Political humor effects and partisan humor bias: an experimental study in the United States

By Áron József Szászi

Submitted to

Central European University

Department of Political Science

In partial fulfillment of the requirements for the degree of

Master of Arts in Political Science

Supervisors: Gábor Tóka and Gábor Simonovits

Budapest, Hungary

2020

Abstract

This thesis builds an original causal theory for political humor. First, it investigates the effects of hostile political joke exposure on political attitudes. It is theorized that symmetrical hostile joke exposure increases affective polarization and decreases trust in politicians. Second, the thesis develops a theory of partisan humor bias. It is hypothesized that people evaluate better pro-attitudinal jokes (jokes from co-partisan sources and hostile jokes about out-partisans). Furthermore, it is expected that affective polarization positively moderates joke source effect. It is also hypothesized that jokes with female joke sources are evaluated worse (benchmark for partisan source effect). To test these theories, survey experiments were conducted in a pilot study (N=149) and a second survey (N=796). Hypothesis are tested as part of the experiments of the second survey. The questionnaires were distributed on the survey platform Lucid.io, and respondents received financial compensation. Pilot survey shows that respondents found even soft critical and positive jokes moderately hostile (except a positive meme about Joe Biden). Joke target effects are significant, and partial heterogenous gender effect is revealed in the pilot. Republican males are negatively biased, while Democrat males are positively biased toward female joke sources. Joke target effect is significant in the second survey too, but gender effect does not replicate. Other hypotheses are rejected. The most important finding of this thesis is hostile humor exposure's effect on affective polarization. It is significant and substantial, but in opposite direction than expected. Hostile humor exposure also increases trust in politicians, but it does not pass the pre-set inference criteria.

Acknowledgements

So many people helped me some way to accomplish my thesis that it is impossible to list here all the names. First and foremost, I would like to share my deepest gratitude to my supervisors, Professor Gábor Tóka and Professor Gábor Simonovits. They guided me with both practical advices about my research design and abstract insights on my theoretical contribution.

I am thankful for my Professors and classmates at CEU. Oana Lup and her Internet and Politics course gave me the idea for my thesis topic. I learned experimental methods at the ECPR Summer University course of Federico Vegetti, which was essential for my thesis research. Federico also convinced me that my topic is worth to research, which gave me enormous motivation. I would like to thank Levente Littvay for making me understand statistics for the first time in my life. Levi also gave useful feedback to my thesis draft during the Political Psychology CEU Summer School. I am very thankful for the generous support of the CEU Research Grants, which helped me to conduct my survey experiments. I could not accomplish this research without the help of our Department Coordinator, Éva Lafferthon. Thank you!

I am grateful for the community and professors of Rajk College, who inspired me and showed me what passion for socially beneficial knowledge is. I would like to thank Pálma Polyák, for her mentoring, love for social sciences and positivist brainwashing which was essential to get ready for the challenges of CEU. Special thanks to Balázs Muraközy who helped me gain highly useful methodological skills.

I could not accomplish this thesis without the incredible support of my family. Their encouragement, care and curiosity for my studies meant a lot to me.

Finally, I am deeply grateful to Lívia who was a true companion besides me on this intellectual journey. Thank you for your kindness and challenging thoughts!

List of tables

Table 1: OLS regressions, dependent variable: perceived offensiveness of political jokes 39

Table 2: OLS regressions, dependent variable: joke appreciation index of political jokes 41

Table 3: Clustered OLS regressions, dependent variable: pilot joke evaluations 43

Table 4: OLS regressions, Hostile political joke exposure's effect on affective polarization (baseline: non-political joke exposure) 46

Table 5: OLS regressions, Hostile political joke exposure's effect on feelings toward co-partisans and out-partisans (baseline: non-political joke exposure) 48

Table 6: OLS regressions, Hostile political joke exposure's effect on trust variables (baseline: non-political joke exposure) 49

Table 7: OLS regressions, Hostile political joke exposure's effect on political activity (baseline: non-political joke exposure) 50

Table 8: Chi-squared tests, Hostile political joke exposure's effect on Apology variables 51

Table 9: Clustered OLS regressions, joke target congruence's effect on joke evaluation (baseline: co-partisan joke target) 53

Table 10: Clustered OLS regressions, Joke source partisanship's congruence and joke source's gender effects on perceived funniness (baseline partisanship: Democrat) 56

Table 11: Clustered OLS regressions, Joke source partisanship's congruence and joke source gender effects on perceived funniness and moderation by affective polarization 59

Table 12: Clustered OLS regressions, Hostile joke exposure's moderation effect on partisan and gender source effect 62

Table 13: 2SLS regression, Hostile joke exposure's indirect effect on funniness (mediated by affective polarization) and indirect moderation of partisan and gender source effect (by instrumented affective polarization) 63

Table 14: Summary of hypothesis tests (second survey) 66

Tables in the Appendix

Table A1: OLS regressions, partisan and gender joke source effects on the appreciation index of non-political jokes, by gender subgroups (pilot study, baseline: non-political male joke source) ..82

Table A2: OLS regressions, Partisan and gender joke source effects on perceived funniness of political jokes, by gender subgroups (pilot study, baseline: non-political male joke source) 83

Table A3: OLS regressions, hostile joke exposure's effects, mediated by strength of partisanship (baselines: non-political joke exposure, strong partisanship, strong Democrat).....84

Table A4: OLS regressions, Hostile political joke exposure's effect on trust variables, partisanship interaction included (baseline: non-political joke exposure) 85

Table A5: Chi-squared tests on subsamples, by partisanship, Hostile political joke exposure's effect on Trump apology variable.....86

Table A6: Chi-squared tests on subsamples, by partisanship, Hostile political joke exposure's effect on Biden apology variable.....86

Table A7: Clustered OLS regressions, joke target congruence's effect on political jokes' evaluation, mediated by strength of partisanship (Baseline: Co-partisan joke target).....87

Table A8: Clustered OLS regressions, joke source congruence's effect on non-political jokes' evaluation (Baseline: co-partisan joke target) 88

Table A9: Clustered OLS regressions, partisan and gender joke source effects on non-political joke evaluation, by gender subsamples.....89

Table A10: Two-sample t-tests with equal variances, difference between the average funniness of the jokes in the baseline and the treatment group of humor effect experiment.....90

Table A11: OLS regressions, the effect of perceived funniness of jokes on polarization.....91

Table A12: OLS regressions, the effect of perceived funniness of jokes about co-partisans and out-partisans on subcomponents of polarization 91

List of figures

Figure 1: Theorized mechanism of the polarizing joke effect 15

Figure 2: Theorized mechanism of the trust eroding joke effect 18

Figure 3: Theorized mechanism of ripple effects 20

Figure 4: Structure and research design of the second survey 28

Figure 5: Source cues' effects on perceived funniness, by partisanship and gender (adjusted predictions, 95 percent confidence interval) 37

Figure 6: Perceived hostility of the jokes, by respondent's partisanship 38

Figure 7: Appreciation indices of the jokes, by respondent's partisanship Note: This is an index variable, which is mean of the answers to the questions "Are you comfortable with this joke?" and "Is it a good depiction of reality?", both measured on a 100 point scale (0 'Not at all', 50 'Somewhat', 100 'Extremely') 40

Figure 8: Predicted margins of the political jokes' appreciation from a clustered regression, by partisan source congruence 44

Figure 9: Joke target congruence effect on joke appreciation index, by partisanship (adjusted predictions with 95 percent confidence interval) 53

Figure 10: Joke target congruence effect on joke appreciation index, by strength of partisanship (adjusted predictions with 95 percent confidence interval) 54

Figure 11: Joke source congruence effect on joke appreciation index, by partisanship (adjusted predictions with 95 percent confidence interval) 55

Figure 12: Joke source effect on perceived funniness, (recalibrated joke source variables) Note: adjusted predictions with 95 percent confidence interval) 57

Figure 13: Hostile joke exposure's moderation effect on joke source effect, by partisanship Note: adjusted predictions with 95 percent confidence interval 61

Figure 14: Perceived funniness of partisan jokes, by partisanship, strength of partisanship 68

Figures in the Appendix

Figure A1: Research design of the second part of the pilot survey 78

Figure A2: Vignettes of the hostile political joke exposure (2nd survey, humor effect experiment, treatment group) 79

Figure A3: Vignettes of the non-political joke exposure, with different partisan joke source cues (2nd survey, humor effect experiment, baseline group) 80

Figure A4: Vignettes of the non-political joke exposure, with different partisan and gender joke source cues (partisan and gender humor bias experiment) 81

Table of contents

Introduction 1

Chapter 1: Theoretical background 6

1.1 General theories of humor 6

1.2 The light side of political comedy 9

Chapter 2: A new framework for political humor 12

2.1 The dark side of political comedy 12

2.1.1 Polarizing effects of hostile political humor 12

2.1.2 Trust eroding effects of hostile political humor 16

2.1.3 (De)mobilization, as ripple effect of trust erosion and polarization 18

2.2 Political humor bias 20

Chapter 3: Research design 25

3.1 Pilot study 26

3.2 Second survey 27

3.2.1 Humor effects experiment 29

3.2.2 Partisan humor bias experiment 29

3.2.3 Partisan and gender humor bias experiment 31

3.5 Variables and measurement 31

3.5.1 General variables 31

3.5.2 Manipulated variables 32

3.5.3 Measured variables 33

Chapter 4: Analysis 36

4.1 Pilot study 36

4.2 Humor effects experiment 45

4.2.1 Effects on affective polarization 45

4.2.2 Effects on trust variables 48

4.2.3 Further exploratory analyses 50

4.3 Partisan humor bias experiment 52

4.3.2 Joke target effects 52

4.3.2 Joke source effects 54

4.4 Partisan and gender humor bias experiment 55

4.4.1 Joke source effects 55

4.4.2 Joke source moderation 58

4.4.3 Further exploratory analyses 60

Chapter 5: Discussion of the results 65

Conclusion.....71
Bibliography.....74
Appendix.....78

Introduction

Political humor is a historic phenomenon, ancient Greeks and Romans used jokes to discredit their competitors (Benton 1988). President Raegan collected Soviet jokes and told these jokes to his audiences (Roberts and Times 1987). However, some recent political events suggest that comical characters and humorous messages may have become more widespread in politics than before. Pirate and joke parties emerged all over Europe and some of them got elected into local governments and national parliaments. The Iceland joke party's mayor candidate in Reykjavik, the country's capital, even won in 2010 (McGrane 2010). Volodymyr Zelensky, a professional comedian who impersonated a teacher accidentally becoming president in a television series before, was elected as Ukraine's president in 2019 (Ray 2020). Boris Johnson, a stunningly comical figure became London's mayor, campaigned for leaving the EU, and later was chosen by the Conservative Party to be the UK's prime minister in the midst of the Brexit crisis (Encyclopaedia Britannica 2020).

