects participants paid close attention to the experiment, we report the causal effects of messages in Models 4-6 using abortion as the baseline category and only respondent fixed effects. Apparently, participants saw large differences between the offensiveness and inappropriateness of the messages, with the immigration message being rated as 12-13 points more offensive and inappropriate than the abortion message, whereas the capitol and covid19 messages were rated as 12 points less offensive and 9-10 points less inappropriate. Rarity effects are similar, although much smaller.

Table E1. Fixed effects regression results for Study 5

	Dependent variable:					
	Offensive	Inappropriate	Rare	Offensive	Inappropriate	Rare
	(1)	(2)	(3)	(4)	(5)	(6)
Online	-0.01	-0.03**	-0.02**			
	(0.01)	(0.01)	(0.01)			
Public	0.002	-0.01	-0.02**			
	(0.01)	(0.01)	(0.01)			
Message - Capitol				-0.12***	-0.10***	-0.04***
				(0.01)	(0.01)	(0.01)
Message - COVID19				-0.12***	-0.09***	-0.06***
				(0.01)	(0.01)	(0.01)
Message - Immigration				0.13***	0.12***	0.03***
				(0.01)	(0.01)	(0.01)
Observations	5,268	5,268	5,268	5,268	5,268	5,268
Adjusted R^2	0.11	0.12	0.32	0.11	0.12	0.32
Note:				*p<0.05; **p<0.01; ***p<0.001		

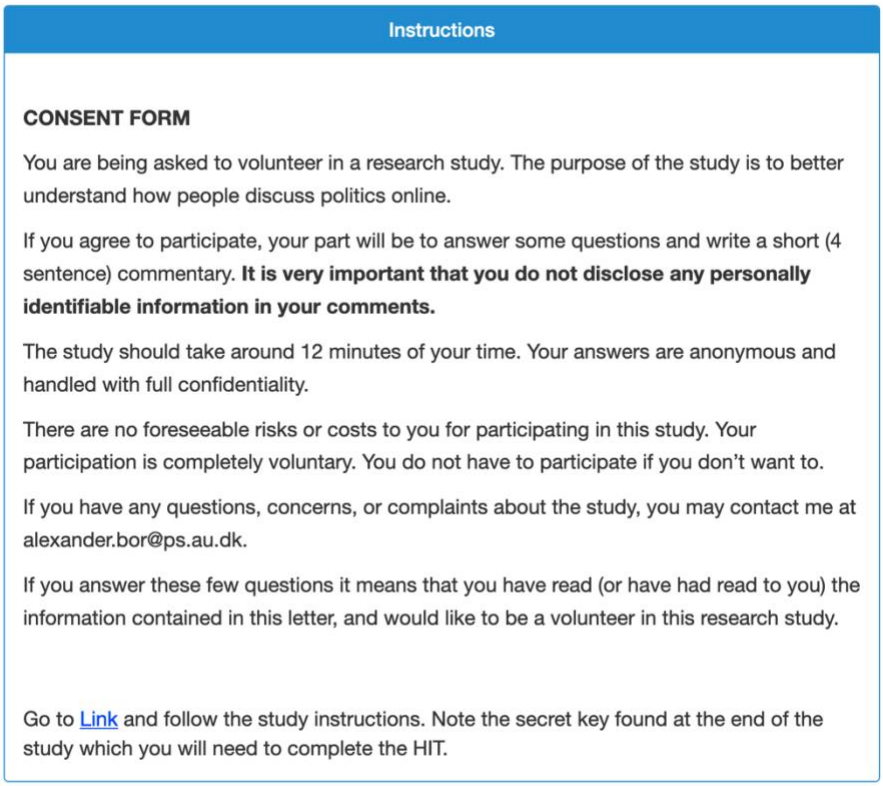
Table E2. No interaction between inappropriateness gap and offline hostility

	Dependent variable:	
	Online Hostility	
	(1)	(2)
Inappropriateness gap	-0.01 (0.02)	-0.01 (0.01)
Offline Hostility		0.84*** (0.02)
Inappropriateness gap × Offline Hostility		-0.02 (0.05)
Constant	0.18*** (0.01)	0.03*** (0.01)
Observations	770	770
Adjusted R^2	-0.001	0.70
Note:	*p<0.05; **p<0.01; ***p<0.001	

Appendix F. Full experimental materials for Studies 6 and 7

F1. Procedure of the experiments in Study 6 and 7

1. [BOTH] Giving informed consent on Amazon's Mechanical Turk upon taking the HIT associated with the studies (See formbelow).


