Polarising punchlines: The influence of inter-group humour on partian affective polarisation

Yara Sleiman¹^{*}, Maria Grasso¹, Charles Pattie², Laura Serra⁴, Carlo Hofer³

¹Queen Mary University of London

 2 University of Sheffield

 3 University of Warwick

⁴London School of Economics and Political Science

March 14, 2024

Abstract

Modern political discourse increasingly transcends traditional platforms, engaging audiences through social media, television, and online forums. This shift has inspired a surge in humour as a primary means of political expression, making it a central element of modern political participation. Nonetheless, the effect of humour on citizens' political attitudes remains largely unexplored. In this study, we field a survey experiment with 2,011 respondents from Great Britain to test the influence of disparaging humour on affective polarisation among partisans. Using a 2x3 factorial design, we manipulate both the tone (humorous vs. nonhumorous) and the target (in-group, out-group, no target) of political content. We find that messages targeting in-partisans exacerbate affective polarisation; however, when these messages are delivered humorously, the effect is reversed resulting in a significant decrease in polarisation. The depolarising effect of humour is attributed to changes in content perception and the discounting of contentious messages.

^{*}Corresponding author email: y.sleiman@qmul.ac.uk.

1 Introduction

In recent decades, affective polarisation has emerged as a significant concern among social scientists due to its negative effects on community cohesion, democratic norms, and the potential to incite violence (Broockman et al. 2023, Iyengar et al. 2019, Kingzette et al. 2021, McCoy et al. 2018). Scholars consistently attribute the rise in polarisation to digital information-sharing patterns that amplify pre-existing beliefs and limit exposure to opposing viewpoints. These studies primarily manipulate the political orientation of online content to gauge its impact on political attitudes (Bakshy et al. 2015, King et al. 2017, Prior 2013). However, online platforms provide a diverse range of methods through which information is communicated, such as articles, videos, and memes. Each medium possesses unique attributes that can significantly shape how individuals interpret information and respond to it (Highfield & Leaver 2016, Shahbaznezhad et al. 2021). Indeed, in many cases, 'the medium is more important than the message' (McLuhan 2017), yet virtually no studies examine how the mode of communication influences the degree to which online political messages can (de)polarise users.

We contribute to the conversation on the polarising impact of social media by studying a growing medium of digital political expression: humour. Humour inherently trivialises its subject, inviting people to approach it in a more light-hearted and casual manner. When aimed at particular social groups or their members, humour conceals prejudice in a layer of entertainment, making it seem socially acceptable (Bill & Naus 1992, Johnson 1990). Today, with the advent of the digital era and heightened sensitivity to offensive speech, the use of such disparaging humour is more pervasive than ever before. Scholars have positioned it as more potent than traditional news in fostering selective exposure and an echo chamber effect (Messing & Westwood 2014). It has also garnered increasing attention from the media surrounding the electoral successes of Donald Trump and Boris Johnson and their strategic employment of humour as a shield against criticism (Karnitschnig 2019, Weiss, Joanna 2019).

In this study, we ask whether inter-group political humour influences affective polarisation. Existing research suggests that humour significantly shapes political attitudes and perceptions of political figures (Baumgartner & Morris 2008, Becker 2012, Xenos et al. 2011). This influence is particularly moderated by partisanship and ideology, which shape both individuals' preferences for humour and their responses to it (Becker 2014, Goldthwaite Young 2004, Hmielowski et al. 2011). However, it remains unclear whether humour, as a medium of expression, uniquely influences polarisation apart from political content more generally. On one hand, humour can place its target as an object of ridicule and sharpen distinctions between groups (Ford et al. 2017). On the other hand, its disarming effects could act as a buffer against inter-group hostility (Gruner 2017, Meyer 2000).

To answer this question, we field a survey experiment in Great Britain where 2,011 participants are randomly exposed to six treatment conditions that vary in message tone (humorous, non-humorous) and target (in-group, out-group, no target). Affective polarisation is measured post-treatment through thermometers, trait ratings, and social-distance items. We also elicit participants' emotional and attitudinal responses to the content to identify potential causal mechanisms.

Consistent with previous studies, we find that exposure to counter-attitudinal messages causes an increase in affective polarisation (e.g. Bail et al. 2018, Garrett 2009, Garrett et al. 2014). Interestingly, this effect reverses when the same message is delivered humorously: compared to the control group, participants exposed to humorous messages mocking their political group exhibited a significant decrease in affective polarisation. A causal mediation analysis shows that humour reduces polarisation by changing how people perceive contentious messages: either making them less likely to take the content seriously (discounting) or improving their views toward the content.

The paper proceeds as follows: Section 2 includes the theoretical framework and research hypotheses. Section 3 describes our experimental design. In Section 4, we discuss the results of our balance and manipulation checks. This is followed by a description of our main estimation strategy in Section 5 and results in Section 6. Section 7 includes robustness checks and Section 8 concludes.

2 Inter-group humour and polarisation

Humour plays an increasingly important role in politics, with a significant portion of the population, particularly younger individuals, citing satirical programs like Saturday Night Live and The Daily Show as key sources of political information (Becker 2012). Despite its prevalence, the exact influence of humour on attitude change remains unclear. Some studies suggest that humour can successfully sway public opinion (Eisend 2009), while others conclude that it can detract from the message's persuasive power (English et al. 2011). This has led to the view that humour is a "double-edge sword" (Meyer 2000) and generalisations about its effects can be misleading (Weinberger & Gulas 1992). Instead, academic discussions have shifted towards identifying moderating variables that might clarify the role of humour in communication. We discuss these factors and the ways in which they may correlate with partian affective polarisation.

First, humour can reduce polarisation through two pathways that we categorise broadly as emotional and attitudinal mechanisms. According to Meyer (2000)'s relief theory, humour alleviates tension and fosters positive emotions, thereby leading to a counter-arousal state. This state can soften negative perceptions towards the target of humour and, by extension, mitigate affective polarisation. This is further supported by Kuiper et al. (1995)'s mood maintenance approach, which indicates that a good mood induced by humour can decrease the production of counterarguments and result in less critical engagement with contentious political content. On the attitudinal side, humour enhances the perceived credibility of the source, even when such credibility is not warranted (Gruner 2017). This might reduce polarisation if the content is perceived as likeable or more trustworthy. Similarly, humorous messages are often discounted as 'just a joke', intended more for entertainment than for serious information dissemination (Nabi et al. 2007). This leads to the exclusion of humorous messages from serious judgement processes, and diminishes the impact of potentially divisive content.

More recent empirical research incorporates Tajfel (1982)'s social identity theory into the analysis of political humour (e.g. Abrams & Bippus 2011, 2014, Baumgartner & Morris 2008, Becker 2021). In this regard, scholars argue that political humour specifically of an inter-group nature - can sway public opinion by sharpening distinctions between groups. The *target* of humour emerges as a key moderating factor within this framework: individuals are more likely to appreciate humour that targets those they oppose, while they may react unfavourably to humour directed at in-group members (Zillmann & Cantor 2017). This suggests that political humour, similar to other political media content, is processed through the lens of inter-group prejudices, where partisanship and ideological leanings significantly shape the consumption and effect of humorous content (e.g. Becker 2014, Goldthwaite Young 2004, Hmielowski et al. 2011), though mixed findings also prevail.

For instance, Baumgartner & Morris (2008) observed that The Colbert Report inadvertently bolstered support for Republican politicians, contrary to its intent to support Democrats. In contrast, Xenos et al. (2011) find that Republican audiences developed a more favourable view of Nancy Pelosi and Congressional Democrats after viewing them being mocked on The Daily Show. Similarly, Becker (2012) noted that both Republican and Democratic viewers evaluated a Republican target more negatively after exposure to critical content from The Colbert Report.

Evidence on non-political inter-group humour is also mixed. For example, Abrams & Bippus (2011) find that sexist jokes contribute to inter-group hostility: both men and women exhibit in-group bias by rating jokes about the opposite gender funnier and more typical than jokes about their own gender. However, in another study, women demonstrate in-group rejection and experience lower in-group favouritism in response to sexist jokes about them (Abrams & Bippus 2014). This discrepancy is also a reflection of the mixed findings in the general political information processing literature, specifically the discourse surrounding the consumption of politically congruent (pro-attitudinal) and incongruent (counter-attitudinal) content (Guess & Coppock 2020).