Political joking became prevalent recently on both political sides in the United States as well. Satirical political television shows are mostly liberal and political psychology researches claim that its ultimate cause is that conservatives have different traits and aesthetic preferences and lower sense of humor (Young 2020; Young et al. 2019). In contrast, the right wing dominates the online meme culture in the US. Data-science driven research showed that an alt-right related site generated the most memes, while a Trump-supporter site was the most effective distributor of memes in the meme ecosystem (Zannettou et al. 2018). Some suggest that heavy usage of memes contributed largely to the 2016 electoral victory of Donald Trump. For instance, an opinion piece on Mother Jones stated that conservative groups had been organizing "meme trainings" for young right wingers for years, while the liberals are not willing to engage in this online fight (Mencimer and Helenowski 2019). The title of the article was "The Left Can't Meme", a phrase used as an insult toward liberals suggesting that they have no sense of humor (Urban Dictionary 2018). Although

the right wing may lead in the meme war, the iconic Biden-Obama (see Williams 2016) and Bernie Sanders memes (Tenbarge 2020) show that the left also "learns to meme".

The growing body of literature about political humor signals the importance of the topic (for a comprehensive review see Young 2018), but most studies focus on the US and the effects of satirical political television shows, such as the Daily Show or the Colbert Report. Many of these researches emphasize entertaining political material consumption's positive effects, primarily on political activity through the increased level of political knowledge. This thesis enrich the short list of political humor experiments which tested humor exposure's direct effects on political attitudes. My research differs in two important features from previous studies. First, it analyzes the effects of political humor in meme format, while previous studies used scenes of political satires as experimental treatments. Second, I use symmetrical joke exposure. Most studies used one-sided treatments, which were either pro- or counter-attitudinal jokes (an exception is Baumgartner and Morris 2006). My experiment of humor effects was based on a hostile joke exposure consisting memes about both Democrats and Republicans.

My thesis aims to build a complex theory of political humor effects and politically motivated humor comprehension. These focal points of my research is summarized in two research questions. The first research question regards the potential effects of political humor exposure.

RQ1: Does political humor exposure influence political attitudes directly and in which directions?

Political humor effects, understood as special types of media effects, potentially influence several political attitudes. The experiment aiming to answer this research question is limited to effects of hostile political humor exposure. It is tested how these jokes affect feelings toward co-partisans and out-partisans, trust in politicians and institutions, and political activity. Most importantly, two effects are hypothesized. It is expected that symmetrical exposure to political mockeries erodes trust in politicians. It is a reasonable hypothesis, since memes and satires are often humorous forms of political criticism. Experimental studies in US context showed that exposure to hostile parodies

of Democrat and Republican politicians reduced their perceived competence, even amongst participants who were on the same side as the joke target (Becker 2012, Becker and Haller 2014). I expect that these effects are additive and symmetrical exposure would result in general distrust.

A potentially more consequential hypothesized effect of hostile jokes can be the polarization of their audience. While political humor became an integral part of entertaining news media and modern campaign strategies, the United States experienced a dramatic increase of partisan divide. Strong partisanship and mutual distrust and social distancing between supporters of opposing political camps, known as affective polarization increased by around 20 percentage points between the 1970s and 2010s (Iyengar et al. 2019: p132). Are political mockeries, such as sarcastic political comedies or online trolling, partially responsible for this negative trend? The answer is not straightforward. To the best of my knowledge, only one study analyzed political jokes polarizing effects and it showed that subgroup of Republicans were polarized by liberal satire (Skoroda 2018).

However, the widescale spread of political humor, especially hostile joking could be the source as well as the consequence of affective polarization. Humorous political hostility might increase polarization, but polarization can be responsible for the politization of humor. Therefore, the second pillar of my thesis analyzes how polarization influences humor comprehension.

RQ2: Can humor appreciation cross partisan/ideological lines?

I theorized that an overspill of affective polarization is partisan humor bias. I argue that in a highly polarized country, as the United States, people are expected to appreciate more pro-attitudinal jokes – jokes which are told by co-partisans and/or make fun of the out-party. Counter-attitudinal jokes – out-partisans' jokes and/or jokes ridiculing co-partisans – are theorized to be less appreciated. Accordingly, it was theorized that the partisanship of the joke target as well as the joke source will have an effect on joke evaluation. Furthermore, it was theorized that individual level of affective polarization will moderate the effect of the joke source's partisanship. It was also hypothesized that

besides partisanship, the gender of the joke source also matters. This expectation was analyzed primarily in order to have a benchmark for potential partisan joke source effect.

Even preliminary answers to these research questions might be great contributions to the literature of political communication, voting behavior and political psychology. To test my theories, I carried out experiments embedded in a pilot survey taken by 169 US respondents and a survey of 796 US respondents. For the surveys, both political and non-political memes were collected from comical websites and these memes were embedded in artificially created Tweet-like vignettes. The pilot tested the effects of joke source cues and whether the selected memes are hostile or friendly enough to constitute a friendly and a hostile joke exposure in the second survey. Only hostile joke exposure treatment was incorporated in the second survey, as respondents evaluated soft and positive jokes hostile too. The pilot showed that respondents were surprisingly emphatic toward the out-party and evaluated even the soft jokes about the opposite side hostile.

The second survey tested the hypotheses of this paper. Hence, the survey contained three experiments: (1) humor effect, (2) partisan humor bias and (3) partisan and gender humor bias experiments. The analysis of the humor effect experiment showed the most surprising results. Significant effects were revealed, but opposite to the expectations. The symmetrical hostile media exposure treatment aimed to simulate a politically intense social media environment and increase polarization. It turned out that the treatment in fact decreased the individual level of affective polarization. This effect was driven by the increased sympathy towards the out-party. Some limited evidence showed that this experimental manipulation also increased trust in politicians, although this effect was only significant at 10 percent not surpassing the preset inference criteria.

The humor bias experiments showed null results, which can be also welcomed as good news. It suggests that the high level of affective polarization does not lead to a joke source-driven partisan humor bias. In other words, non-political jokes are evaluated based on their actual funniness and regardless the partisanship of the joke source. Regarding political jokes, on the other hand, the joke

target matters. Not surprisingly, respondents evaluated those jokes significantly better which made fun of the opposite political side.

My thesis is structured as follows. Chapter 1 introduces general theories of humor and the positive effect of political humor consumption. Chapter 2 presents my own theory of political humor effects and partisan humor bias. These theoretical models are partially based on political communication and political psychology studies. My hypothesis as well as planned exploratory analyses are based on these causal theories. Chapter 3 presents the research design, including the structure of my surveys and experiments, and operationalization of my variables. Chapter 3 describes the statistical analysis of the experimental data and Chapter 5 discusses the results. Finally, the thesis ends with concluding thoughts.

Chapter 1: Theoretical background

This chapter introduces the wide theoretical context of political humor. First, general theories of humor are briefly reviewed, presenting psychological and sociological perspectives. The mechanism of humor comprehension, the origins and social functions of humor and comical genres are introduced. Then, positive effects of political humor consumption are reviewed and discussed.

1.1 General theories of humor

Jokes and humor in general are fascinating social phenomena, which have been a fruitful research topic for many disciplines. In the early days of humor research, primarily psychologists and sociologists tried to understand how humor works and what purpose it serves for individuals and human communities.

Modern psychology aimed to open up the mechanism of humor comprehension to understand its origin and role. Martin (2007b) enhances that both three elements of humor, "cognitive (nonserious incongruity), emotional (mirth), and expressive (laughter)" (114) are deeply embedded in social context. Jokes usually describe people and their interactions, the feeling of humor made mirth is experienced with others and laughter is a communicative act. However, it is not clear whether political humor is inclusive or exclusive by nature regarding identity and group coherence building. It seems a rational strategy (following the spatial model) to remain open to a wide range of voters. On the other hand, Campbell's (1992) theory of "othering" suggest that identity creation is often built on the construction of a group of different people. In contrast, Koestler's sociological approach considers humor as a communication tool, which helps avoiding and resolving conflicts (Koestler 1964 in Martin 2007b: p 114-116). His evolutionary theory argues that humorous style, in contrast to more serious, goal-orientated way of communication, leads to the celebration of the inherent differences between people and it makes social dynamics smooth.

For instance, observational and experimental studies show that telling or laughing at "dirty jokes" often reveal attraction to the other person, or even a sexual offer (Walle 1976, Davis and Farina 1970 in Martin 2007b: p 118). Walle's study argues that humorous style makes the situation less serious and may protect both parties from harms of perceived harassment and rejection. Making fun of other groups, telling for instance racist or sexist jokes can be used to measure the tolerance level of others. Humor can be a successful mean of conflict management. For instance, a study of floor hockey players showed that humor was a common tool to handle verbal and physical clashes during the game and especially older people used humor, possibly because friendly relationships are more important for them (Palmer 1993 in Martin 2007b: p 119). These examples point to humor's role in deescalating conflicts, and friendly teasing can strengthen human relations. However, Martin's review (2007b) also emphasizes that humor can be used to hide hostile and inappropriate behavior.

It is also worth to understand the emotional explanations of humor's mechanism. Humor and laughter are genetically coded in humans and some ape races and they serve as a communication tool of social play (see the literature review of Martin 2007a). These cheerful physical interactions are considered to get evolved to humor with the development of language, which remained to play a role in social play, but humor also became general social tool. Creation, comprehension and response to humor involves cognitive and emotional processes. Koestler called the cognitive process when one understands two different meaning of the same thing bisociation, which is a common source of humor, usually appearing in puns (Koestler 1964 in Martin 2007a: p 7). Apter calls synergy the cognitive process when the same thing recalls two contradicting pictures, which causes a pleasurable state of mind (Apter 1982 in Martin 2007a: p 7). Humor is associated with mirth and neurological studies indeed prove that humor and laughter cause positive emotions. Not only one's own laughter, but hearing others laughing causes good feelings.