Instructions

CONSENT FORM

You are being asked to volunteer in a research study. The purpose of the study is to better understand how people discuss politics online.

If you agree to participate, your part will be to answer some questions and write a short (4 sentence) commentary. **It is very important that you do not disclose any personally identifiable information in your comments.**

The study should take around 12 minutes of your time. Your answers are anonymous and handled with full confidentiality.

There are no foreseeable risks or costs to you for participating in this study. Your participation is completely voluntary. You do not have to participate if you don't want to.

If you have any questions, concerns, or complaints about the study, you may contact me at alexander.bor@ps.au.dk.

If you answer these few questions it means that you have read (or have had read to you) the information contained in this letter, and would like to be a volunteer in this research study.

Go to Link and follow the study instructions. Note the secret key found at the end of the study which you will need to complete the HIT.

2. [BOTH] Give a pledge to carefully read and follow instructions. Participants selecting the "I do not understand" button were screened out.
3. [BOTH] Basic demographics (age, sex, level of education, party ID)
4. [Study 3] Political attitudes: self-reported hostility, talking about politics (as reported in Appendix A).
5. [BOTH] Attitudes towards immigration following Gallup poll question[3]

   "Well over a million immigrants arrive to the United States each year. There is a big divide between the two major political parties about this issue: most Republican politicians are dissatisfied with the current level of immigration into the US and argue that it should be decreased. Meanwhile, most Democratic politicians are satisfied with the current level or even argue that the level of immigration should increase.

   What are your personal views on the current levels of immigration to the US?
   - I'm satisfied with current levels of immigration or want levels to increase
   - I'm dissatisfied with current levels of immigration or want levels to decrease
   - Neither satisfied nor dissatisfied"
6. [BOTH] Personality measures: Status-driven risk taking, Difficulties in emotion regulation, Trait aggression (as reported in Appendix A).
7. [Study 6] Choice between political and non-political post
8. [BOTH] Rating the hostility of the (chosen) post

[3] Tarrance, V. L. (2017). Can a "Nation of Immigrants" Reform 21st-Century Immigration? Retrieved August 23, 2019, from https://news.gallup.com/opinion/polling-matters/205304/nation-immigrants-reform-21st-century-immigration.aspx

9. [BOTH] Writing a reply to the comment as if it appeared on the respondent's own feed (Study 6) / matching the tone of the original comment (Study 7)

   [Study 6]
   "Now, please write a comment to the post you just read. Write about your own views just as you would if this was a real conversation on Facebook.

   Write 4 sentences and use your own words!"

   [Study 7]
   "Now, please write a comment to the post you just read. Write about your own views but do your best to match the tone of the post as closely as you can! We are not interested in what tone you would normally use, but in your ability to match the tone as closely as possible.

   Write 4 sentences and use your own words! Your answer will be verified to make sure you do not repeat too much from above."

10. [Study 7] Evaluate the hostility of the post from memory
11. [BOTH] Comprehension check

    Finally, we want to make sure you understood the comment writing task correctly.
    Please select from the list below, all instructions you were given!

    1. Match the tone of the comment
    2. Use your own words
    3. Be polite
    4. Be hostile
    5. [Study 6] Use your usual tone
    6. [Study 7] Write about your own views
    7. Repeat many words from the comment

    Note: Participants selecting neither option 2 or 6 in Study 6 and neither option 1 or 2 in Study 7 were excluded from the sample.

12. [Study 6] Self-reported honesty in replying as in real life
13. [BOTH] Debriefing. Participants were informed about the purpose of our research, reassured that they saw excerpts from a fictitious debate and encouraged to adhere to norms of civil public discourse in their life.