The purpose of this article is not to adjudicate among these studies but instead to (1) extend the scope of analysis to the study of partisan affective polarisation and (2) address empirical limitations that confound the effects of inter-group humour in previous studies. Prior experimental work analyses viewer responses to pre-existing video clips focused on particular political figures or campaigns. This approach might not fully capture enduring political sentiments and may not effectively separate the impact of the comedic content from associated source cues. Moreover, most of these studies rely on student samples, which limits their generalisability to broader populations. Our study uses a 2x3 factorial design with a uniform set of stimuli on a representative British sample, ensuring that the independent influence of humour is isolated from confounding influences. Importantly, our design uniquely tests the varying influence of humour on affective polarisation - contingent upon its intended target. We outline our pre-registered¹ hypotheses with regards to the (de)polarising influence of humour and its target below.

2.1 Research hypotheses

In our analysis of the main effect of humour on affective polarisation, we propose competing hypotheses to capture the dual function of humour. On one hand, its ability to reinforce inter-group biases could intensify affective partian divides. On the other, its disarming effects and potential to induce positive emotions could serve as a buffer

¹Our pre-analysis plan can be found here.

against inter-group animosity. Therefore, we test the following competing hypotheses with regards to the main effect of humorous content:

- H1a: Humorous content decreases affective polarisation compared to non-humorous content.
- H1b: Humorous content increases affective polarisation compared to non-humorous content.

Further, we expect that content with a partial slant - regardless of its nature - will increase affective polarisation compared to neutral content given the tendency of partial media to reinforce group identities and exacerbate inter-group distinctions (e.g. Levendusky 2013):

• H2: Both humorous and non-humorous content that targets a political group will significantly increase affective polarisation compared to neutral content.

However, given the varied evidence in existing literature, we hold no strong a priori assumptions regarding differences between in- and out-partisan content on affective polarisation and its interaction with humour. Thus, we present these predictions as null hypotheses.

- H3: There are no significant differences between in-group and out-group targeted messages in their effect on affective polarisation.
- H4: The impact of humour on affective polarisation does not differ significantly across the different target groups.

3 Experimental design

We employ a 2x3 factorial design to examine the influence of the tone and target of political messages on partian affective polarization. Participants are randomly assigned to one of six treatment arms generated by crossing two primary variables: the tone of message (humorous and non-humorous) and the target of the message (in-group, out-group, and non-political). Table 1 below presents our experimental conditions and Appendix E and F include the full questionnaire and stimulus material.

We first collect baseline demographic and political affiliation information before exposing participants to the stimulus material. Post-treatment, participants are asked to report their emotional and attitudinal responses to the content, alongside a series of conventional questions designed to measure affective polarisation through feeling thermometer ratings, social distance measures, and trait ratings. The experiment is administered on a nationally representative sample of 2,011 respondents from Great Britain aged 18 and over, recruited via Deltapoll².

Target Tone	Right-wing	Left-wing	None
Humoroug	1	2	3
numorous	(n = 328)	(n = 337)	(n=336)
Non humorous	4	5	6
mon-numorous	(n = 343)	(n = 332)	(n = 335)

Table 1: Experimental conditions

The humorous content, aimed at both left-wing and right-wing participants, consists of 8 jokes that highlight common stereotypes and criticisms, using satire and exaggeration. The non-humorous content mirrors the themes of the jokes but is presented in a straightforward, serious tone. The political stimulus material is selected to evoke humour through slightly disparaging narratives, with the aim of engaging participants with content that humorously critiques political groups and ideologies. This approach is meant to activate inter-group sentiments more effectively than softer forms of humour. Control conditions expose participants to either humorous (condition 3) or non-humorous (condition 6) content unrelated to politics, focusing instead on the British weather.

The classification of treatments as targeting the "in-group" or "out-group" is tied to participants' self-reported political ideology. Participants indicate their ideological placement on a scale where 0 represents the left end of the political spectrum and 10 represents the right. For those who select "5" (indicating a central position), "Don't know", or "Prefer not to answer", a follow-up question asks them to clarify if they lean more towards the left or the right. Those who identify as left-wing and are exposed to

 $^{^{2}}$ All responses are anonymised to ensure confidentiality. Participants are compensated for their time and are informed about their rights. Before participating, they provide consent, which includes understanding their ability to withdraw from the study at any time without penalty. Exclusions are applied for failed attention checks or illogical responses.

treatment conditions 2 or 5 are classified as having viewed an "in-group" stimulus. Leftwing participants exposed to treatment conditions 1 or 4 are classified as having viewed an "out-group" stimulus. The same logic applies inversely for participants identifying as right-wing. The analysis that follows excludes centrists (n = 518) unless otherwise specified.

4 Balance and manipulation checks

To ensure that the treatment effect is not confounded by pre-treatment differences across groups, we conducted balance checks on a range of covariates including education, gender, region, ideology, ethnicity, and political attention. The findings from these checks are detailed in Appendix A. Overall, we only find modest differences in pretreatment covariates across the experimental conditions. Despite this, we maintained ethnicity, ideology, and political attention as control variables in the analyses due to their statistical significance in preliminary tests. However, both excluding these covariates and including the remaining covariates does not substantively alter our results (see Robustness Checks in Section 7.)

Next, we conduct manipulation checks to assess the effectiveness of our experimental treatment, specifically examining the impact of the humorous treatment on participants' perceptions of humour. These checks are important for validating that the intervention was perceived as intended by the study participants. We test the influence of the humorous treatments, relative to the non-humorous conditions on two key variables: 'Perceived Humour' and 'Joke Perception' (Table A2 in Appendix A). The former is a 10-point scale that measures the extent to which participants found the content humorous, while the latter measures whether participants believed the authors of the messages were joking on a 5-point agreement scale. Overall, the results provide strong evidence that our experimental manipulation successfully influenced participants' perceptions in the intended directions. Participants perceived the humorous content as significantly funnier ($\beta = 1.044, \rho < 0.01$) and were more likely to view it as a joke ($\beta = 0.569, \rho < 0.01$) compared to the non-humorous content.

5 Estimation

Our 2x3 design allows us to test the main effects of humour and target group and their interaction (Humour * Target) on affective polarisation. We measure affective polarisation in three different ways using feeling thermometer ratings, social distance measures, and trait attributions. All variables are rescaled to range from 0 to 1. Appendix B describes how we measure the outcome variables in more detail. Given the continuous nature of our outcome variables, we employ linear regression models specified as follows:

$$Y_{ijk} = \beta_0 + \beta_1 \operatorname{Humor}_i + \beta_2 \operatorname{Target}_j + \beta_3 (\operatorname{Humor}_i \times \operatorname{Target}_j) + \mathbf{X}'_{ijk} \beta + \varepsilon_{ijk}$$
(1)

Where Y_{ijk} represents the affective polarisation measure for participant k in the treatment condition defined by the *i*th level of humor (humorous or non-humorous) and the *j*th target type (in-group, out-group, or no target). The terms $Humor_i$ and $Target_j$ denote respectively the indicator for the type of humour and the 3-level factor variable for the target of the message, while $Humor_i \times Target_j$ captures their interaction, allowing us to test the combined effect of message type and target on affective polarisation. \mathbf{X}'_{ijk} represents the vector of control variables and ε_{ijk} is the error term. We also employ the same specification to test the influence of the treatments of the components that make up the affective polarisation measures, namely out-group hostility and in-group favouritism.

6 Results

Table 2 presents the effects of humour and message targeting on the three distinct measures of affective polarisation: thermometer ratings, traits distance, and social distance. In Table 3, we present the regression output for the individual components of affective polarisation - out-group hostility (Panel A) and in-group favouritism (Panel B). We include the full regression output with the control variables in Tables A9, A10, and A11 of Appendix G.