It often depends on social position who is the joke-teller, who is the target of the joke and what is considered acceptable. A statistical research of interpersonal conversations revealed that those people who dominated discussion also gave relatively more humorous remarks (Robinson and Smith-Lowin 2001 in Martin 2007b: p 120-121). A sociological study of psychologists showed that senior staff joked much more, and their jokes targeted younger colleges who, on the other hand, made fun of themselves or other people, not their colleagues (Coser 1960 in Martin 2007b: p 120). The observational study of Spradley and Mann (1975 in Martin 2007b: p 120) showed that bartenders insulted and made fun of waitresses as part of a domination strategy. While waitresses were expected not to take the jokes seriously, bartenders could warn waitresses about the inappropriateness of their jokes.

However, not only it might be misleading to present on overarching theory of humor. Martin (2007b) also points out that "humor is such a broad topic, the greatest empirical advances will likely be achieved by developing more narrowly focused theoretical models concerning specific humor components and processes" (p 151). Therefore, it is important to know what type of humor one aims to investigate. Different type of jokes does not necessarily lead to the same outcome.

Ruch et al. differentiates 8 categories of comical styles: fun, benevolent humor, nonsense, wit, irony, satire, sarcasm and cynicism, originally coming from Schmitt-Hidding's categorization (Schmitt-Hidding 1963 in Ruch et al. 2018). Four of these styles are considered "dark" by Ruch and his colleagues because they contain some form of ridicule or mockery. Sarcasm outstands regarding hurting others and it aims to point out how crooked the world is. Its performer tends to pose himself as superior to the audience. Cynicism differs in its subject, as it aims to ridicule certain socially accepted norms and values. Irony is a performative style which gives additional meaning to the literal message. It creates a community of the knowledgeable audience who feel superior towards those who do not recognize the joke. Satire is another aggressive type of humor, but it is an exception amongst "dark humor styles".

Ruch and Heintz labels satire or corrective humor, alongside with benevolent humor, virtue-related comical styles (Ruch and Heintz 2016 in Ruch et al. 2018). These two types of humor are highly normative and aim to raise awareness to everyday human and institutional mistakes. Benevolent humor is more understanding, while satire "mocks" people to push them toward the right direction. However, both share a hidden moral quest, instead of seeking goalless joy by harming others. The cross-national psychometric study of Heintz et al. (2018), which primarily tested BenCor, a recently founded measure system for virtue-related humor, showed that engagement in benevolent humor is increasing with age, while it is the opposite in the case of corrective humor.

This sophisticated categorization could be used for studying political humor as well, but some of the literature use much simpler classification, which is advisable to follow for the sake of simplicity of research. In this experimental research, hostile (hard and soft) and positive jokes will be distinguished based on the categorization of Ruch et al. (2018). I spend special attention to the comical categories introduced above, since different styles may activate different branches of the theorized humor effect's causal mechanisms.

1.2 The light side of political comedy

Although many positive social aspects of humor are known, the growing success of joke parties, and comical figures can be associated some kind of social and political paralysis. As Will Rogers put it, "the problem with practical jokes is that they very often get elected" (Rogers n.d. in Galles 2018). One may argue that "the elections of clowns" are sings of public disenchantment with politics, the erosion of substantive public discourse, or counterselection of politicians. These argumentations reflect a cynical view based on the underlying assumption that the general level of humor in politics correlates with low political quality. The next chapter elaborates how humor may play a role in growing partisan divide, lowering trust in politicians and public disengagement from politics. However, there is a different interpretation of the success of comical genres in political communication. It would simply suggest that humor has a serious mobilizing power. In other

words, not those politicians succeed whom you could only laugh at rather the ones who are good at making you laugh.

Social psychology lists "a wide variety of social uses of humor, including self-disclosure and social probing, decommitment and conflict de-escalation, enforcing social norms and exerting social control, establishing and maintaining status, enhancing group cohesion and identity, discourse management, and social play" (Martin 2007b: 155). It seems a reasonable assumption that such uses of humor appears not only in everyday life, but in political mass communication as well. Martin calls most situations, which involve laughing alone, such as TV-watching or reading "pseudo-social situations" (113). Therefore, political humor in the media might be understood as efforts to use fun, not only to discredit political competitors, but also unite the party, reinforce support, or reduce internal tensions.

The scholarly literature highlights the several positive effects of political humor consumption. Plevriti (2014) argues that the consumption as well as the creation of political memes are effective tools of political criticism, hence it can increase civic engagement. She based her conclusions on qualitative research. Baum's (2002) "gateway hypothesis" states that soft news increases political knowledge (primarily about foreign policy) and therefore these entertaining genres can lead to higher level of political engagement. First, Baumgartner and Morris (2006) applied this theory on political satires. Their laboratory experiment showed that exposure to political satire scenes increased the level of political efficacy. It was named "The Daily Show effect". (They also revealed negative effects of this treatment, which I introduce in the next chapter.) Xenos and Becker (2009) gave further empirical support for the "The Daily Show effect" with an experiment carried out on a US university student sample. From these studies, it is not clear whether exclusively humor, or other genre characteristics, or their combination has positive effect on political learning. Psychological studies, summarized by (Martin 2007b), suggest that humorous contents are easier

to remember. This cognitive effect of humor reflects a selective bias, or in other words one memorizes funny things because they attract attention.

These studies conclude that these entertaining forms of news consumption indirectly increase political activity by reducing the costs of information collection. This thesis is rather interested in direct effects of political humor exposure. My main dependent variables are affective polarization and trust in politicians. Originally, it was planned that the effects of soft and virtue-related jokes on these variables will be tested. It was expected that exposure to the "light side" of political comedy can regain trust in politicians and it can be an anti-dote of polarization. However, the results of the pilot study showed that soft jokes and even positive jokes were perceived hostile. Therefore, I turned toward the investigation of hostile joke effects.

Chapter 2: A new framework for political humor

This chapter introduces original causal theories of political humor, partially based on the review of political communication and political psychology researches. The hypotheses of this thesis are based on these theories and introduced in this chapter. First, hypothesized humor effects, then the theory of partisan humor bias are introduced.

2.1 The dark side of political comedy

In this section, the negative effects of hostile political jokes are theorized. First, it it drawn up how symmetrical hostile political humor exposure would lead to increased level of affective polarization. Second, the mechanism of hypothesized trust eroding joke effects are described. The last subsection speculates on how hostile humor exposure would affect political activity, primary indirectly through the changed level of affective polarization and trust in politicians.

2.1.1 Polarizing effects of hostile political humor

One potential social effect of the widespread usage of political satires, political memes and other humorous political contents can be the exacerbation of partisan divides, more precisely affective polarization. It is a quite new research topic, and many of the studies focuses on the United States, which is considered a seriously polarized country. Iyengar et al. (2019) gives an overview of these studies of affecting polarization. In contrast to ideological polarization (distance between issue positions), affective polarization regards citizens' general negative attitudes toward the other party's/parties' members. It often involves hostility, stereotyping, social distance and lack of trust toward out-partisans (idib: p 130-131).

The literature suggests that ideological polarization could be an important driver of affective polarization, although empirics show that these two types of polarization do not necessarily move together (Rogowski & Sutherland 2016, Webster & Abramowitz 2017 in Iyengar et al. 2019). Other explanations of increasing affective polarization are the strengthening of partisanship as a social

identity, partisan polarization of the media environment, political campaigns (especially attack advertisements against opponents), increasing political homogeneity of families and social environment (ibid 134-136). Iyengar and his colleagues argue that there are mixed evidence behind most of these potential causes of affective polarization, however the study highlights social identity, as a credible starting point for affective polarization, especially when political, ethnic, religious and other identity groups overlap with each other. One of the goals of this thesis is to examine whether political humor plays a role in the increasing level of affective polarization in the United States.

Exposure to mean political jokes in the traditional and new media as well, could be an accelerator of affective polarization. Potential "joke effects" could be considered as special types of media effects. The experiments of Levendusky (2013) showed that not only media exposure congruent with participants' political views increase polarization, but non-congruent media exposure too. This "boomerang" effect, however, was partial and only affected already polarized participants. Potential "joke effects" on polarization may work likewise. Similar to congruent media effects, hostile jokes about the out-party could reinforce stereotypes and negative attitudes and strengthen partisan identity by creating a sense of moral superiority. In the same way, mockery of the in-party by out-partisan can also have a "boomerang effect". The joke may be perceived unrealistic or falsely generalizing (hence trigger a sense of injustice), or as a spot-on insult, "punch below the belt" (which may be even more humiliating). In any cases, negative attitudes toward the "bully" joke-teller and the political side he or she may represent. Therefore, "polarizing joke hypothesis" is the first one this thesis tests:

(H1) Hostile political humor exposure leads to higher level of affective polarization

To my knowledge, only one study tested political humor exposure's effect on affective polarization so far. The experiment of Skoroda (2018) showed that individual level of affective polarization of a subgroup of Republican participants was increased by the exposure to a scene of the Colbert Report, a liberal political satire show. The Colbert Report only affected those Republicans who

watched the scene in entertainment orientation, and it did not affect Republicans who were seeking information and neither Democrats. The entertainment orientation's moderation affect suggests that both cognitive and emotional channels could be activated. Republican respondents realized that the serious tone of Colbert makes fun of Fox News One could conclude from this result that the mockery of the in-party by out-partisans was what increased affective polarization.

What these results hide is to what extent the lack of any polarizing effect among Republicans with information seeking orientation can be attributed to different emotional and cognitive responses to the liberal satire exposure, or simply the failure to understand liberal satire. As mentioned before, people may have problems identifying irony if the joke comes from the opposing political side (Bowyer, Kahne, and Middaugh 2017). Nevertheless, the revealed polarizing effect of liberal satire on Republicans in entertainment orientation shows that political humor effects, as a subgroup of media effects on polarization is worth further investigation.

In the studies showed above (Levendusky 2013, Skoroda 2018), participants were exposed to one-sided media content, either congruent or non-congruent with their political beliefs. It is reasonable to test the effect of symmetrical exposure (containing both congruent and non-congruent contents, which are jokes in this thesis), since it can be a more realistic simulation of real world media consumption. A famous study about news exposure on Facebook claims that in contrary to the beliefs about echo chambers, Facebook algorithms actually show political news from out-partisan news source to users (Bakshy, Messing, and Adamic 2015). This study was widely criticized (e.g. Sandvig 2015), mainly because of the lack of data transparency and accusations that Facebook's interests motivated the results. However, it seems a reasonable claim to think that users encounter out-partisan contents, especially in swing states during campaign periods.