F2. On the target posts

Our motivation with attributing the target posts to Jackie Bennett was to have a generic, unisex name. We relied on a fivethirtyeight.com post[4] on unisex names, which suggests that Jackie is the 4th most common unisex name in the United states with a near even split (42% male). Jessie in Study 6 comes from the same place ranking 3rd and a similar even split (48% male). We conducted a small pilot test (N = 200) to ensure that the name of the author (Jackie Bennett vs Jessie Johnson) did not affect the likelihood of choosing the political comments. The target posts were created relying on a free online tool on Simitator.com to imitate Facebook's design.

[4] Flowers, Andrew. 2015. "The Most Common Unisex Names In America: Is Yours One Of Them?", FiveThirtyEight, Retreived on August 23, 2019. https://fivethirtyeight.com/features/there-are-922-unisex-names-in-america-is-yours-one-of-them/

Table F1 below displays the full wording of each of our 16 target FB posts. Figure F1 shows mean ratings hostility ratings for each of these posts from the pre-tests. Figure F2 offers a snapshot from the actual experiment, demonstrating the design participants interacted with.

Table F1. Full target post wordings from Studies 3 and 4.

	Pro-immigration	**Anti-immigration**
1	It is a **mistake** to think that current levels of immigration are not sustainable. Cheap immigrant labor benefits everyone. We should respect the rights of people living in this country more.	It is a **mistake** to think that current levels of immigration are sustainable. Crime and unemployment due to immigration hurt everyone. We should respect the culture of people living in this country more.
2	It is a **mistake** to think that current levels of immigration are not sustainable. Cheap immigrant labor benefits Republicans as much as everyone else. Police and border patrols should respect the rights of people living in this country more.	It is a **mistake** to think that current levels of immigration are sustainable. Crime and unemployment due to immigration hurt Democrats as much as everyone else. Immigrants should respect the culture of people living in this country more.
3	It is a **mistake** to think that current levels of immigration are not sustainable. Cheap immigrant labor benefits Republicans as much as everyone else. The **misguided** police and border patrols **bother** innocent people. They should respect the rights of people living in this country more.	It is a **mistake** to think that current levels of immigration are sustainable. Crime and unemployment due to immigration hurt Democrats as much as everyone else. **Misguided** immigrants **bother** innocent people. They should respect the culture of people living in this country more.
4	It is **absurd** to think that current levels of immigration are not sustainable. Cheap immigrant labor benefits Republicans as much as everyone else. The **misguided** police and border patrols **bother** innocent people. They should respect the rights of people living in this country more.	It is **absurd** to think that current levels of immigration are sustainable. Crime and unemployment due to immigration hurt Democrats as much as everyone else. **Misguided** immigrants **bother** innocent people. They should respect the culture of people living in this country more.
5	It is **absurd** to think that current levels of immigration are not sustainable. Cheap immigrant labor benefits t**wo-faced** Republicans as much as everyone else. The **misguided** police and border patrols **bother** innocent people. They should respect the rights of people living in this country more.	It is **absurd** to think that current levels of immigration are sustainable. Crime and unemployment due to immigration hurt **two-faced** Democrats as much as everyone else. **Misguided** immigrants **bother** innocent people. They should respect the culture of people living in this country more.
6	It is **stupid** to think that current levels of immigration are not sustainable. Cheap immigrant labor benefits t**wo-faced** Republicans as much as everyone else. **Ferocious** police and border patrols **harass** innocent people. They should respect the rights of people living in this country more.	It is **stupid** to think that current levels of immigration are sustainable. Crime and unemployment due to immigration hurt **two-faced** Democrats as much as everyone else. **Ferocious** immigrants **harass** innocent people. They should respect the culture of people living in this country more.
7	It is **extremely stupid** to think that current levels of immigration are not sustainable. Cheap immigrant labor benefits **hypocrite** Republicans as much as everyone else. **Ferocious** police and border patrols **harass** innocent people. They should respect the rights of people living in this country more.	It is **extremely stupid** to think that current levels of immigration are sustainable. Crime and unemployment due to immigration hurt **hypocrite** Democrats as much as everyone else. **Ferocious** immigrants **harass** innocent people. They should respect the culture of people living in this country more.
8	If you think that the current levels of immigration are not sustainable **you are stupid and irresponsible.** Cheap immigrant labor benefits **hypocrite** Republicans as much as everyone else. **Wicked and sadistic** police and border patrols **harass** innocent people. They should respect the rights of people living in this country more.	If you think that the current levels of immigration are sustainable **you are stupid and irresponsible.** Crime and unemployment due to immigration hurt **hypocrite** Democrats as much as everyone else. **Wicked and sadistic** immigrants **harass** innocent people. They should respect the culture of people living in this country more.