	Dependent variable: Affective Polarisation			
	Thermometer	Traits	Social Distance	
	(1)	(2)	(3)	
Target				
In-group	0.035	0.059***	0.021	
	(0.026)	(0.018)	(0.018)	
Out-group	0.008	0.014	0.015	
	(0.026)	(0.018)	(0.018)	
Tone				
Humorous	0.024	0.026	0.005	
	(0.026)	(0.018)	(0.018)	
Target*Tone				
In-group*Humorous	-0.071^{*}	-0.070^{***}	-0.011	
	(0.036)	(0.025)	(0.025)	
Out-group*Humorous	-0.007	-0.001	0.018	
	(0.037)	(0.025)	(0.026)	
Observations	1,488	1,488	1,488	
Adjusted \mathbb{R}^2	0.071	0.115	0.033	

Table 2: Effects of humour and target group on affective polarisation

Note: This table presents the estimated coefficients from linear regression models testing the influence of humour and its target on affective polarisation. The reference groups are 'non-humorous' and 'no target' respectively. Affective polarisation is measured using three different scales: thermometers, trait ratings, and social distance. The models control for political attention, left-right self-placement, and ethnicity. Survey weights are applied. Standard errors are in parentheses. Refer to Table A9 of Appendix G for the full regression output.

$$p^* < 0.1; p^* < 0.05; p^* < 0.01$$

Our analysis reveals three key findings regarding the influence of humour and target on affective polarisation and its components. First, contrary to our initial hypotheses (H1a and H1b), we find that humour alone does not significantly influence affective polarisation. This observation suggests that the mere exposure to humour neither inherently polarises nor depolarises political attitudes. Second, we observe a significant increase in affective polarisation when messages target one's own political group. This is primarily attributed to heightened out-group hostility (In Panel A $\beta_{\text{In-group}} = 0.059^{***}$ (thermometer ratings) and 0.044^{***} (trait ratings)), indicating that in-group targeted messages exacerbate negative perceptions of out-partisans. These findings support our hypothesis that messages targeting a political group would increase polarisation compared to neutral content (H2) and we reject the null that there are no significant differences between the effects of in-group and out-group targeted political messages on affective polarisation (H3). Additional analyses presented in Appendix G, where the *Target* variable is re-coded into binary categories (in-group vs. out-group and targeted vs. non-targeted), further support these hypotheses.



Figure 1: Effect of humour on affective polarisation by target group *Note:* This figure presents the interaction effects of target group and humour treatments on affective polarisation (from Table 2 Model 2). The x-axis represents the humour treatment, with 'Non-humorous' and 'Humorous' categories. The y-axis shows the estimated fit for affective polarisation. The lines represent the different target groups: 'Control', 'In-group target', and 'Out-group target'.

Third, and most notably, humour significantly moderates the impact of in-group targeted messages on polarisation. This finding is presented more clearly in Figure 1 where we plot the predictive margins of target group from Model 2, separated by humorous and non-humorous treatments. Compared to the control, affective polarisation increases by 5.9 percentage points when the messages target the in-group non-humorously. This effect reverses when humour is introduced: affective polarisation drops by 4.37 percentage points from the non-humorous in-group baseline. The results in Table 3 indicate that this occurs through a dual mechanism: humour targeting the

in-group decreases polarisation by reducing out-group hostility (-0.078^{**}) and curbing in-group favouritism (-0.049^{**}) . Humorous messages targeting out-groups do not exhibit the same de-polarising effect. Instead, respondents maintain a level of polarisation comparable to the control, shifting insignificantly from 20.90 percent to 23.40 percent.

	Panel A: Out-group Hostility				
	Thermometer	Traits	Social Distance		
	(1)	(2)	(3)		
Target					
In-group	0.059***	0.044***	0.042^{*}		
	(0.022)	(0.015)	(0.021)		
Out-group	-0.021	0.017	0.009		
	(0.021)	(0.015)	(0.021)		
Tone					
Humorous	0.009	0.003	-0.013		
	(0.021)	(0.015)	(0.021)		
Target*Tone					
In-group*Humorous	-0.078^{**}	-0.032	0.008		
	(0.030)	(0.022)	(0.030)		
Outgroup*Humorous	0.054^{*}	0.019	0.058^{*}		
	(0.031)	(0.022)	(0.030)		
Observations	1.488	1,488	1.488		
Adjusted \mathbb{R}^2	0.052	0.126	0.020		
	Panel B: In-group Favouritism				
	Thermometer	Traits	Social Distance		
	(1)	(2)	(3)		
Target					
In-group	-0.033^{*}	0.016	-0.016		
	(0.020)	(0.015)	(0.020)		
Out-group	-0.031	-0.031^{**}	-0.010		
	(0.020)	(0.015)	(0.019)		
Tone					
Humorous	-0.010	0.018	0.006		
	(0.020)	(0.015)	(0.020)		
Target*Tone					
In-group*Humorous	0.018	-0.049^{**}	-0.033		
	(0.028)	(0.021)	(0.028)		
Outgroup*Humorous	0.035	0.017	-0.012		
	(0.028)	(0.022)	(0.028)		
Observations	1,488	1,488	1,488		
Adjusted \mathbb{R}^2	$_{0.079}13$	0.068	0.073		

Table 3: Effects of humour and target group on out-group hostility and in-group favouritism

*p<0.1; **p<0.05; ***p<0.01

6.1 Causal mechanisms

The main findings suggest that exposure to counter-attitudinal information (i.e. ingroup targeted) can reduce affective polarisation when presented humorously. However, the causal mechanisms behind this effect remain unclear. Does humour mitigate polarisation by modifying participants' emotional responses, or by changing their attitudes toward the content? To answer this question, we employ the causal mediation approach developed by (Imai, Keele & Tingley 2010, Imai et al. 2011), using the mediation package in R (Tingley et al. 2014). This approach allows us to separate the direct effect of humour from its indirect effects - i.e. the mediating roles of emotional and attitudinal reactions.

To accurately identify the mechanisms behind the influence of humour, we focus our analysis specifically on comparing in-group humorous versus in-group non-humorous conditions as the primary treatment variable. The key estimand of interest is the indirect mediating effect, or the average causal mediation effect (ACME), which captures the influence of humour on polarisation that operates through changes in emotions and attitudes toward the content. We present this estimate for each of our mediating variables in Figure 2 below. The emotional mechanism is captured through two variables. *Positive emotions* combines responses to questions measuring feelings of happiness and hopefulness elicited by the content. *Negative emotions* combines responses to questions on feelings of anger, sadness, frustration, and anxiety after viewing the content. For the attitudinal mechanism, we focus on message discounting and content perceptions. Discounting is a scale derived from combining responses to questions on whether the content is perceived as mere joking, entertainment, serious commentary, or dismissible humour. This scale captures the extent to which content is downplayed or dismissed. Content perception is measured through two variables: *Content accuracy* is gauged by participants' agreement with the statement that the content accurately describes its subject matter. *Content approval* is measured by participants' level of agreement with the statement 'My overall feeling toward the content is positive'. All variables have been re-scaled to range from 0 to 1.



Figure 2: ACME of attitudinal and emotional mediators

Note: This figure presents the results of separate mediation analyses examining the indirect effects of in-group humorous (vs. non-humorous) content on affective polarisation through various mediators. The traits-based measure of affective polarisation is used here. The average causal mediation effect (ACME) is depicted for each mediator, along with 95% confidence intervals. The mediators included in the analysis are discounting, negative emotions, positive emotions, content perception, and content accuracy. The dashed line at y = 0 indicates no mediation effect. Survey weights are applied. Refer to Appendix C for the full mediation results.

Overall, we find evidence in support of the attitudinal mechanism. We observe significant mediation effects for both discounting ($ACME = -0.0109, \rho < 0.001$) and content perception ($ACME = -0.0167, \rho < 0.001$). Discounting explains 21.83 percent and content perception explains 36.97 percent of the total effect of humour. Essentially, humour reduces polarisation by changing how people perceive contentious messages: either making them less likely to take the content seriously (discounting) or improving their views toward the content (content perception). We find no evidence in support of the emotional mechanism.

7 Robustness checks

We conduct robustness checks aimed at addressing potential concerns related to model specification, treatment effect heterogeneity, and the influence of unmeasured confounders in the mediation analysis. First, we show that the results are robust to variations in model specification by comparing outcomes across three scenarios: without control variables, with the demographic controls excluded from the main analysis, and with the full set of control variables (Table A12 in Appendix G). The results show no substantive differences in the treatment effect across model specifications. Second, we examine treatment effect heterogeneity by incorporating interactions between our treatment variables and the following variables: gender, ideology, age, and ethnicity (Appendix D). We find no significant differences that detract from the substantive conclusions drawn from the main results, suggesting that the treatment effects are broadly consistent across different subgroups. Sub-setting the analysis to centrists - participants who do not identify as right-wing or left-wing - revealed no significant treatment effects (Table A15). This is unsurprising since the treatments were designed to evoke intergroup sentiments, inherently targeting partisan biases. Centrists lack strong alignment with either political group and would logically be less responsive to messages framed within an inter-group context.