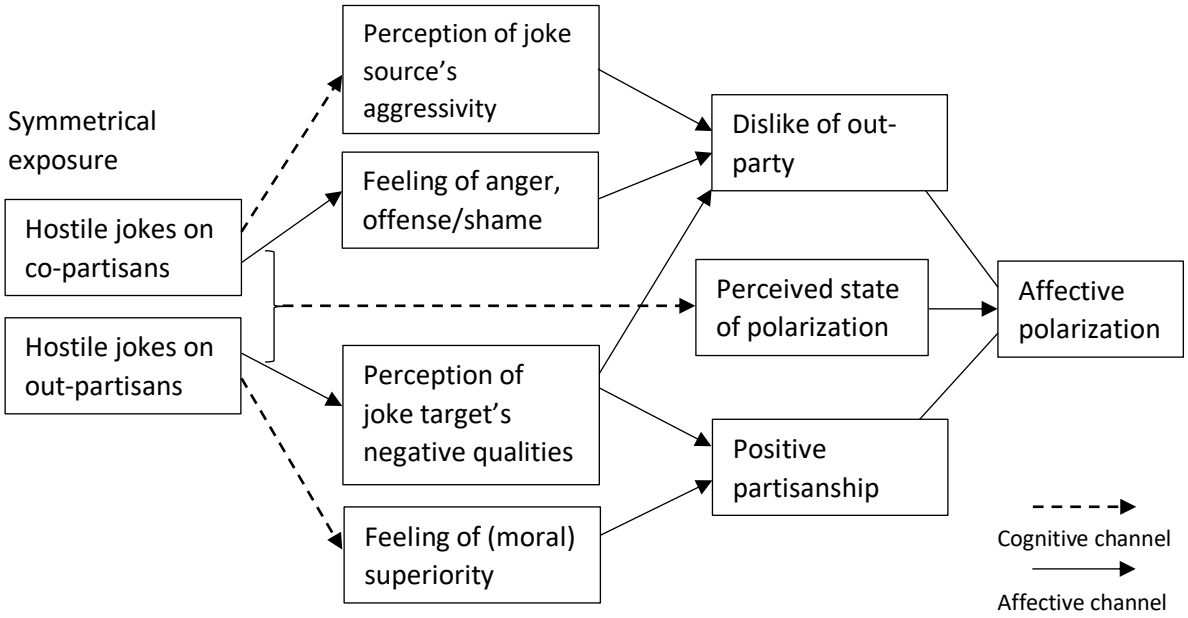

Figure 1: Theorized mechanism of the polarizing joke effect

Regardless of whether citizens are exposed to the state of homogenous or diverse political materials, once non-congruent media effects are analyzed, it is better to embed such materials in a symmetrical exposure treatment. The causal mechanism of symmetrical exposure's effect may be more than the sum of the causal chains of congruent and non-congruent treatments (see: Figure 1). Such diverse political messages could influence affective polarization through an additional channel. Symmetrical exposure is a core element of a politically intense social media environment. Previous study showed that the media coverage of polarization itself increases affective polarization, though decreases ideological polarization (Levendusky & Malhotry 2016a in Iyengar et al. 2019). It is expected that when people see Democrats and Republicans teasing each other, they would be reassured that the level of polarization is high, which may increase their individual level of polarization.

An observational study analyzing the electorate of the US and Israel showed that pro-attitudinal media consumption increased, while counter-attitudinal exposure decreased affective polarization (Garrett et al. 2014). Pro-attitudinal media effect fits well, but counter-attitudinal effect contradicts

my theorized causal mechanism. Mockery can be more easily perceived as an insult compared to sophisticated criticism from out-partisan media. Second, these effects were one sided. Weak evidence from the same study, however, showed that pro- and counter-attitudinal media exposure together slightly increased affective polarization in Israel. Such simultaneous media exposure is conceptually similar to symmetric joke exposure. This result support the supposition that symmetrical exposure has a separate causal chain.

2.1.2 Trust eroding effects of hostile political humor

Political humor may fuel tensions and increase polarization, but somewhat contradictory to these suppositions, it may also disengage people from politics. A straightforward consequence of political criticism can be the erosion of trust in institutions and politicians.

Such criticism may be even more effective when it is embedded in humorous messages. The experiment of Becker and Haller (2014) seems to corroborate this hypothesized effect. Their study relied on a historical categorization of humor, which distinguishes two types of parody: "(a) Juvenalian satire that focuses on the key elements of aggression and judgment and (2) Horatian satire, a form that places greater emphasis on the elements of play and laughter" (Holbert et al. 2011 in Becker and Haller 2014: 36-37). They found out that dark and light types of humor influence differently the evaluation of the comic target David Paterson, New York state's first blind governor. Those experiment subjects who were exposed to a "Juvenalian" type of hostile parody of Paterson evaluated him significantly worse regardless their ideological predisposition than those who saw the governor's "Horatian" self-ridicule. Despite the evidence of the harming effects of "Juvenalian" attacks, the respondents evaluated more negatively this type of humor.

A previous version showed similar results. It compared the effects of the same two types of parodies, but in this study target was Republican senator and presidential candidate John McCain (Becker 2012). These two experiments showed that hard mockery worsens the evaluation of both Democrat and Republican politicians even amongst their supporters. What both studies lacks is a

neutral control group. Therefore, it is hard to determine to what extent the positive effect of self-parody or the negative effect of mockery drove the results. The previously mentioned "The Daily Show effect" experiment used similarly negative, but non-humorous news exposure as a benchmark group and no exposure as a control group for the satire treatment (Baumgartner and Morris 2006). The humor exposure was symmetrical – Daily Show scenes attacking George Bush and John Kerry – and negatively affected both politicians' evaluations. Different general evaluation indices were the dependent variables of these researches, while I examine hostile humor's effect on trust in politicians. I assume that the two variables strongly correlates.

An important assumption is that these negative effects toward co-partisan and out-partisan politicians are additive. Therefore, I expect that a symmetrical exposure to hostile jokes (targeting both Democrats and Republicans) would cause a general, cynical distrust in politicians. Accordingly, "trust eroding jokes hypothesis" describes the second expected effect of hostile political joke exposure.

(H2) Hostile political humor exposure leads to lower level of trust in politicians

Similarly to the potential polarizing effects, trust eroding effects' causal mechanism are expected to incorporate both affective and cognitive channels. Most dark comical styles' core is the criticism of its target (see: Ruch et al. 2018). Humor may make otherwise non-comfortable messages more acceptable. Therefore, mockery may be a more effective tool of persuasion than simple criticism, especially if its target is liked by the audience. Even if this cognitive channel fails, dark humor may alienate the audience from its target, as it is highly uncomfortable to identify yourself with the subject of laughter.

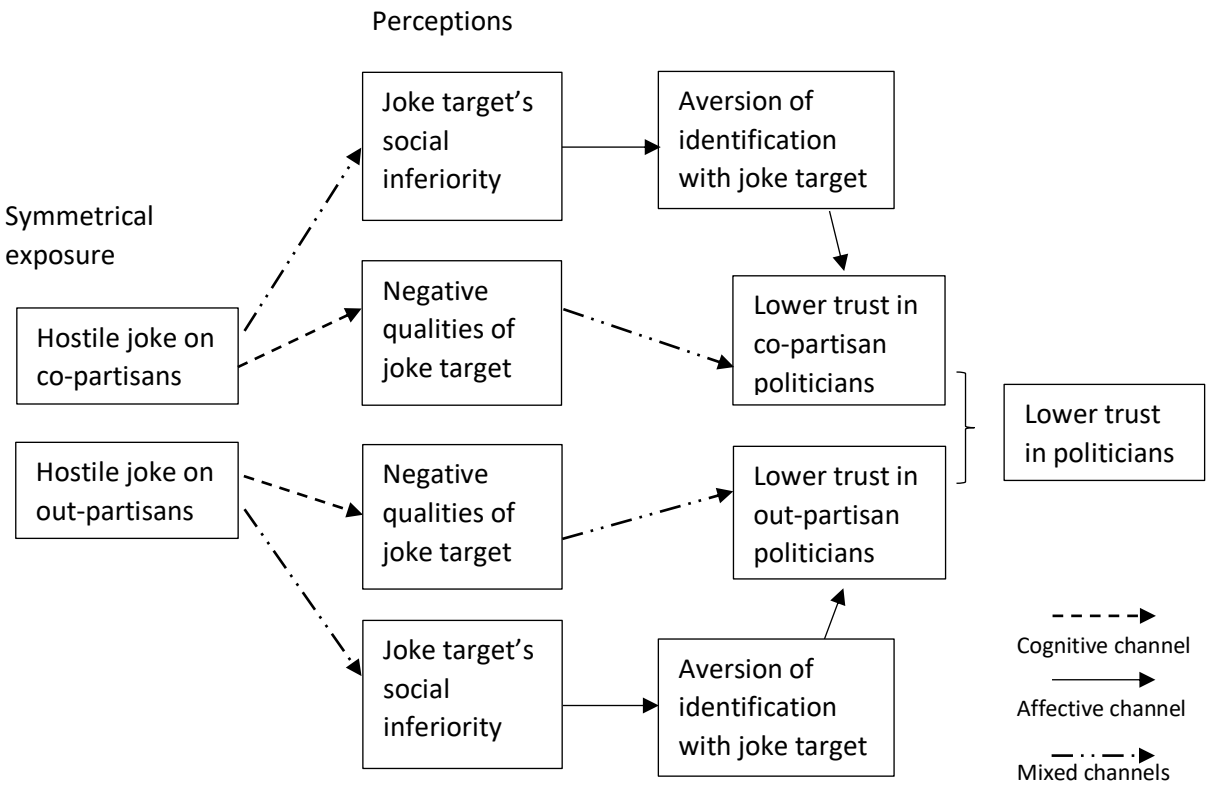

Figure 2: Theorized mechanism of the trust eroding joke effect

Later, the cognitive channel will be observed in the humor effects experiment as an exploratory research. It will be analyzed whether exposure to two jokes (one about Joe Biden's behavior with women and Donald Trump's comments on infecting disinfectant) affected attitudes specifically about the joke target's criticized behaviors.