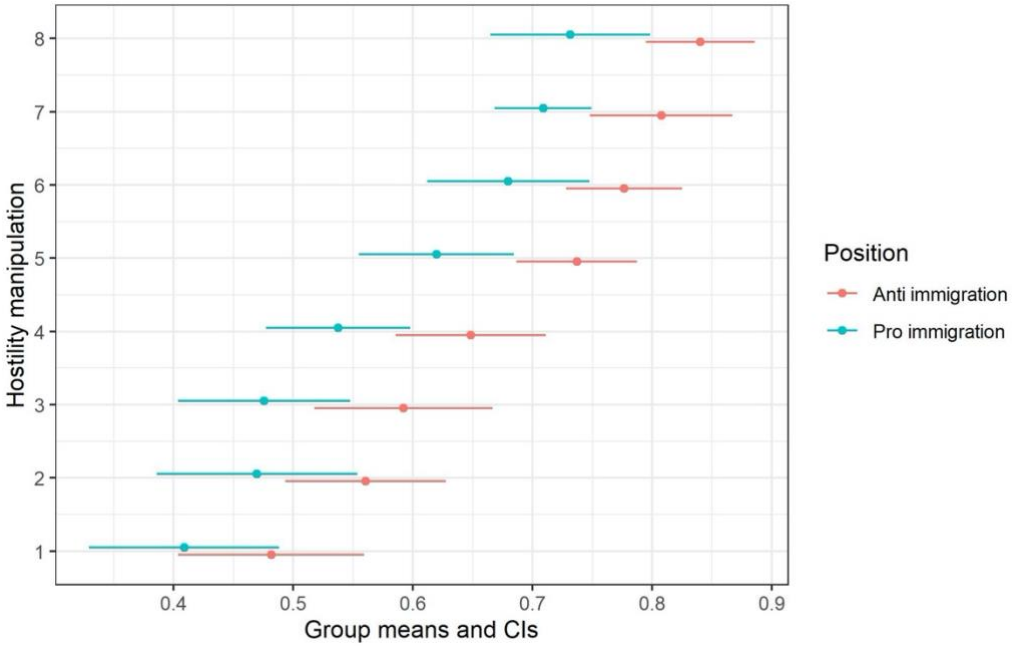


Figure *F1 Mean hostility ratings for each of the 16 target FB posts from the pre-test*