Third, we conduct a sensitivity analysis for the causal mediation to address concerns related to sequential ignorability - a core assumption in causal mediation analysis that entails two main conditions: the as-if random assignment of treatment and the absence of unmeasured confounders influencing both the mediator and the outcome (Imai, Keele & Yamamoto 2010). While experimental designs inherently satisfy the first condition, we perform the sensitivity analysis to examine the robustness of our ACME values against potential unmeasured confounders that could affect both the mediator and the outcome. This involves computing the maximum correlation an unmeasured confounder might have with both mediator and outcome without invalidating our observed mediation effects. The results are presented in Table A4 of Appendix C. The findings demonstrate that the significant mediation effects observed for discounting and content perception remain robust unless there exists an unmeasured confounder with correlations exceeding -0.15 and -0.3, respectively. This suggests a moderate-to-strong resistance of our findings to potential unmeasured confounders.

8 Discussion

This research set out to test the influence of inter-group political humour on affective partisan polarisation. Taken together, the results suggest that humour can act as a depolarising influence in political communication, particularly when it targets the political in-group: Participants who encountered jokes about their own political group showed a significant reduction in affective polarisation compared to those in the control group. Notably, humour moderates the impact of in-group targeted messages on polarisation by reducing both out-group hostility and in-group favouritism. Humorous messages aimed at out-groups did not demonstrate a similar depolarising effect. A causal mediation analysis shows that humour reduces polarisation by changing how people perceive contentious messages: either making them less likely to take the content seriously (discounting) or improving their views toward the content.

Overall, the study makes two key contributions to the literature on political communication and affective polarisation. First, while previous research has extensively documented how consumption of partisan media exacerbates polarisation (e.g. Bakshy et al. 2015, King et al. 2017, Prior 2013), the relative influence of different mediums of expression within the media has remained less understood. This study is the first to consider inter-group disparaging humour as a unique medium of political expression that influences affective partisan polarisation. By focusing on humour - a medium growing in prominence and distinct from traditional news formats - we extend the current understanding of the media's role in shaping political attitudes.

Our second contribution engages with the contentious debate surrounding the effects of counter-attitudinal media on political polarisation. This literature consistently shows that homogeneous media environments increase polarisation (Levendusky 2013, Stroud 2010, Sunstein 1999), yet the effects of counter-attitudinal information is contested. While some contend that exposure to counter-attitudinal content exacerbates polarisation (Bail et al. 2018, Garrett 2009, Garrett et al. 2014), evidence from other studies suggests this reaction fails to materialise (Guess & Coppock 2020), with some noting a reduction in polarisation (Kim 2015). In response, scholars have called for a shift in focus to the *conditions* under which counter-attitudinal information exacerbates or mitigates polarisation (Kubin & Von Sikorski 2021). We find that humour is one such determining condition: counter-attitudinal messaging (i.e. in-group targeted) increase polarisation, but this effect reverses when the messages are delivered humorously.

The limitations of this study pertain mainly to its ecological and external validity. While the current design allowed for a controlled examination of causal relationships, its ecological validity is constrained due to the inherent differences with the real-world media environment in which political humour is typically encountered. Ideally, to enhance ecological validity, future research should aim to test the influence of humour within participants' natural media environments (Bail et al. 2018), such as manipulating exposure to political humour or memes organically within their own social media feeds. Similarly, while our study uses a representative British sample, extrapolating these findings to other contexts must be approached with caution. The consumption and affinity to humour is profoundly culture-specific (Jiang et al. 2019); thus, its influence might vary across different contexts. Despite these limitations, our study represents a significant step forward in terms of external validity, especially when compared to prior work that hinges on the context of specific television programs like 'The Colbert Report' or 'Who is America?'. By using jokes that span the prevalent stereotypes across political groups, our results are less tied to the peculiarities of a particular show or media personality and more reflective of the structural components of humour itself.

Ultimately, it is clear that the importance of research on the influence of diverse political media exposure will only continue to grow as our media space becomes increasingly fragmented and the partisan divide deepens. In this context, humour emerges as an accessible tool that can bridge these widening partisan gaps. Unlike traditional reconciliation approaches that demand sustained, positive interactions (Pettigrew & Tropp 2013), humour offers a singular, accessible moment that could alleviate inter-group tensions. By making light of our differences, humour could serve to subvert rather than reinforce them.

References

- Abrams, J. R. & Bippus, A. (2011), 'An intergroup investigation of disparaging humor', Journal of Language and Social Psychology 30(2), 193–201.
- Abrams, J. R. & Bippus, A. M. (2014), 'Gendering jokes: Intergroup bias in reactions to same-versus opposite-gender humor', *Journal of Language and Social Psychology* 33(6), 692–702.
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F. & Volfovsky, A. (2018), 'Exposure to opposing views on social media can increase political polarization', *Proceedings of the National Academy* of Sciences 115(37), 9216–9221.

- Bakshy, E., Messing, S. & Adamic, L. A. (2015), 'Exposure to ideologically diverse news and opinion on facebook', *Science* **348**(6239), 1130–1132.
- Baumgartner, J. C. & Morris, J. S. (2008), 'One "nation," under stephen? the effects of the colbert report on american youth', *Journal of Broadcasting & Electronic Media* 52(4), 622–643.
- Becker, A. B. (2012), 'Comedy types and political campaigns: The differential influence of other-directed hostile humor and self-ridicule on candidate evaluations', *Mass Communication and Society* 15(6), 791–812.
- Becker, A. B. (2014), 'Playing with politics: Online political parody, affinity for political humor, anxiety reduction, and implications for political efficacy', *Mass Communication and Society* 17(3), 424–445.
- Becker, A. B. (2021), 'When comedy goes to extremes: The influence of ideology and social identity on source liking, credibility, and counterarguing.', *Psychology of Popular Media* 10(1), 39.
- Bill, B. & Naus, P. (1992), 'The role of humor in the interpretation of sexist incidents', Sex Roles 27, 645–664.
- Broockman, D. E., Kalla, J. L. & Westwood, S. J. (2023), 'Does affective polarization undermine democratic norms or accountability? maybe not', American Journal of Political Science 67(3), 808–828.
- Eisend, M. (2009), 'A meta-analysis of humor in advertising', *Journal of the Academy* of Marketing Science **37**, 191–203.
- English, K., Sweetser, K. D. & Ancu, M. (2011), 'Youtube-ification of political talk: An examination of persuasion appeals in viral video', *American Behavioral Scientist* 55(6), 733–748.
- Ford, T. E., Breeden, C. J., O'Connor, E. C. & Banos, N. C. (2017), Jokes and humor in intergroup relations, *in* 'Oxford Research Encyclopedia of Communication'.
- Garrett, R. K. (2009), 'Echo chambers online?: Politically motivated selective exposure among internet news users', Journal of computer-mediated communication 14(2), 265–285.

- Garrett, R. K., Gvirsman, S. D., Johnson, B. K., Tsfati, Y., Neo, R. & Dal, A. (2014), 'Implications of pro-and counterattitudinal information exposure for affective polarization', *Human communication research* 40(3), 309–332.
- Goldthwaite Young, D. (2004), 'Late-night comedy in election 2000: Its influence on candidate trait ratings and the moderating effects of political knowledge and partisanship', *Journal of Broadcasting & Electronic Media* **48**(1), 1–22.
- Gruner, C. R. (2017), The game of humor: A comprehensive theory of why we laugh, Routledge.
- Guess, A. & Coppock, A. (2020), 'Does counter-attitudinal information cause backlash? results from three large survey experiments', *British Journal of Political Science* 50(4), 1497–1515.
- Highfield, T. & Leaver, T. (2016), 'Instagrammatics and digital methods: Studying visual social media, from selfies and gifs to memes and emoji', *Communication research* and practice 2(1), 47–62.
- Hmielowski, J. D., Holbert, R. L. & Lee, J. (2011), 'Predicting the consumption of political tv satire: Affinity for political humor, the daily show, and the colbert report', *Communication monographs* 78(1), 96–114.
- Imai, K., Keele, L. & Tingley, D. (2010), 'A general approach to causal mediation analysis.', *Psychological methods* 15(4), 309.
- Imai, K., Keele, L., Tingley, D. & Yamamoto, T. (2011), 'Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies', American Political Science Review 105(4), 765–789.
- Imai, K., Keele, L. & Yamamoto, T. (2010), 'Identification, inference and sensitivity analysis for causal mediation effects'.
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. (2019), 'The origins and consequences of affective polarization in the united states', *Annual review* of political science 22, 129–146.
- Jiang, T., Li, H. & Hou, Y. (2019), 'Cultural differences in humor perception, usage, and implications', Frontiers in psychology 10, 438919.

Johnson, A. M. (1990), 'The "only joking" defense: Attribution bias or impression management?', *Psychological Reports* **67**(3), 1051–1056.