2.1.3 (De)mobilization, as ripple effect of trust erosion and polarization

Both trust erosion and polarization by hostile political comedy may have ripple effects. Dark jokes may demobilize citizens indirectly. If political mockery indeed discredits politicians and political processes, lower level of trust will decrease the perceived benefits and increase the emotional costs of political engagement. Many liberal democracies experienced a declining level of electoral turnout (Dassonneville and Hooghe 2017) and the United States is the most famous example of declining civic engagement (Putnam 2000). In many cases, the young generations are those who consume

political satires (Holbert et al. 2007) and also who are disengaged from politics (Mann 1999, Mattson 2003, Cammaerts et al. 2014). It is far from being obvious whether there is a connection between the two phenomena, but political humor consumption can be one of the many factors which push people away from political participation.

On the other hand, the logical indirect effect of a polarizing joke would be the opposite. It is a reasonable assumption that if people become more fanatic partisans, they will be more likely to be politically active. It is supported by political psychology researches of expressive partisanship, which is said to be dominant in the United States and Western Europe (Huddy, Bankert, and Davies 2018). "Expressive partisanship motivates a defense of the party in the face of challenging information, leads to the vilification of threatening out-parties, and generates action-oriented emotions that result in heightened political activity" (ibid: 174). In other words, expressive partisanship is a precondition of threats' polarizing and mobilizing effects. Threats to partisanship, as social identity increase both affective polarization and political activity. I expect that this relationship is not one directional, and increased affective polarization also influences political activity.

Previous studies also show contradictory results. The results of the previously mentioned "Daily Show effect" experiment showed that symmetrical satire exposure decreased trust in the electoral system (Baumgartner and Morris 2006). The experiment of Chen, Gan, and Sun (2017) carried out in Hong Kong shows that satire exposure have mobilizing effect. The results showed that counter-attitudinal satires effect was even larger. This mobilization was mediated by increased level of anger. It is not straightforward whether the activation of the emotion represents a separate causal chain, or it is simply the affective channel of polarizing effect (increased dislike of the out-group). It supports the separate channel argument that partisanship is not necessarily and exclusively a social identity. Depending on political context, people may develop instrumental partisanship (besides or instead expressive partisanship) meaning that their partisan support only reflect their issue/policy

preferences (Huddy, Bankert, and Davies 2018). Therefore, humor can have non-group-related direct behavioral effects as well.

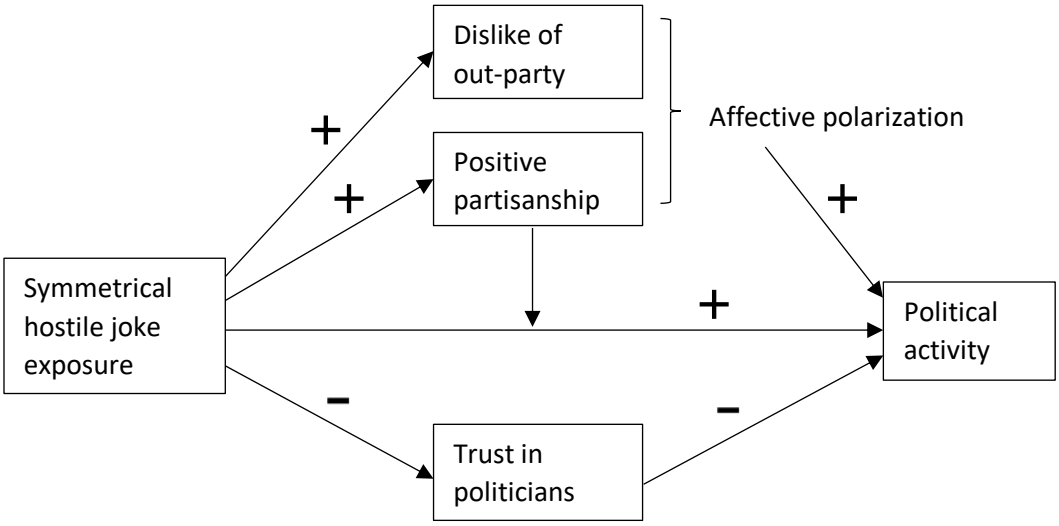

Figure 3: Theorized mechanism of ripple effects

Potential (de)mobilizing effects of hostile joke exposure are analyzed as part of exploratory research. Hypothesis was not formulated for two reasons. As described above and summarized in Figure 3, the expected direct effect, and ripple effects of polarization and trust erosion are contradictory. Furthermore, self-declared political activity is a behavior-oriented attitude that is believed to be harder to influence by experimental manipulation.

2.2 Political humor bias

In the previous chapters, political humor exposure's potential effects were theorized. The first hypothesis identified hostile political humor as a potential cause of affective polarization. The theory introduced in this chapter turns the direction of causation. Reversed causation is a common problem of affective polarization research. For instance, the media consumption is believed to be one of the causes of polarization, but it is also considered a consequence, since highly partisan individuals are more likely to consume partisan media (Arceneaux & Johnson 2013 in Iyengar et al. 2019). Similarly, widescale usage of hostile political humor can be a special form of polarizing

media effects. On the other hand, the politicization of comedy may be the result of growing partisan divide.

These two suppositions do not necessarily contradict each other, political humor can be amongst the causes and consequences of polarization simultaneously. The very recent book of Young (2020) seems to reflect this idea. The book states that satires on the liberal side and outrage opinion shows on the conservative side play the same role. She states that both genres offers political information in a very polarizing manner. She also argues that besides media deregulation and new technologies, polarization was another important factor which opened the stage for these genres.

The second research question of this thesis regards how polarized political humor became. It is investigated whether humor appreciation can cross partisan/ideological lines. It leads to an important question: do people with different political views find different messages humorous? Even if there is a difference between the sense of humor between supporters of different parties, it is not clear what is the underlying cause.

Young (2020) in her previously mentioned book states that there are differences in psychological traits and aesthetic preferences between conservatives and liberals. She argues that these differences explain why liberals are not successful radio hosts and conservatives rarely make political satires. A student sample experiment showed that conservatives found both exaggerating and ironic non-political jokes less funny. The study argued that conservativism's negative effect was primarily mediated by their lower level of need for cognition and sense of humor.

Another explanation can be the lack of background information of the jokes and misperception of tone of the humorous messages which creates "partisan humor". The online experiment of Bowyer, Kahne, and Middaugh (2017) showed that young people could not identify satire in online videos about immigration, if its message opposed their prior stance on the issue. Therefore, they did not understand the message of a political jokes either. Although, the validity of this study is arguable, it seems a reasonable assumption that information effect exists in political humor. In

other words, specific jokes might only function well within the boundaries of the relevant epistemological group. To make it more obvious, let assume that engineers understand mechanic jokes better, as well as Democrats laugh more on satires about government surveillance, solely because they know more about these topics. One can suggest that it is not only information, but different political groups' different levels of sensitivity towards certain issues what drives receptivity to specific types of humor.

If we return to comical styles' categorization (Schmitt-Hidding 1963 in Ruch et al. 2018), one can find an obvious explanation for the existence of these "partisan humor". It is possible that some of these jokes are not satirical, rather ironic, so they never meant to be understood by the outsiders. It seems that the border between different comic styles is only clear in theory, if we talk about political humor. It is possible that any difference between receptiveness to political jokes comes not only from the lack of understanding, but resistance of hostile messages toward one's own political group, represented by its leader. In my research, I pursue filtering out these information causes in order to examine the more direct effects of joke features' effects on appreciation of humor.

This thesis argues that the appreciation and perceived funniness of jokes, especially the ones concerning politics and politicians, depends on whether the jokes are pro or counter attitudinal. Attitudinal congruence was split into two components. It was distinguished whether the joke source and joke target are congruent or not. Congruence means that source/target is on the same or opposite political side as the audience, or non-political. I intend to reveal the hypothesized effects of target and source congruence, which I address as "partisan humor bias". Four groups of humor bias hypotheses were created.

The joke target effect hypotheses state that mockeries will be evaluated worse when they target co-partisans than out-partisans. It seems a straightforward expectation based on the defense

mechanism of expressive partisanship introduced in the previous subsection (Huddy, Bankert, and Davies 2018).

H3.1: People appreciate hostile political jokes less, if it is about their preferred political side compared to a joke about the opposite political side

H3.2: People find hostile political jokes less funny, if it is about their preferred political side compared to a joke about the opposite political side

The two hypotheses differs in the relevant dependent variable: while H3.2 analyzes simply perceived funniness, H3.2 regards more complex attitude toward the joke (see the description of Joke appreciation variable in the Variables and measurement subsection of the 3rd chapter). Testing these hypotheses also serves as a manipulation check of the humor effect experiment. Although the hypothesis tests only regard hostile political jokes, the effect of target congruence of soft and positive jokes is also analyzed in the pilot study as exploratory research.

Partisan joke source hypotheses forms the next group of expectations. I expect that people will appreciate jokes more if it comes from their own side and less if the joke teller is from the opposite political camp. Joke source effects amongst both political and non-political jokes were analyzed as exploratory research in the pilot. Based on the results, hypotheses are limited to only non-political jokes.

H4.1: People find non-political jokes less funny, if the joke source is out-partisan compared to co-partisan joke source

H4.2: People find non-political jokes less funny, if the joke source is out-partisan compared to non-partisan joke source

H4.3: People find non-political jokes funnier, if the joke source is co-partisan compared to non-partisan joke source

Both significant and null results were considered to be meaningful beforehand the research. These hypothesis tests can reveal whether humor can be an escape from partisan fights or polarization is so high that out-partisan source spoil even good jokes. The hypothesized joke source effect is conceptualized as partially a behavioral overspill of affective polarization. Therefore, affective polarization can amplify joke source effect. This expectation is formulated in the partisan source moderation hypothesis:

> *H5: Joke source effect is moderated by affective polarization – the higher individual level of affective polarization is, the stronger partisan joke source effect is.*

Finally, I looked for a non-political factor, which can probably influence the evaluation of the jokes in a magnitude comparable to partisanship. Gender serves as an ideal benchmark for partisan joke source effect. It divides the US population approximately in the same ratio as partisanship. Furthermore, humor studies found gender patterns. Heintz et el. (2018) showed that men use more often corrective humor, but there is no difference regarding benevolent humor. Tosun, Faghihi, and Vaid (2018) showed that men are considered to embody the ideal sense of humor in the United States, Iran and Turkey. Gender source hypothesis is the last one to be tested.