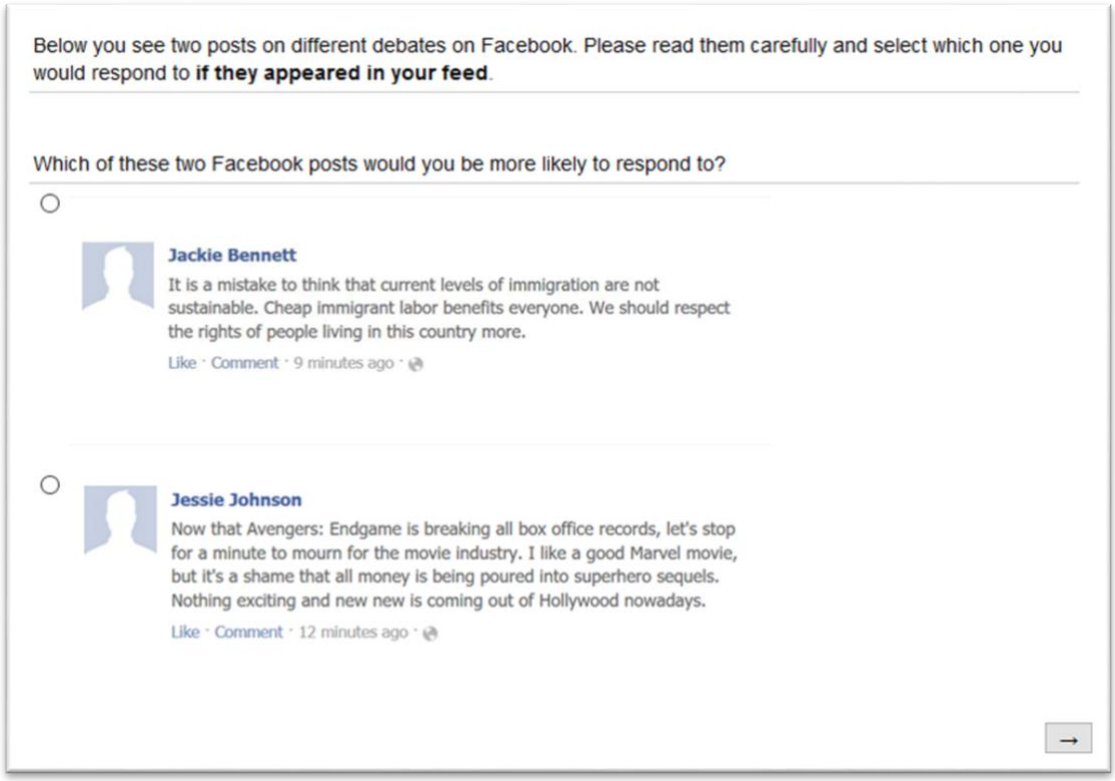


Figure F2 A snapshot from our experiment

F3. On crowdsourcing

Studies 6 and 7 rely on a behavioral measure of hostility. We asked participants to write a comment to a Facebook post experimentally manipulated on its issue position and level of hostility. To rate the hostility of these responses, we employ crowd-sourcing as a cheap and reliable option to reduce measurement error.

In the first phase of the crowd-sourcing, we asked MTurkers to rate the hostility of our target Facebook posts. We invited 1430 participants to rate 34 target posts in two rounds. Each participants rated only one comment to mimic the design of the experiment and to avoid demand effects due to seeing multiple versions of the stimulus. Specifically, we displayed the unformatted text and provided the following instructions. We deliberately avoided defining tone or hostility for participants, because we are most interested in participants' intuitive understanding of these concepts.

> *Below you see a comment related to the debate about current levels of immigration to the United States. Please read the comment carefully and rate its tone.*
> *[Stimulus Text]*
> *How would you rate the tone of this comment on a scale from 1 (extremely polite) to 7 (extremely hostile)? Extremely polite 1 2 3 4 5 6 7 Extremely hostile.*

Based on this exercise, we picked the 16 comments which were evenly distributed along the scale. These final comments were rated by 42.5 raters on average (range: 34-63).

Next participants, too, rated the hostility of the target comments, before they crafted their response. Here we employed a slightly modified scale:

> *How hostile is this post on a scale from 0 (not at all) to 100 (extremely)? (With slider anchored at 0) 0 Not at all; 25 Slightly; 50 Moderately; 75 Very; 100 Extremely*

As 781 people selected and rated a political comment in Study 6, and 1640 people participated in Study 7. This yields another 151 ratings per target on average (range: 116-172).

Finally, all responses from the two studies were crowd-sourced on MTurk relying on the latter scale. Because here each comment was unique, we could rely on multiple ratings per participant. We set the number of texts to 12 for each participant to avoid fatigue and attrition.

To determine the ideal number of ratings per comment, we ran a simulation based on the target posts' ratings. In particular, we bootstrapped two samples for each of the 16 target posts and calculated Krippendorff's alpha across these two sets of crowd-sourced hostility scores. Importantly, we varied the size of these two samples from 2 to 50 in small increments, to estimate how higher samples reduce measurement error and increase the reliability of the ratings. We found that the inflection point is at around 10 raters. Beyond these, adding new raters has quickly diminishing returns in precision. The reliability of the ratings passes acceptable levels (Krippendorff's alpha > 0.8) already at 5 ratings per comment. We commissioned 10 and 12 ratings per comment on average in Study 4 and Study 5 respectively (range: Study 4: 3-23, Study 5: 1-26). 98% and 99% of the comments received at least 5 ratings in Study 4 and 5, respectively.