Karnitschnig, M. (2019), 'Why europe can't stop laughing at boris johnson', POLITICO

URL: https://www.politico.eu/article/why-europe-cant-stop-laughing-at-borisjohnson/

- Kim, Y. (2015), 'Does disagreement mitigate polarization? how selective exposure and disagreement affect political polarization', *Journalism & Mass Communication Quarterly* 92(4), 915–937.
- King, G., Schneer, B. & White, A. (2017), 'How the news media activate public expression and influence national agendas', *Science* 358(6364), 776–780.
- Kingzette, J., Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M. & Ryan, J. B. (2021), 'How affective polarization undermines support for democratic norms', *Public Opinion Quarterly* 85(2), 663–677.
- Kubin, E. & Von Sikorski, C. (2021), 'The role of (social) media in political polarization: a systematic review', Annals of the International Communication Association 45(3), 188–206.
- Kuiper, N. A., McKenzie, S. D. & Belanger, K. A. (1995), 'Cognitive appraisals and individual differences in sense of humor: Motivational and affective implications', *Personality and individual differences* 19(3), 359–372.
- Levendusky, M. S. (2013), 'Why do partisan media polarize viewers?', American journal of political science 57(3), 611–623.
- McCoy, J., Rahman, T. & Somer, M. (2018), 'Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities', *American Behavioral Scientist* 62(1), 16–42.
- McLuhan, M. (2017), The medium is the message, *in* 'Communication theory', Routledge, pp. 390–402.
- Messing, S. & Westwood, S. J. (2014), 'Selective exposure in the age of social media: Endorsements trump partial source affiliation when selecting news online', *Communication research* 41(8), 1042–1063.

- Meyer, J. C. (2000), 'Humor as a double-edged sword: Four functions of humor in communication', *Communication theory* **10**(3), 310–331.
- Nabi, R. L., Moyer-Gusé, E. & Byrne, S. (2007), 'All joking aside: A serious investigation into the persuasive effect of funny social issue messages', *Communication Monographs* 74(1), 29–54.
- Pettigrew, T. F. & Tropp, L. R. (2013), Does intergroup contact reduce prejudice? recent meta-analytic findings, *in* 'Reducing prejudice and discrimination', Psychology Press, pp. 93–114.
- Prior, M. (2013), 'Media and political polarization', Annual review of political science 16, 101–127.
- Shahbaznezhad, H., Dolan, R. & Rashidirad, M. (2021), 'The role of social media content format and platform in users' engagement behavior', *Journal of Interactive Marketing* 53(1), 47–65.
- Stroud, N. J. (2010), 'Polarization and partial selective exposure', Journal of communication 60(3), 556–576.
- Sunstein, C. R. (1999), 'The law of group polarization', University of Chicago Law School, John M. Olin Law & Economics Working Paper (91).
- Tajfel, H. (1982), *Social identity and intergroup relations*, Vol. 7, Cambridge University Press.
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. (2014), 'Mediation: R package for causal mediation analysis'.
- Weinberger, M. G. & Gulas, C. S. (1992), 'The impact of humor in advertising: A review', *Journal of advertising* **21**(4), 35–59.
- Weiss, Joanna (2019), 'Trump pokes fun at himself: Why do only some people see it?', POLITICO.

URL: https://www.politico.eu/article/trump-pokes-fun-at-himself-why-do-onlysome-people-see-it/

Xenos, M. A., Moy, P. & Becker, A. (2011), 'Making sense of the daily show', The Stewart/Colbert effect: Essays on the real impacts of fake news pp. 47–62. Zillmann, D. & Cantor, J. R. (2017), A disposition theory of humour and mirth, *in* 'Humor and laughter', Routledge, pp. 93–115.

Online Appendix

Α	Balance and manipulation checks	26
в	Description of outcome variables	27
С	Mediation analysis	28
D	Heterogeneity analysis	30
\mathbf{E}	Questionnaire	34
F	Stimulus material	38
\mathbf{G}	Supplementary tables	42

A Balance and manipulation checks

	1	2	3	4	5	6
Education						
None	0.01	0.01	0.01	0.00	0.01	0.01
Secondary	0.06	0.06	0.06	0.06	0.05	0.05
Vocational	0.03	0.02	0.03	0.03	0.03	0.03
Higher education	0.08	0.07	0.06	0.08	0.08	0.07
Other	0.00	0.00	0.00	0.00	0.00	0.01
Gender						
Male	0.09	0.09	0.09	0.08	0.09	0.09
Female	0.08	0.06	0.08	0.09	0.08	0.08
Other	0.00	0.00	0.00	0.00	0.00	0.00
Region						
North West	0.02	0.02	0.02	0.02	0.02	0.02
North East	0.01	0.01	0.01	0.01	0.01	0.01
Yorkshire and the Humber	0.02	0.01	0.01	0.01	0.02	0.01
East of England	0.01	0.01	0.02	0.01	0.01	0.02
West Midlands	0.01	0.01	0.02	0.01	0.02	0.01
East Midlands	0.01	0.01	0.01	0.01	0.01	0.01
Wales	0.01	0.01	0.01	0.01	0.01	0.01
Scotland	0.02	0.01	0.02	0.02	0.02	0.01
South East	0.03	0.02	0.02	0.03	0.02	0.02
London	0.02	0.02	0.02	0.03	0.03	0.03
South West	0.01	0.02	0.01	0.02	0.01	0.02
Ideology						
Left-wing	0.08	0.07	0.08	0.07	0.07	0.07
Right-wing	0.09	0.09	0.09	0.11	0.10	0.10
Ethnicity						
White	0.14	0.12	0.14	0.14	0.15	0.13
Black or Black British	0.01	0.01	0.01	0.01	0.01	0.01
Asian or Asian British	0.01	0.02	0.01	0.02	0.01	0.02
Mixed race	0.00	0.00	0.00	0.00	0.00	0.00
Chinese	0.00	0.00	0.00	0.00	0.00	0.00
Other	0.00	0.00	0.00	0.00	0.00	0.00
Age	47.01	47.17	47.95	46.57	47.91	46.43
Political attention	7.74	7.79	7.72	7.70	7.76	7.86

Table A1: Balance checks across treatment conditions

	Dependent variable:		
	Perceived Humour	Joke Perception	
Humour Treatment	1.044***	0.569***	
	(0.153)	(0.057)	
Constant	4.766^{***}	2.820***	
	(0.108)	(0.040)	
Observations	1,472	1,480	
Adjusted \mathbb{R}^2	0.030	0.063	

Table A2: Manipulation checks

*p < 0.1; **p < 0.05; ***p < 0.01

Note: This table presents the estimated coefficients from linear regression models testing the influence of the humourous treatments (vs. non-humorous) on perceived humour and joke perception. Perceived humour is a 10-point scale that measures the extent to which participants found the content humorous. Joke perception measures whether participants believed the authors of the messages were joking on a 5-point agreement scale. Survey weights are applied. Standard errors are in parentheses.

B Description of outcome variables

Post-treatment, respondents are presented with a series of questions designed to measure their political attitudes and feelings towards different political groups. These questions are used to construct measures of in-group favouritism and out-group hostility. For all measures, affective polarisation is calculated as the absolute difference in positive ratings between the in-group and out-group.

- 1. Thermometer Ratings: Respondents rate their feelings towards Labour and Conservative identifiers on a scale from 0 (very cold) to 100 (very warm). For in-group favouritism, we use the thermometer scores directed towards the respondent's own political group. For out-group hostility, we use the inverted scores directed towards out-partisans.
- 2. Traits Ratings: Respondents rate how well various traits apply to Labour and Conservative supporters on a scale from 1 (not at all well) to 5 (extremely well). For in-group favouritism, we aggregate respondents' ratings on positive traits (e.g., Patriotic, Intelligent, Honest) and the inverted ratings of negative traits

(e.g., Hypocritical, Selfish, Mean) for their own group. These combined scores form a scale where higher values indicate more favourable views towards the ingroup. Out-group hostility is then measured by applying the same aggregate measure of out-partisan ratings, but inverted so that higher values represent less favourable views toward the out-group.

3. Social Distance: Respondents rate how comfortable they are having close personal friends, neighbours, or in-laws who are Labour or Conservative on a scale from 1 (Not at all comfortable) to 4 (Extremely comfortable). In-group favouritism is measured by the aggregate comfort level with in-partisans and out-group hostility is measured by discomfort levels with out-partisans.