> *H6: People find non-political jokes less funny, if the joke source is (non-partisan) female compared to a (non-partisan) male joke source.*

The (non-)existence of humor bias effects complement well the humor effect theories introduced in the previous chapter. On one hand, humor biases can strengthen polarizing joke effect, especially when people are exposed to symmetrical jokes. A biased partisan would react negatively to counter-attitudinal jokes, and welcome pro-attitudinal jokes. In this way, the magnitude of polarization can be even higher. On the other hand, if it turns out that humor bias is not prevalent, trust eroding effect is likely to be stronger. If people are open to humorous criticism of both co-partisans and out-partisans, exposure to hostile political humor would cause more effectively a general disenchantment from politics.

Chapter 3: Research design

Experimental design is suitable to my research questions and hypotheses, as I aim to reveal causal relationships about political humor's effects. Further reason for an original survey is that voters' political humor consumption habits similarly to political beliefs and voting history can barely be found in existing secondary data. Even if such confidential data is available, these two types of data are not linked with each other on the individual level. The research was planned to be a survey experiment, as it is a cost-efficient way of gaining a representative sample of an acceptable size. This thesis can be considered a special media effect study as well as a political psychology research. My research design resembles the tools used in the experimental researches of these disciplines. It is based on the assumption that the analyzed short-term humor effects are special types of media effects, which are mediated by cognitive and emotional responses of the treated participants of the experiment.

Two separate surveys were conducted, a pilot survey on a limited sample (N=149) and a second survey on a larger sample (N=796). The size of the samples were limited because of budget constraints. The surveys were distributed on the survey platform Lucid.io.. Respondents received financial compensation for filling out the survey.

I created "Tweet-like" vignettes for both surveys. In each vignette, there was a fake profile picture and name, a sentence written by me, and a non-political or political meme. The experimental manipulations were partisan cues (Republican or Democrat) on the profile pictures, or gender cues (male or female name and profile picture). The memes in the vignettes were collected both by general Internet search and more focused search on online comical sites. (Vignettes used in the second survey are presented in Figure A2, Figure A3, Figure A4 in the Appendix). The profile pictures and the names of these vignettes were different for each meme the respondents encountered. Stock photos of young and middle-aged Caucasian males were used in order to avoid any other potential confounding effects, such as effect of ethnicity or age on the relevant dependent

variables. Stock photos of females were only used in the non-political joke of the pilot and the gender and humor bias experiment of the second survey. In both cases, it served as an experimental manipulation (gender effect as benchmark for partisan effect).

In all experiments, simple randomization were used. Respondents had equal chances to get into any of the groups or get any of the cues. Within separate blocks of the surveys, the sequence of memes was also randomized in order to avoid potential priming effects (see Figure 4 and Figure A1 in the Appendix). The design of the second survey, the hypotheses and most statistical analyses were registered as a pre-registration plan on the website of the Center for Open Science (see the pre-registration plan: Szászi 2020). In the following, it is marked in footnotes where any additional, non-registered variables and analyses of the experimental data of the second survey are conducted.

3.1 Pilot study

In the pilot, 9 meme vignettes were shown to the respondents, which was built up by two parts. The first part involved one non-political joke and the second part a series of political jokes. The first part was a simple between subject experiment with 4 groups. Experimental manipulations were partisan and gender joke source cues. The structure of the second part followed a between subject, factorial design (see Figure A1 in the Appendix). There are 8 or 2 levels based on different conceptualization. 8 different jokes are shown to respondents (hard memes about Trump and Biden, hard memes about Republicans and Democrats, positive jokes about Trump and Biden, soft critical jokes about Republicans and Democrats). However, it is simplified to 2 levels in some statistical analysis (co-partisan or out-partisan joke target). There are 2 factors (co-partisan and out-partisan joke source cue).

In the pilot survey, the sentences were slightly different in some vignettes with the same meme but different partisan cues to seem a believable and authentic Tweet. (For instance, a hostile meme about the Republicans with a Republican source cue was below a self-criticizing comment, while it was below an "other criticizing" comment when the joke source cue was Democrat.)

The aim of the pilot survey was twofold. Firstly, I aimed to test the planed effectiveness of the vignettes I constructed, more precisely whether they are perceived as hostile or friendly enough to simulate a politically intense social media environment and a political friendly social media environment. Secondly, it aimed to examine the hypothesized humor bias (joke source effect and joke target effect) on a limited sample first. The second survey was re-designed by using the lessons from the analysis of the pilot survey.

3.2 Second survey

The second survey consisted three separate experiments (for the structure of the survey see Figure 5). The partisan humor bias experiment was embedded in the treatment and baseline groups of the humor effect experiment. The partisan and gender humor bias experiment was after the follow-up questions of the humor effect experiment. In the humor effect experiment, respondents were randomly assigned to either the treatment group (hostile political meme exposure) or the baseline group (non-political meme exposure). Within each group, the sequence of the memes was randomized. The joke target effect part of the partisan humor bias experiment was the treatment group of the humor effect experiment. The joke source effect part of the partisan humor bias experiment was the baseline group of the humor effect experiment. It contained non-political jokes with randomized partisan cues (Republican or Democrat joke source) for each meme. In the partisan and gender humor bias experiment, the cues (Republican male, Democrat male, non-partisan male, non-partisan female) and the sequence were both randomized.

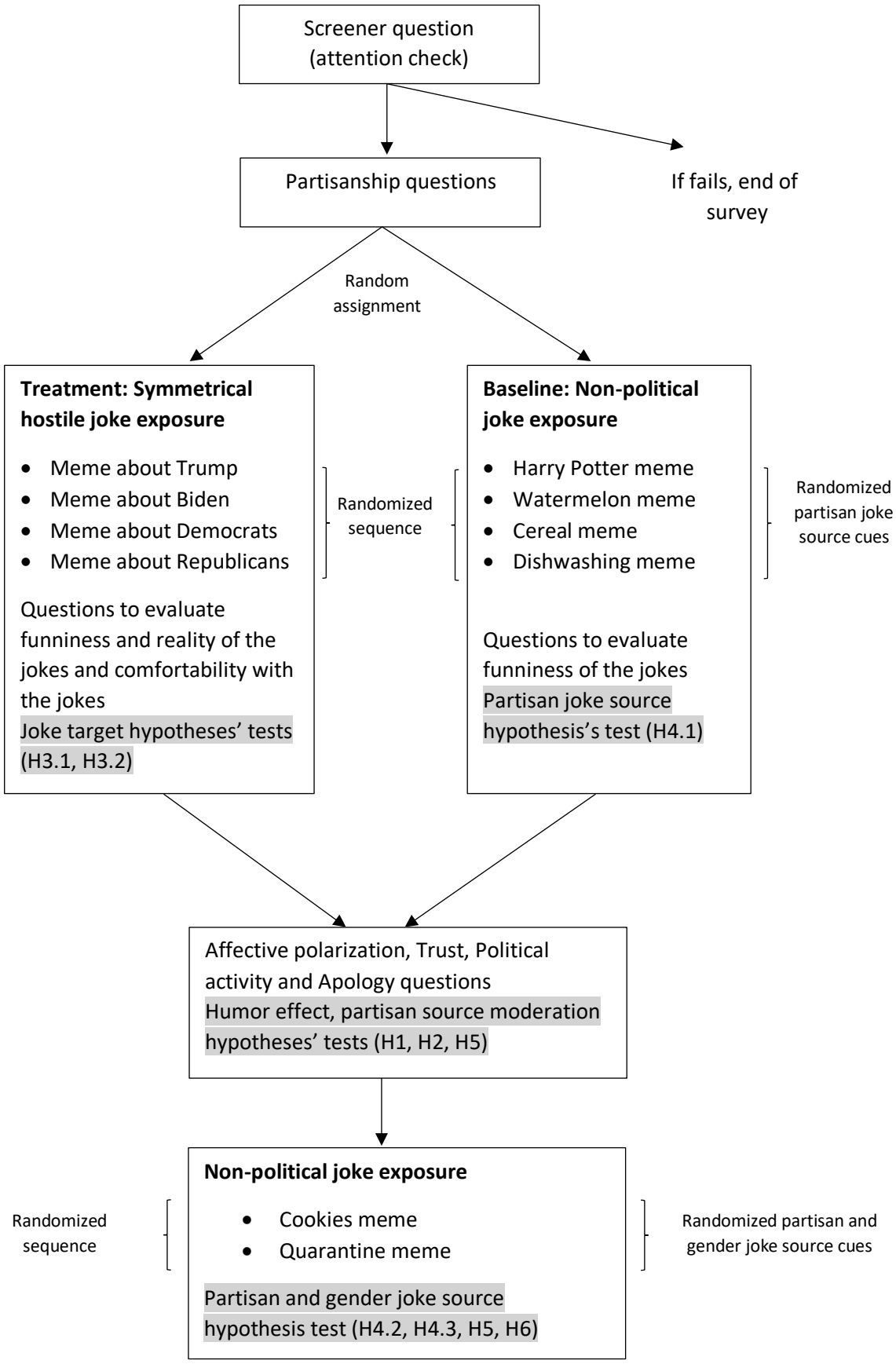

Figure 4: Structure and research design of the second survey

3.2.1 Humor effects experiment

The first experiment was a two group, between-design experiment. In this experiment the treatment group was exposed to four hostile political memes (targeting both Republican and Democratic voters and President Donald Trump and presidential candidate Joe Biden). Four non-political memes were shown to the baseline group. All memes were embedded in "Tweet-like" images and respondents are asked to evaluate them.

After the "meme exposure", the survey continued with affective polarization and trust questions (measuring the main dependent variables of this experiment). The next question measured political activity[1]. Then, two questions came in randomized order, one asked President Trump should have apologized for his comments about injecting disinfectant, and one about whether Vice President Biden should have apologized for hist past behavior with woman. These specific questions were asked because the treatment group's hostile memes on Donald Trump and Joe Biden made fun of these acts and accusations. Although these questions were preregistered, they serve the purpose of an exploratory analysis too. It aims to check the theorized cognitive channel of humor effect on trust in politicians.