F4. On relying on MTurkers

While MTurk samples are not representative of national populations, 1) they have been shown to replicate findings based on online representative samples (Coppock et al 2018 PNAS), and 2) they offer a diverse and computer-savvy sample of respondents to participate in our experiments. Relying on untrained raters to assess hostility mean we could let regular people to define what they find hostile, instead of coming up with a coding scheme which may or may not pan onto citizens' experiences. We compensated for the lower reliability of untrained raters by employing more of them.

Appendix G. Full model details and additional analyses for Studies 6 and 7

G1. Validation of experimental paradigm and self-reported behavioral DV

Studies 4 and 5 provide an opportunity to ask two important questions: First, is there a meaningful statistical relationship between the observed behavioral hostility within the experimental paradigm and well-established psychological constructs known to affect hostile behavior? Finding this relationship would established the face-validity of the experimental paradigm. Second, is there a meaningful statistical relationship between the observed behavioral hostility in the experiment and the novel self-reported hostility scale, which constitute the backbone of Studies 1-3? Finding this relationship would increase our confidence that the unexpected results from these studies are not due to measurement bias.

Do people with (vs without) aggressive personality traits write more hostile comments? Yes. As a validation of our experimental paradigm, we regress the crowd-sourced hostility ratings of the comments from Study 6 on three personality measures known to correlate with hostility: status-driven risk taking, difficulties in emotion regulation and trait aggression. As before, the independent variables are centered at their mean and divided by two standard deviations. Reassuringly, we find consistent correlations between these measures and the hostility of the responses (see Table G1,

columns 1-3). Trait aggression (column 3) has the largest effect at 5 percentage points ($p < 0.001$) corresponding to a two standard deviation change in the independent. A similar change in SDRT and DERS correspond to a roughly 2 percentage points change in the comment hostility, on average ($p = 0.057$ and $p < 0.05$, respectively). Although these effects may seem small, it is worth remembering that our dependent measure is a single, short comment in an experimental setting. We would expect, these relationships to be stronger in longer interactions with real-stakes. The important takeaway is that the behavioral dependent measure in our novel experimental paradigm has face-validity.

Table G1. Study 6 - Regressing hostility of comment on personality measures

	Dependent variable:			
	Hostility of comment			
	(1)	(2)	(3)	(4)
Status drive	0.06 (0.03)*			
Emotion regulation		0.05 (0.03)**		
Trait aggression			0.16 (0.03)***	
Self-reported hostility				0.13 (0.04)***
Age	0.05 (0.04)	0.05 (0.04)	0.07 (0.04)*	0.05 (0.04)
Female	0.005 (0.01)	-0.004 (0.01)	0.003 (0.01)	0.003 (0.01)
Party ID	0.07 (0.02)***	0.07 (0.01)***	0.07 (0.01)***	0.07 (0.01)***
Higher educated	-0.02 (0.01)**	-0.02 (0.01)*	-0.02 (0.01)	-0.02 (0.01)**
Constant	0.24 (0.02)***	0.24 (0.02)***	0.18 (0.03)***	0.25 (0.02)***
Observations	781	781	781	777
Log Likelihood	344.10	344.13	352.31	347.03
Akaike Inf. Crit.	-670.21	-670.27	-686.62	-676.06
Bayesian Inf. Crit.	-628.26	-628.32	-644.68	-634.16
Note:				*p<0.1; **p<0.05; ***p<0.01

G2. Selection hypothesis

The choice between the manipulated political post with varying levels of hostility and the non-political posts also allows to test open questions related to the selection hypothesis. Thus, if non-hostile individuals are specifically motivated to avoid hostile political discussions, we should find (3) that non-hostile individuals are more likely to select the political post at lower levels of hostility, whereas hostile individuals are less influenced by the hostility of the post. In fact, following the original formulation of the selection hypothesis, we might find that hostile individuals are drawn to hostile political debates and, therefore, even more likely to select the political post when it is more hostile. In contrast, if the main difference between hostile and non-hostile individuals is that non-hostile individuals are motivated to avoid online political discussions altogether, we should find that measures of hostile predispositions influence the choice of the political post, irrespective of its hostility.