C Mediation analysis

Mediator	Total Effect	Direct Effect	Indirect Effect	Percent Mediated	Sample Size
Discounting	-0.047**	-0.036*	-0.011***	22.571**	490
Negative emotions	-0.042**	-0.031	-0.011**	24.599*	493
Positive emotions	-0.045**	-0.04**	-0.005	10.559	496
Content perception	-0.045***	-0.029	-0.017***	36.966***	495
Content accuracy	-0.044**	-0.041**	-0.003	7.141	497
			*p < 0.1	; **p < 0.05; ***p < 0.05; **	0.01

Table A3: Mediation analysis summary

Note: This table summarises the results of mediation analyses examining the indirect effects of in-group humorous (vs. non-humorous) content on affective polarisation through various mediators. The traitsbased measure of affective polarisation is used here. The direct effect represents the effect of humour on affective polarisation, controlling for the mediator. The indirect effect is the effect of humour on the polarisation that operates through the mediator. The total effect represents the combined influence of both direct and indirect effects. Percent mediated indicates the percentage of the total effect mediated by each mediator. Survey weights are applied.

Mediator	ρ	$R_M^2 \times R_Y^2$	$R^2_{\tilde{M}} \times R^2_{\tilde{Y}}$	
Discounting	-0.15	0.02	0.02	
Content perception	-0.30	0.09	0.06	
		*p	< 0.1; **p <	0.05; ***p < 0.01

Table A4: Sensitivity analysis for mediation effects

Note: This table presents sensitivity analysis outcomes for the mediation effects in Table A3 using a ρ increment of 0.05. ρ quantifies potential bias from unobserved confounders. $R_M^2 \times R_Y^2$ shows the variance explained by observed variables, excluding unmeasured confounders. $R_{\tilde{M}}^2 \times R_{\tilde{Y}}^2$ indicates the variance explained when considering potential unmeasured confounders. Negative ρ values suggest an inverse relationship with unmeasured confounders.

D Heterogeneity analysis

	Dependent variable: Affective Polarisation (Traits)
In-group	0.11**
	(0.05)
Out-group	0.06
	(0.05)
Humorous	0.05
	(0.05)
In-group*Humorous	-0.09
	(0.07)
Out-group*Humorous	0.03
	(0.07)
Age	0.002***
	(0.001)
In-group*Age	-0.001
	(0.001)
Out-group*Age	-0.001
	(0.001)
Humorous*Age	-0.001
	(0.001)
In-group*Humorous*Age	0.0005
	(0.001)
Out-group*Humorous*Age	-0.001
	(0.001)
Constant	0.01
	(0.04)
Observations	1,488

Table A5: Heterogeneity in treatment effects by age

*p < 0.1; **p < 0.05; ***p < 0.01

Note: This table presents the estimated coefficients from linear regression models testing the influence of humour, target group, and their interactions with age on affective polarisation. The reference groups are 'non-humorous' and 'no target' respectively. Affective polarisation is measured using three different scales: thermometers, trait ratings, and social distance. The models control for political attention, left-right self-placement, and ethnicity. Survey weights are applied. Standard errors are in parentheses.

	Dependent variable: Affective Polarisation (Traits)
In-group	0.09***
	(0.03)
Out-group	0.05**
	(0.03)
Humorous	0.04^{*}
	(0.03)
In-group*Humorous	-0.11^{***}
	(0.04)
Out-group*Humorous	-0.03
	(0.04)
Gender	0.03
	(0.03)
In-group*Gender	-0.06^{*}
	(0.04)
Out-group*Gender	-0.07^{**}
	(0.04)
Humorous*Gender	-0.03
	(0.04)
In-group*Humorous*Gender	0.06
	(0.05)
Out-group*Humorous*Gender	0.05
	(0.05)
Constant	0.07***
	(0.02)
Observations	1,479

Table A6: Heterogeneity in treatment effects by gender

*p < 0.1; **p < 0.05; ***p < 0.01

Note: This table presents the estimated coefficients from linear regression models testing the influence of humour, target group, and their interactions with gender on affective polarisation. The reference groups are 'non-humorous' and 'no target' respectively. Gender is a binary variable which takes on the values (1) male and (0) female. Affective polarisation is measured using three different scales: thermometers, trait ratings, and social distance. The models control for political attention, left-right self-placement, and ethnicity. Survey weights are applied. Standard errors are in parentheses.

	Dependent variable: Affective Polarisation (Traits)
In-group	0.12**
	(0.05)
Out-group	0.03
	(0.05)
Humorous	0.08*
	(0.05)
In-group*Humorous	-0.18***
	(0.06)
Out-group*Humorous	-0.03
	(0.06)
Ethnicity	0.08**
0	(0.04)
In-group*Ethnicity	-0.06
	(0.05)
Out-group*Ethnicity	-0.02
	(0.05)
Humorous*Ethnicity	-0.07
	(0.05)
In-group*Humorous*Ethnicity	0.14*
	(0.07)
Out-group*Humorous*Ethnicity	0.03
s and a star framerous Dominorby	(0.07)
Constant	0.02
	(0.02)

Table A7: Heterogeneity in treatment effects by ethnicity

*p < 0.1; **p < 0.05; ***p < 0.01

Note: This table presents the estimated coefficients from linear regression models testing the influence of humour, target group, and their interactions with ethnicity on affective polarisation. The reference groups are 'non-humorous' and 'no target' respectively. Ethnicity is a binary variable which takes on the values (1) white and (0) non-white. Affective polarisation is measured using three different scales: thermometers, trait ratings, and social distance. The models control for political attention and left-right self-placement. Survey weights are applied. Standard errors are in parentheses.

	Dependent variable: Affective Polarisation (Traits)
In-group	0.09***
	(0.03)
Out-group	0.03
	(0.03)
Humorous	0.05^{*}
	(0.03)
In-group*Humorous	-0.12^{***}
	(0.04)
Out-group*Humorous	-0.002
	(0.04)
Ideology	-0.06^{**}
	(0.03)
In-group*Ideology	-0.06
	(0.04)
Out-group*Ideology	-0.03
	(0.04)
Humorous*Ideology	-0.04
	(0.04)
In-group*Humorous*Ideology	0.09^{*}
	(0.05)
Out-group*Humorous*Ideology	0.004
	(0.05)
Constant	0.08***
	(0.03)
Observations	1,488

Table A8: Heterogeneity in treatment effects by ideology

 $p^* < 0.1; p^* < 0.05; p^* < 0.01$

Note: This table presents the estimated coefficients from linear regression models testing the influence of humour, target group, and their interactions with ideology on affective polarisation. The reference groups are 'non-humorous' and 'no target' respectively. Ideology is a binary variable which takes on the values (1) right-wing and (0) left-wing. Affective polarisation is measured using three different scales: thermometers, trait ratings, and social distance. The models control for political attention and ethnicity. Survey weights are applied. Standard errors are in parentheses.

E Questionnaire

Demographic information

- 1. How old are you? (end survey if age < 18)
- 2. Which of the following best describes how you think of yourself?
 - Female
 - Male
 - I identify in another way
 - Prefer not to answer
- 3. Which one of the following ethnic groups do you consider yourself to belong to?
 - White
 - Black or Black British
 - Asian or Asian British
 - Mixed race
 - Prefer not to answer
 - Other
- 4. What is the highest educational level that you have achieved?
 - Secondary school
 - University degree or equivalent professional qualification (NVQ level 4, etc.)
 - Higher university degree (doctorate, MBA, NVQ level 5, etc.)
 - Still in full time education
 - No formal education
 - Don't know
 - Prefer not to answer
- 5. Region
 - North West

- North East
- Yorkshire and the Humber
- East of England
- West Midlands
- East Midlands
- Wales
- Scotland
- South East
- London
- South West
- Northern Ireland

Political ideology

- 6. In politics people sometimes talk of left and right. Where would you place yourself on the following scale where 0 is Left and 10 is Right?
- 7. If you had to pick one, would you say you lean more towards the left-wing or the right-wing?
 - Left wing
 - Right wing
 - Neither

Voting behaviour

- 8. Talking to people about the General Election on Dec 12th, 2019, we have found that a lot of people didn't manage to vote. How about you did you manage to vote in the General Election in December 2019? (Randomize)
 - No, I chose not to vote
 - No, I was prevented from voting
 - No, I was not old enough to vote

- No, I was not registered to vote
- Yes, I voted
- I don't remember
- 9. Which party did you vote for in the General Election on December 12th, 2019?
 - Conservative
 - Labour
 - Liberal Democrat
 - Scottish National Party
 - Plaid Cymru
 - Brexit Party
 - UK Independence Party
 - Green Party
 - Some other party
 - Don't remember
 - Did not vote
 - Prefer not to answer