The polarizing joke hypothesis (H1) and the trust eroding joke hypothesis (H2) were tested with simple OLS regressions with the following equations:

$$Affective\ polarization_i = ß0 + ß1 * Humor\ exposure\ type_i + Ɛ$$

$$Trust\ in\ politicians_i = ß0 + ß1 * Humor\ exposure\ type_i + Ɛ_i$$

3.2.2 Partisan humor bias experiment

The second experiment is embedded in the baseline and treatments groups of the humor effect experiment. Partisan humor bias experiment had two parts: one analyzing the hypothesized joke source effect and one analyzing the joke target effect.

[1] This question was added to the survey after submitting later to the survey, therefore it does not appear in the pre-registration plan.

Respondents assigned to the baseline group of the humor effect experiment participated in the joke source effect part of this experiment. It was registered as a factorial design experiment with 2 factors (co-partisan or out-partisan joke source cues) and 2 levels (funny and not funny jokes). For "funny jokes", I selected memes which had many positive ratings on comical websites. For "not funny jokes", I selected memes which were a few days old, but received only few positive ratings. However, there was no significant difference between the evaluation of jokes, which I considered "funny" and "not funny". Therefore, I did not include this flawed joke funniness variable in the analysis.

It was analyzed whether the congruence of the joke source's and the respondent's partisanship influence the evaluation of the joke (measured by *Joke appreciation* index and *Perceived funniness*, see at Variables and measurement). I run clustered OLS regression with the following equation to test the first joke source effect hypotheses (H4.1, H4.2).

$$Percieved\ funniness_{i,j} = ß0 + ß1 * Joke\ source\ congruence_{i,j} + \varepsilon_{i,j}$$

The observations were the given jokes (marked by j) evaluation by the given respondents (marked by i). The regressions were clustered by respondent ID because one respondent appeared several times in the analyzed database (multiplied by the number of analyzed jokes) and this way I could filter out individual effects.

Respondents assigned to the treatment group of the humor effect experiment participated in the target effect part of the humor bias experiment. The test of target effect hypotheses (H3.1, H3.2) is more observational than experimental, as the effect of the congruence of the political jokes' targets and respondent's partisanship is analyzed. These hypotheses were tested with the following clustered OLS regressions.

$$Joke\ appreciation_{i,j} = ß0 + ß1 * Joke\ target\ congruence_{i,j} + \varepsilon_{i,j}$$

$$Percieved\ funniness_{i,j} = ß0 + ß1 * Joke\ target\ congruence_{i,j} + \varepsilon_{i,j}$$

3.2.3 Partisan and gender humor bias experiment

The third experiment was right after the follow-up questions of the humor effect experiment and all respondents were part of it. It is a four group[2], between design experiment. Two additional non-political memes are shown to the respondents in randomized sequence. The groups differ in the joke source cues. Randomized partisan and gender joke source cues were used (male Democrat, male Republican, male non-partisan, female non-partisan joke sources). Participants were similarly asked to evaluate the funniness of these memes.

The remaining partisan joke source hypotheses (H4.2, H4.3) and the gender source hypothesis (H6) were tested with another clustered analysis with the following equation.

$$Percieved\ funniness_{i,j} = ß0 + ß1 * Partisan\ congr.and\ gender\ cue_{i,j} + Ɛ_{i,j}$$

The joke source effect moderation hypothesis was also tested with clustered OLS regression with the following equation.

$$Percieved\ funniness_{i,j} = ß0 + ß1 * Partisan\ congr.and\ gender\ cue_{i,j} + ß2 * Affective\ polarization_{i,j} + ß3 * Partisan\ congr.and\ gender\ cue_{i,j} * Affective\ polarization_{i,j} + Ɛ_{i,j}$$

3.5 Variables and measurement

3.5.1 General variables

Partisanship is an important control variable. In the pilot, it was constructed based on the survey platform provided individual data. *Partisanship* variable is Democrat when respondents were previously categorized as one the following categories: Strong Democrat, Not very strong Democrat, Other -Leaning Democrat, Independent Democrat. It is Republican when respondents were previously categorized as one the following categories: Strong Republican, Not very Strong

[2] It has been preregistered as a two group between design experiment by mistake, but the cue-based groups were specified correctly.

Republican, Other - Leaning Republican, Independent Republican. Other – neither and Independent categories were excluded from the analysis of the pilot. In the second survey, *Partisanship* was constructed based on my own survey questions, as the survey platform provided partisanship does not follow standard procedure. *Partisanship strength* variable was also used in the second survey, it was either strong or weak. It is strong for respondents, who identified themselves as Strong Democrats/Republicans. It is weak for respondents who identified themselves as Not so strong Democrats/Republicans. It is also weak for respondents who did not identify themselves as Democrats/Republicans first but chose one of the parties in a follow-up question about which party is closer to them. It is Independent for respondents who identified themselves as Independent or Other and chose the "None of them" option in the follow-up question.

3.5.2 Manipulated variables

In the humor effect experiment, the main independent variable is *Humor exposure type*. It is a dichotomous variable which can be either non-political humor exposure (baseline), hostile political humor exposure (treatment).

Joke type is a categorical variable and it is defined by which vignette respondents evaluated. It can be either Joke on Democrats, Joke on Republicans, Joke on Biden, Joke on Trump, or Neutral.

Joke target congruence is a dichotomous variable constructed from the *Joke type* and *Partisanship* variables. It is only created for the respondents who are in the treatment group of the Humor effect experiment. It is Congruent if the *Joke type* is Joke on Republicans or Joke on Trump and *Partisanship* is Republican; or *Joke type* is Joke on Democrats or Joke on Biden and *Partisanship* is Democrat. In any other case (if *Joke type* or *Partisanship* is not missing), it is Non-congruent.

Partisan cue (Republican or Democrat) and *Partisan and gender cue* (Republican male, Democrat male, Non-partisan male, Non-partisan female) are randomly created categorical variables. These variables and *Partisanship* are used to construct the *Partisan cue congruence* and *Partisan congruence and gender cue* variables (see below).

In the Partisan humor bias experiment, the main independent variable is *Partisan cue congruence*: congruence between the partisanship of the respondent and the given meme's source (joke source partisan cue). It is a dichotomous variable (congruent or non-congruent). It is constructed the following way. When *Partisanship* and *Partisan cue* are the same, it is Congruent. When it is not (one is Democrat, the other is Republican and vice versa), it is Non-congruent. (When either *Partisanship* or *Partisan cue* is missing, missing data is generated.) This variable is used for the analysis of the Partisan humor bias experiment.

In the partisan and gender humor bias experiment, the main independent variable is the *Partisan congruence and gender cue* of the meme source. It is a categorical variable, it can be either congruent male, non-congruent male, non-partisan male or non-partisan female. It is constructed similarly to *Partisan cue congruence*. It is constructed using *Partisanship* and *Partisan and gender cue*. The only difference is that there are two additional categories. When *Partisan and gender cue* is Non-partisan male or Non-partisan female, it is the same for *Partisan congruence and gender cue* variable. Therefore, it can be Congruent, Non-congruent, Non-partisan male (baseline) or Non-partisan female. This variable is used for the analysis of the partisan and gender humor bias experiment.

3.5.3 Measured variables

Most dependent variables are measured on a 100 point scale (0 'not at all', 50 'somewhat', 100 'very'). In the pilot, different labels were used (0 'not at all', 50 'somewhat', 100 'extremely'), which was changed because it seemed that there was a relatively small variance of the responses given to these questions.

Humor effect experiment:

There are two main dependent variables in the humor effect experiment: *Affective polarization* and *Trust in politicians*. *Affective polarization* variable is an index of several answers to the question "How do you feel about these groups and individuals?". In two separate blocks, respondents are asked to evaluate their feelings toward presidential candidates and their supporters (Donald Trump, Joe

Biden, Trump supporters, Biden supporters), and parties' politicians and voters (Republican voters, Democratic voters, Republican politicians, Democratic politicians). These feelings are measured on a 100 point scale (0 'extremely negative', 50 'neither positive nor negative', 100 'extremely positive'). Affective polarization is measured as the difference of the means of the feelings toward co-partisan and out-partisan subjects.

The sub-components of *Affective polarization* variable were used for additional analysis. *Candidate polarization* is the mean of the differences between co-partisan and out-partisan presidential candidates and their co-partisan and out-partisan supporters. *Party polarization* is the mean of the differences between co-partisan and out-partisan politicians, and co-partisan and out-partisan voters. Another categorization of affective polarization was also used. *Feelings towards in-party* and *Feelings toward out-party* were simply the means of the evaluations of presidential candidates and their supporters, partisan politicians and voters (co-partisan and out-partisan accordingly).

Trust in politicians variable is the answer to the question "How much do you trust politicians?", it is a simple 100 point scale as described above.

As a benchmark, additional dependent variables are used, namely *Trust in judiciary*, *Trust in police*, *Trust in media.* (Measured the same way as trust in politicians.)

Additional dependent variables are *Apology Biden* and *Apology Trump*, which are dichotomous variables (Yes or No). These are answers to the following questions accordingly: "Do you think President Trump should have apologized for his comments about injecting disinfectant?" and "Do you think Vice President Biden should have apologized for his past behavior with women?"

Another additional variable is *Political activity*. This variable is not in the preregistration plan, as the question measuring political activity was added to the survey after the submission of the preregistration. It is an mean of the answers to several questions, namely "How likely would you … ?" "engage in an online political debate", "vote in presidential election", "volunteer in a political

campaign" and "protest for an issue important for you". It was measured on a 100 point scale too (0 'very unlikely", 25 'unlikely'. 50 'neither likely, nor unlikely', 75 'likely', 100 'very likely').

Partisan humor bias experiment; partisan and gender humor bias experiment:

Perceived funniness is the main dependent variable. It is the answer to the question "how funny is this joke?", it is a simple 100 point scale as described above.

Joke appreciation is an index variable, which tests H3 and also serves as a manipulation check for the humor effect experiment. It is only measured in case of political jokes (treatment group of Humor effect experiment). It is constructed as the mean of the answers to the following two questions: "Are you comfortable with this joke?" and "Is it a good depiction of reality?" These questions are in the same block with the funniness question and measured the same way (0 "Not at all" 50 "Somewhat" 100 "Very"). It is conceptually different from funniness, as it aims to better capture negative vs. positive attitudes towards a political joke than a simple funniness question.

Chapter 4: Analysis

The statistical analysis of the experimental data from the pilot study and the second survey serves the purposes of both exploratory study and hypothesis testing. The pilot study is exclusively exploratory besides preparing the second survey. The second survey was pre-registered, the empirical strategy of the hypothesis testing and most of the exploratory analysis were conducted according to the pre-registration plan. Some additional exploratory analysis were added after the submission of the pre-registration, these cases are marked in footnotes in the following.