Does the hostility of political messages reduce the likelihood that people with non-hostile predispositions select into a political discussion? We build a hierarchical, linear probability model by regressing a dummy variable – coded as 1 if the political post was chosen – on status-driven risk taking and trait aggression, adding varying intercepts separately for the two experimental manipulations. Neither adding varying slopes for status-driven risk taking nor varying intercepts for all 16 combinations of the two manipulations improves model fit ($p = 0.64$). To avoid overfitting and for the sake of simplicity, we do not add these terms.

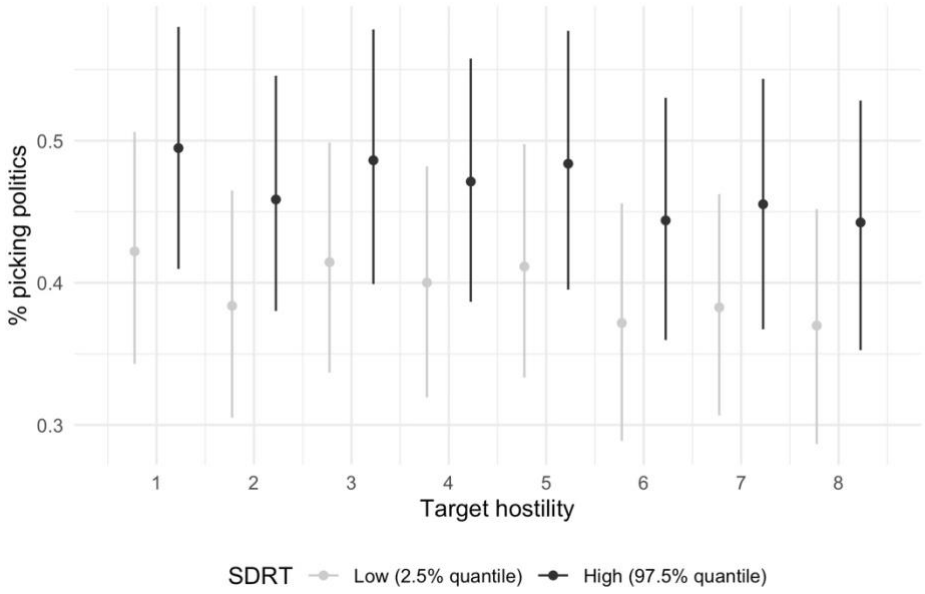


Figure G1. *Predicted Probability of Choosing the Political Target Post as a Function of Status-Driven Risk Taking (shade) and the Hostility of the Target (x)*

Figure G1 shows the predicted probability of choosing the political post (y-axis) as a function of status-driven risk taking (shades of grey) and the eight hostility versions of the target post (x). For the sake of simplicity, we average over the issue position (pro- vs anti-immigration) of the target, although it is worth nothing that the post taking an anti-immigrant stance was chosen three percentage points more often.

We find that people higher in status-driven risk taking ($\beta = 0.11$, $p = 0.08$) and trait aggression ($\beta = 0.18$, $p < 0.05$) are more likely to pick the political post. Consistent with the results of Studies 1 and 2, hostile people have an enhanced preference for participating in political discussions. However, the hostility of the target has an equally (weak) negative effect on people high and low on status-driven risk taking and trait aggression alike. There is a six-percentage-point difference in the average probability of choosing the least and the most hostile target post in our design (bootstrapped 95% CI: 0.00, 0.12). That is, non-hostile individuals are not more likely to select into non-hostile political discussions and hostile individuals are not more likely to select into hostile discussions. Thus, the selection results from Studies 1 and 2 most likely reflects that non-hostile individuals are motivated to avoid every political discussion but find it more difficult to do so in offline environments.