Treatment

Attitudes about content

- 10. How funny did you find the content you just read on a scale of 0 (not funny at all) to 10 (extremely funny)?
- 11. Thinking of the content you just read, to what extent do you agree with the following statements on a scale of 1 (strongly disagree) to 5 (strongly agree):
 - The authors of the messages were just joking
 - The messages were intended more to entertain than to persuade
 - The authors were serious about advancing their views in the messages
 - It would be easy to dismiss these messages as simply a joke

- The content presents an accurate description of its subject matter
- My overall feeling toward the content is positive
- 12. Thinking of the content you just read, to what extent do you agree with the following statements on a scale of 1 (strongly disagree) to 5 (strongly agree): While reading the content I felt:
 - Happy
 - Hopeful
 - Angry
 - Sad
 - Frustrated
 - Anxious
 - Indifferent

Polarisation

- 11. Thermometers: We'd like you to rate how you feel towards people who identify as left-wing and right-wing on a scale of 0 to 100, which we call a "feeling thermometer". On this feeling thermometer scale, ratings between 0 to 49 degrees mean that you feel unfavourable and cold (with 0 being the most unfavourable). Ratings between 51 and 100 degrees mean that you feel favourable and warm (with 100 being the most favourable). A rating of 50 means you have no feelings one way or the other.
- 12. Social distance: How comfortable are you having [close personal friends] [neighbours] [son/daughter-in-law] who are on the right of politics [on the left of politics]? Indicate your response on a scale from (1) Not at all comfortable to (Extremely Comfortable).
- 13. **Trait ratings:** Now we'd like to know more about what you think about [Labour supporters] [Conservative supporters]. Below, we've given a list of words that some people might use to describe them. For each item, please indicate how well you think it applies to them on a scale from (1) Not at all well to (5) Extremely well:

- Patriotic
- Intelligent
- Honest
- Open-minded
- Generous
- Hypocritical
- Selfish
- Mean

F Stimulus material

1. Humorous; Target Labour/Left

In this section, you will be presented with some political jokes. You will be asked to reflect on your thoughts and feelings about the content you just read. (Randomise order of jokes)

- 1. Like most lefties, I will do anything for the working classes, anything apart from mix with them.
- 2. What's the difference between Rudolph the red-nosed reindeer and a leftie? Rudolph can overcome adversity without any government assistance.
- 3. Labour was winning the UK election on Thursday... until work finished at 5 and everyone voted Conservative.
- 4. Why do lefties work seven days a week? So you don't have to retrain them on Monday.
- 5. How do you turn a leftie into a Tory? Move them out of their parents' basement.
- 6. How many lefties does it take to change a light bulb? None, they're too busy changing their gender.
- 7. Lefties are very broadminded: they are always willing to give careful consideration to both sides of the same side.

8. If the British people won't vote Labour, the Labour party will import people who will!

2. Non-humorous; Target Labour/Left

In this section, you will be presented with some political statements. You will be asked to reflect on your thoughts and feelings about the content you just read. (Randomise order of jokes)

- 1. Lefties say they are dedicated to the working classes but do not actually mix with them.
- 2. Lefties cannot overcome adversity without government assistance.
- 3. Lefties can't get jobs.
- 4. Lefties are not efficient or skilled in their work.
- 5. Lefties still rely on their parents for money.
- 6. Lefties are too preoccupied with trivial things like gender identity.
- 7. Lefties are closed-minded.
- 8. The left relies on immigrants to get votes.

3. Humorous; Target Conservative/Right

In this section, you will be presented with some political jokes. You will be asked to reflect on your thoughts and feelings about the content you just read. (Randomise order of jokes)

- 1. Tories feel they deserve everything they've stolen.
- 2. What do you call a criminal Tory? Lord or Sir.
- 3. What do you call a Tory with an IQ of 75? Minister. What do you call a Tory with an IQ of 50? Prime Minister.
- 4. What sort of exercise does a Tory do? They run this country into the ground.

- 5. The only way you can ever accuse a Tory of hypocrisy is if they walk past a homeless person without kicking him in the face.
- 6. Tory MPs have provided excellent examples of how hard-working people just need to roll up their sleeves and be born into a billionaire's family.
- 7. Why do transplant patients prefer Tory hearts to others? Because they've been used so little!
- 8. The Conservative Party is the party that says government doesn't work and then gets elected and proves it.

4. Non-humorous; Target Conservative/Right

In this section, you will be presented with some political statements. You will be asked to reflect on your thoughts and feelings about the content you just read. (Randomise order of jokes)

- 1. Tories unjustly claim ownership of wealth that has not been rightfully earned.
- 2. Tories with titles like Lord or Sir are criminals.
- 3. Tory ministers are incompetent.
- 4. Tories are running this country into the ground.
- 5. Tories lack empathy and concern for those who are less fortunate.
- 6. Tories are disconnected from the struggles of hard-working people and underestimate the barriers to success faced by the average person.
- 7. Tories are heartless.
- 8. The Conservative party accomplishes nothing during its time in government.

5. Humorous; Non-political

In this section, you will be presented with some jokes. You will then be asked to reflect on your thoughts and feelings about the content you just read. (Randomise order of jokes)

- 1. What's the best thing about British weather? It keeps the tourist crowds away!
- 2. I asked my British friend how he handles the weather. He said "Quite simple, really. I just carry an umbrella, sunglasses, a snow shovel, and sunscreen at all times."
- 3. British weather is the perfect employer: it offers a minimum of four seasons of work every day, and the job description always changes without notice.
- 4. What do you call a day in the UK with clear skies and sunshine? A photo-shopped postcard!
- 5. Why don't British people need to travel? Because if you wait long enough, the British weather brings every part of the globe to your doorstep typically in the wrong order.
- 6. What do you call a month's worth of rain? England.
- 7. What do you call it when the sun comes out in Scotland? A solar eclipse!
- 8. UK weather is so passive aggressive! Oh you want summer? I suppose you expect sunshine too?!

5. Non-humorous; Non-political

Please read the excerpt below carefully. You will then be asked to reflect on your thoughts and feelings about the content you just read.

More than nine in 10 Brits have talked about the weather in the last six hours. But is this unusual – and if so, is it their culture or the climate that makes them so obsessed?



scar Wilde said conversation about it was the last refuge of the unimaginative, while Bill Bryson noted that its most striking characteristic is that there isn't much of it. The weather – and the British obsession with talking about it – has been puzzling outsiders for decades.

According to recent research, 94% of British respondents admit to having conversed about the weather in the past six hours, while 38% say they have in the past 60 minutes. "This means at almost any moment in this country, at least a third of the population is either talking about the weather, has already done so or is about to do so," says social anthropologist Kate Fox, who performed the studies in 2010 for an update of her book Watching the English.

G Supplementary tables

	Dependent variable: Affective Polarisation			
	Thermometer	Social Distance		
	(1)	(2)	(3)	
In-group	0.035	0.059***	0.021	
	(0.026)	(0.018)	(0.018)	
Out-group	0.008	0.014	0.015	
	(0.026)	(0.018)	(0.018)	
Humorous	0.024	0.026	0.005	
	(0.026)	(0.018)	(0.018)	
In-group*Humorous	-0.071^{*}	-0.070^{***}	-0.011	
	(0.036)	(0.025)	(0.025)	
Out-group*Humorous	-0.007	-0.001	0.018	
	(0.037)	(0.025)	(0.026)	
Political attention	0.022***	0.021***	0.006***	
	(0.003)	(0.002)	(0.002)	
Ideology	-0.116^{***}	-0.097^{***}	-0.067^{***}	
	(0.015)	(0.011)	(0.011)	
Black	-0.068^{**}	-0.040^{*}	-0.024	
	(0.034)	(0.024)	(0.024)	
Asian	-0.091^{***}	-0.050^{**}	-0.031	
	(0.028)	(0.020)	(0.020)	
Mixed race	0.025	-0.021	0.074^{*}	
	(0.056)	(0.039)	(0.039)	
Chinese	0.003	-0.102^{*}	-0.008	
	(0.082)	(0.057)	(0.058)	
Other Ethnicity	-0.012	-0.030	0.082	
	(0.074)	(0.051)	(0.052)	
Constant	0.228^{***}	0.095***	0.151^{***}	
	(0.030)	(0.021)	(0.021)	
Observations	1,488	1,488	1,488	
Adjusted \mathbb{R}^2	0.071	0.115	0.033	

Table A9: Effects of humour and target group on affective polarisation

Note:

*p<0.1; **p<0.05; ***p<0.01

	Dependent variable: Out-group Hostility			
	Thermometer	Social Distance		
	(1)	(2)	(3)	
In-group	0.059***	0.044***	0.042*	
	(0.022)	(0.015)	(0.021)	
Out-group	-0.021	0.017	0.009	
-	(0.021)	(0.015)	(0.021)	
Humorous	0.009	0.003	-0.013	
	(0.021)	(0.015)	(0.021)	
In-group*Humorous	-0.078**	-0.032	0.008	
	(0.030)	(0.022)	(0.030)	
Out-group*Humorous	0.054^{*}	0.019	0.058^{*}	
	(0.031)	(0.022)	(0.030)	
Political attention	0.009***	0.009***	-0.007***	
	(0.003)	(0.002)	(0.003)	
Ideology	-0.080***	-0.106***	-0.049***	
	(0.013)	(0.009)	(0.013)	
Black	-0.086***	-0.091***	0.004	
	(0.029)	(0.020)	(0.028)	
Asian	-0.065^{***}	-0.074***	-0.011	
	(0.024)	(0.017)	(0.023)	
Mixed race	0.036	0.003	0.030	
	(0.046)	(0.033)	(0.046)	
Chinese	-0.027	-0.048	0.021	
	(0.069)	(0.049)	(0.068)	
Other Ethnicity	0.028	-0.088**	-0.012	
·	(0.062)	(0.044)	(0.061)	
Constant	0.584***	0.557***	0.408***	
	(0.025)	(0.018)	(0.025)	
Observations	1,488	1,488	1,488	
Adjusted R ²	0.052	0.126	0.020	
Note:	*p<0.1; **p<0.05; ***p<0.01			

Table A10: Effects of humour and target group on out-group hostility

	Dependent variable: In-group Favouritism			
	Thermometer	Social Distance		
	(1)	(2)	(3)	
In-group	-0.033^{*}	0.016	-0.016	
	(0.020)	(0.015)	(0.020)	
Out-group	-0.031	-0.031^{**}	-0.010	
	(0.020)	(0.015)	(0.019)	
Humorous	-0.010	0.018	0.006	
	(0.020)	(0.015)	(0.020)	
In-group*Humorous	0.018	-0.049^{**}	-0.033	
	(0.028)	(0.021)	(0.028)	
Out-group*Humorous	0.035	0.017	-0.012	
	(0.028)	(0.022)	(0.028)	
Political attention	0.021***	0.013***	0.023***	
	(0.002)	(0.002)	(0.002)	
Ideology	-0.076^{***}	-0.069^{***}	-0.043^{***}	
	(0.012)	(0.009)	(0.012)	
Black	-0.054^{**}	0.016	-0.026	
	(0.026)	(0.020)	(0.026)	
Asian	-0.053^{**}	0.006	-0.026	
	(0.022)	(0.017)	(0.021)	
Mixed race	0.015	-0.017	-0.017	
	(0.043)	(0.033)	(0.042)	
Chinese	-0.037	-0.012	-0.106^{*}	
	(0.063)	(0.048)	(0.062)	
Other Ethnicity	-0.045	0.047	0.011	
	(0.057)	(0.044)	(0.056)	
Constant	0.550***	0.507***	0.647***	
	(0.023)	(0.018)	(0.023)	
Observations	1,488	1,488	1,488	
Adjusted R ²	0.079	0.068	0.073	
Note:	*p<0.1; **p<0.05; ***p<0.01			

Table A11:	Effects	of	humour	and	target	group	on	in-group	favouri	tism

	Dependent variable: Affective Polarisation (Traits)			
	No Controls	Demographic Controls	All Controls	
	(1)	(2)	(3)	
In-group	0.053***	0.057***	0.062***	
	(0.019)	(0.019)	(0.018)	
Out-group	0.009	0.012	0.017	
	(0.019)	(0.019)	(0.018)	
Humorous	0.026	0.028	0.029	
	(0.019)	(0.019)	(0.018)	
In-group*Humorous	-0.061^{**}	-0.063**	-0.072^{***}	
	(0.027)	(0.026)	(0.025)	
Out-group*Humorous	0.004	-0.002	-0.007	
	(0.027)	(0.026)	(0.025)	
Constant	0.195***	0.118***	0.090***	
Observations	1,493	1,483	$1,\!479$	
Adjusted \mathbb{R}^2	0.004	0.036	0.126	

Table A12: Treatment effects under different control conditions

*p < 0.1; **p < 0.05; ***p < 0.01

Note: This table presents the estimated coefficients from linear regression models testing the influence of humour and its target group on affective polarisation. The reference groups are 'non-humorous' and 'no target' respectively. Model 1 includes no control variables. Model 2 controls for gender, age, education, and region, while Model 3 includes all control variables from Model 2 along with ethnicity, ideology, and political attention (the variables used in the main analysis). The traits-based measure of affective polarisation is used in this analysis. Survey weights are applied. Standard errors are reported in parentheses.

	Dependent variable: Affective Polarisation			
	Thermometer	Traits	Social Distance	
Targeted	0.022	0.036**	0.018	
	(0.022)	(0.016)	(0.016)	
Humorous	0.024	0.026	0.005	
	(0.026)	(0.018)	(0.018)	
Targeted*Humorous	-0.039	-0.036	0.003	
	(0.032)	(0.022)	(0.022)	
Constant	0.228^{***}	0.095^{***}	0.150^{***}	
	(0.030)	(0.021)	(0.021)	
Observations	1,488	1,488	1,488	
Adjusted \mathbb{R}^2	0.071	0.112	0.034	

Table A13: Effects of humour and target group (no target vs. target) on polarisation

p < 0.1; p < 0.05; p < 0.05; p < 0.01

Note: This table presents the estimated coefficients from linear regression models testing the influence of humour and its target on affective polarisation. The reference groups are 'non-humorous' and 'no target' respectively. Affective polarisation is measured using three different scales: thermometers, trait ratings, and social distance. The models control for political attention, left-right self-placement, and ethnicity. Survey weights are applied. Standard errors are in parentheses.

	Dependent variable: Affective Polarisation			
	Thermometer	Traits	Social Distance	
In-group	0.027	0.044**	0.006	
	(0.026)	(0.018)	(0.018)	
Humorous	0.019	0.025	0.023	
	(0.026)	(0.018)	(0.019)	
In-group*Humorous	-0.067^{*}	-0.069^{***}	-0.029	
	(0.036)	(0.026)	(0.026)	
Constant	0.234^{***}	0.121^{***}	0.171^{***}	
	(0.035)	(0.025)	(0.025)	
Observations	994	994	994	
Adjusted \mathbb{R}^2	0.080	0.117	0.021	

Table A14: Effects of humour and target group (in-group vs. out-group) on polarisation

-

p < 0.1; p < 0.05; p < 0.05; p < 0.01

Note: This table presents the estimated coefficients from linear regression models testing the influence of humour and its target on affective polarisation. The reference groups are 'non-humorous' and 'out-group target' respectively. Affective polarisation is measured using three different scales: thermometers, trait ratings, and social distance. The models control for political attention, left-right self-placement, and ethnicity. Survey weights are applied. Standard errors are in parentheses.

	Dependent variable:			
	Thermometer	Traits	Social Distance	
	(1)	(2)	(3)	
Target				
Labour	0.064^{*}	0.012	0.022	
	(0.039)	(0.038)	(0.027)	
Conservative	0.023	-0.008	-0.010	
	(0.039)	(0.038)	(0.027)	
Tone				
Humorous	0.027	0.014	-0.040	
	(0.040)	(0.039)	(0.028)	
Target*Tone				
Labour*Humorous	-0.083	-0.022	0.019	
	(0.056)	(0.054)	(0.039)	
Conservative*Humorous	-0.006	0.017	0.058	
	(0.055)	(0.053)	(0.038)	
Constant	0.008	0.094**	0.042	
	(0.042)	(0.041)	(0.029)	
Observations	303	303	303	
Adjusted \mathbb{R}^2	0.008	0.013	-0.018	

Table A15: Treatment effects among centrists

*p < 0.1; **p < 0.05; ***p < 0.01

Note: This table presents the estimated coefficients from linear regression models testing the influence of humour and its target on affective polarisation among centrists. The reference groups are 'non-humorous' and 'no target' respectively. Affective polarisation is measured using three different scales: thermometers, trait ratings, and social distance. The models control for political attention, 2019 vote choice, and ethnicity. Survey weights are applied. Standard errors are in parentheses.