Table G2. Multilevel logistic regression on selecting political post

	Dependent variable:	
	Selecting political post	
	(1)	(2)
Status drive	0.11 (0.06)*	
Aggression		0.18 (0.07)**
Age	0.61 (0.08)***	0.61 (0.08)***
Female	0.004 (0.02)	-0.001 (0.02)
Party ID	-0.02 (0.03)	-0.02 (0.03)
Higher education	-0.01 (0.02)	0.002 (0.02)
Constant	0.24 (0.05)***	0.20 (0.05)***
Observations	1,922	1,922

Log Likelihood	-1,347.24	-1,345.35
Akaike Inf. Crit.	2,712.48	2,708.69
Bayesian Inf. Crit.	2,762.53	2,758.74
Note:		*p<0.1; **p<0.05; ***p<0.01

Table G3. Varying intercepts for the models in Table G2

Random effect terms	Levels	Model 1	Model 2
Hostility	1	0.273	0.231
	2	0.235	0.193
	3	0.265	0.224
	4	0.249	0.208
	5	0.263	0.221
	6	0.222	0.181
	7	0.233	0.192
	8	0.218	0.178
Position	Anti	0.264	0.223
	Pro	0.225	0.184

G3. Change and perception hypotheses

Average tone matching in Study 7

Our paper reports hostility difference scores relying on the crowd-sourced hostility ratings of the 16 target posts from Study 7. Here, we replicate these results with relying on the hostility ratings of the 781 participants who selected the political comment in Study 6. It is notable that despite the fact the pool of raters in Study 6 self-selected into the politics topic (albeit not necessarily conscious of the choice that they will be asked to rate its hostility), their average hostility ratings are very highly correlated with the ratings of Study 7 participants ($r = 0.96$, $p < 0.001$). Consequently, it is not surprising that results replicate using Study 6 participants' hostility ratings as a benchmark for tone-matching ($m = -0.12$, $t(1639) = 27.1$, $p < 0.001$).

Full model details for multilevel models addressing individual differences in tone-matched comments' hostility

Table G4. Regressing tone matched comments' hostility on SDRT and DERS

	Dependent variable:	
	Response hostility	
	(1)	(2)
SDRT	0.01 (0.02)	
DERS		0.003 (0.02)
Age	0.02 (0.02)	0.02 (0.02)
Female	-0.002 (0.01)	-0.003 (0.01)
Party ID	0.07 (0.01)***	0.07 (0.01)***
Higher education	-0.01 (0.01)	-0.01 (0.01)
Constant	0.29 (0.04)***	0.29 (0.04)***
Observations	1,638	1,638

Log Likelihood	698.17	697.98
Akaike Inf. Crit.	-1,378.34	-1,377.96
Bayesian Inf. Crit.	-1,329.73	-1,329.35
Note:		*p<0.1; **p<0.05; ***p<0.01

G4. Relationship between encoded and recalled hostility in Study 7

After they finished writing their comment, participants in Study 7 rated the hostility of the target post again, but this time without seeing the comment again, relying on their memory. This measure allows us to compare encoded and recalled hostility perceptions and reveal that participants are highly accurate at recalling their earlier hostility ratings (Pearson's $r = 0.93$, $p < 0.001$). Consequently, although the process of recall could create an opportunity to various biases to creep in, this does not appear to be the case. We test for this possibility in two ways. Table G5 below demonstrates that the partial correlation between response hostility and the perceived hostility of the target is *not* moderated by difficulties in emotion regulation. Yet, as linear interaction effects take many assumptions, we also calculate the simple correlation between subjective target assessment and response hostility at low, medium and high levels of DERS after creating three equally sized bins. We find that if anything, medium and high DERS respondents calibrate their responses slightly better then low DERS respondents (low: r (95%CI) = 0.34, (0.27, 0.41), medium: r (95%CI) = 0.42, (0.34, 0.49), high: r (95%CI) = 0.4, (0.33, 0.47))

Table G5. Relationship between subjective hostility of the target and response hostility

	Dependent variable:
	Response hostility
Perceived target hostility	0.10 (0.02)***
Emotion regulation	-0.03 (0.04)
Age	0.02 (0.02)
Female	-0.002 (0.01)
Party ID	0.07 (0.01)***
Higher education	-0.01 (0.01)
P. t. Hostility × Emotion regulation	0.03 (0.07)
Constant	0.26 (0.03)***
Observations	1,638
Log Likelihood	714.93
Akaike Inf. Crit.	-1,407.86
Bayesian Inf. Crit.	-1,348.45
Note:	*p<0.1; **p<0.05; ***p<0.01