4.1 Pilot study

169 full responses were collected for the pilot survey. As part of data cleaning, impartial responses were deleted. Independent respondents (20) were excluded from statistical analysis, therefore 149 respondents remained. Because of a bug in the survey, respondent's evaluation of the hard, hostile joke about Trump from the first day of data collection were excluded too.

First, the evaluation of the neutral joke was analyzed. None of the joke source cues influenced significantly the joke appreciation index (see: Table A1 in the Appendix). I found statistically significant effect of the experimental manipulations in only one of the models with perceived funniness as dependent variable (see: Table A2 in the Appendix and Figure 5 bellow). The model analyzed the subset of male respondents, and female joke source had a significant and heterogenous effect. Female source was significant at 10 percent, while its interaction with partisanship was significant at 1 percent. Democrat males found 27.5 points funnier the same joke from a female non-partisan joke source than from a male non-partisan joke source (which is the baseline group in the following). Republicans, on the other hand, found the same joke 28.5 less funny from a female than from a male non-political joke source.

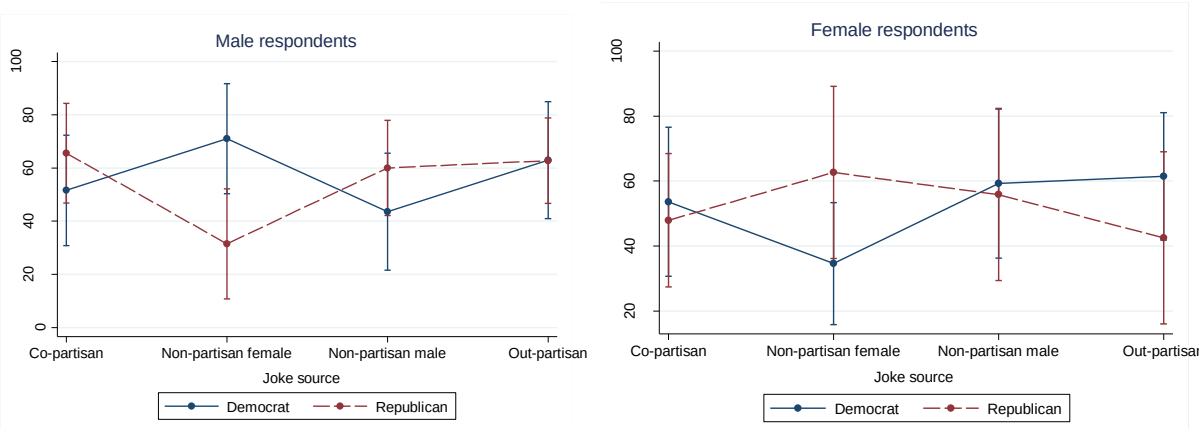

Figure 5: Source cues' effects on perceived funniness, by partisanship and gender (adjusted predictions, 95 percent confidence interval)

Although no other coefficient were statistically significant, some of the effects' sizes were substantial. It can be assumed that the main reason behind the lack of significant results was the small sample size. Just to mention some of the notable results, female Democrats evaluated 24.5 points less funny the joke from female source. Female Republicans, on the other hand, evaluated the joke less funny from partisan sources (both co- and out-partisans), but not from non-partisan female. Democrat males evaluated the joke from out-partisans funnier. These findings deserve further empirical analysis, therefore non-political jokes with simple partisan cue manipulations as well as partisan and gender cue manipulations were built-in the second survey.

Then, political jokes' evaluation was analyzed. It was checked whether there is a difference in the perceived hostility of the political jokes between respondents with different partisanship. Democrats found the jokes on average less hostile towards the jokes targets (see: Figure 6). They found the most hostile the hard joke about Republicans, Democrats and Biden and soft joke about Democrats, though these are still only slightly more hostile values than the central "somewhat hostile" (50 point). Republicans found all jokes more hostile (compared to Democrats). They found the jokes about Republicans slightly more hostile.

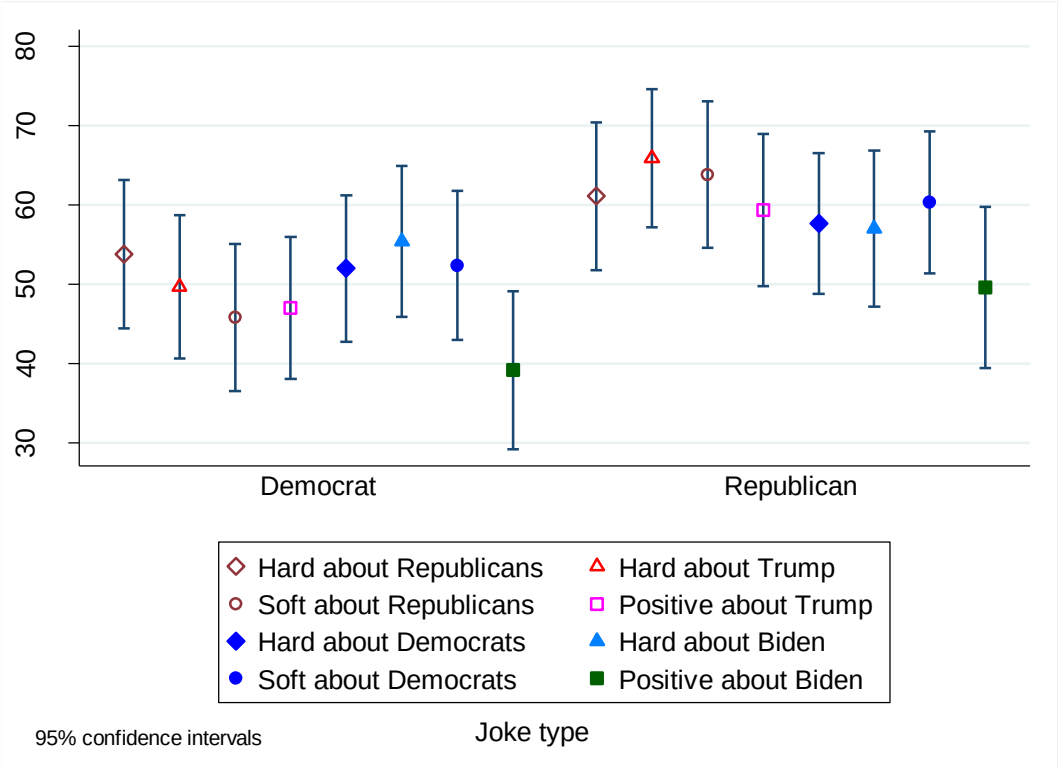

Figure 6: Perceived hostility of the jokes, by respondent's partisanship

Note: This was measured by the question "How offensive this joke is for Democrats/Republicans? (according to the joke targets) on a 100 point scale (0 'Not at all', 50 'Somewhat', 100 'Extremely')

It should be noted that these differences are only amongst point estimates and there are relatively large standard errors overlapping with each other. Therefore, all jokes were analyzed in separate OLS models first. It was analyzed to what extent differences in perceived hostility can be explained by the congruence of the respondent's and the joke target's partisanship, and the congruence of the respondent's and the joke source's partisanship. There was only significant difference in two cases. Republican respondents, whom the soft joke about Republicans was shown with a Democrat joke source, found it 25 points more hostile on average. Republicans also found the hard joke on Trump more offensive (21 points) regardless of the joke source's partisanship.

It is surprising how emphatic respondents were evaluating the offensiveness jokes about out-partisans. The main reason behind this could be the question formation, which may prime respondents to evaluate the jokes from the perspective of the political group it is about. Therefore, different evaluations of the jokes (appreciation of the joke index) were also analyzed.

Table 1: OLS regressions, dependent variable: perceived offensiveness of political jokes

	How offensive is the joke for Republicans/Democrats? (accordingly to the joke's target)							
VARIABLES	'Hard Rep'	'Hard Dem'	'Soft Rep'	'Soft Dem'	'Hard Trump'	'Hard Biden'	'Soft Trump'	'Soft Biden'
Out-partisan joke source	-2.120	-4.871	-8.192	0.493	14.15	6.126	1.492	8.841
	(7.913)	(7.974)	(7.959)	(8.177)	(8.781)	(8.221)	(8.143)	(8.698)
Republican	9.353	-1.292	-3.105	8.075	21.23**	-11.75	2.657	6.591
	(7.913)	(7.974)	(7.910)	(7.775)	(8.702)	(8.271)	(7.910)	(8.801)
Out-partisan joke source X Republican	-4.564	-5.866	24.90**	-9.449	-9.730	9.637	3.955	-4.578
	(11.07)	(11.16)	(11.14)	(11.43)	(12.49)	(11.50)	(11.39)	(12.15)
Constant	55.86***	59.03***	52.77***	52.32***	42.59***	55.62***	47.95***	34.91***
	(5.709)	(5.753)	(5.743)	(5.498)	(6.209)	(6.009)	(5.557)	(6.438)
Observations	149	149	149	149	107	149	149	149
R-squared	0.017	0.019	0.055	0.011	0.085	0.039	0.008	0.012

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: First day of data collection of hard Trump joke excluded because of error in the survey

Based on the point estimates, it seems that appreciation index of the jokes (especially hard ones) is driven by the congruence between partisanship of the respondent and the joke target (see: Figure 7). Democrat respondents appreciated the most hard jokes about Republicans and Trump, and the soft joke about Biden. They appreciated the least the hard joke about Democrats, they appreciated slightly more hard joke about Biden and soft joke about Democrats. They appreciated "somewhat" soft jokes about Republicans (around 50 points) and Trump. Republican respondents appreciated the most hard and soft jokes about Biden and soft joke about Democrats and slightly less the hard joke on Democrats. They appreciated the least the hard joke about Republicans, and other jokes targeting their side (hard and soft joke on Trump, soft joke on Republicans) were slightly bellow 50 points.

Joke appreciation index was analyzed in separate OLS regression models too, which showed that Republican appreciated less all jokes about their side less regardless of the partisanship of the joke

source. This effect was both substantial and statistically significant at 1 percent. This effect was larger in the cases of hard jokes: Republicans appreciated hard joke on Republicans 36 point less, hard joke on Republicans 24 point less, while they appreciated soft jokes on Republicans 15 point less and soft joke on Trump 14.5 point less.

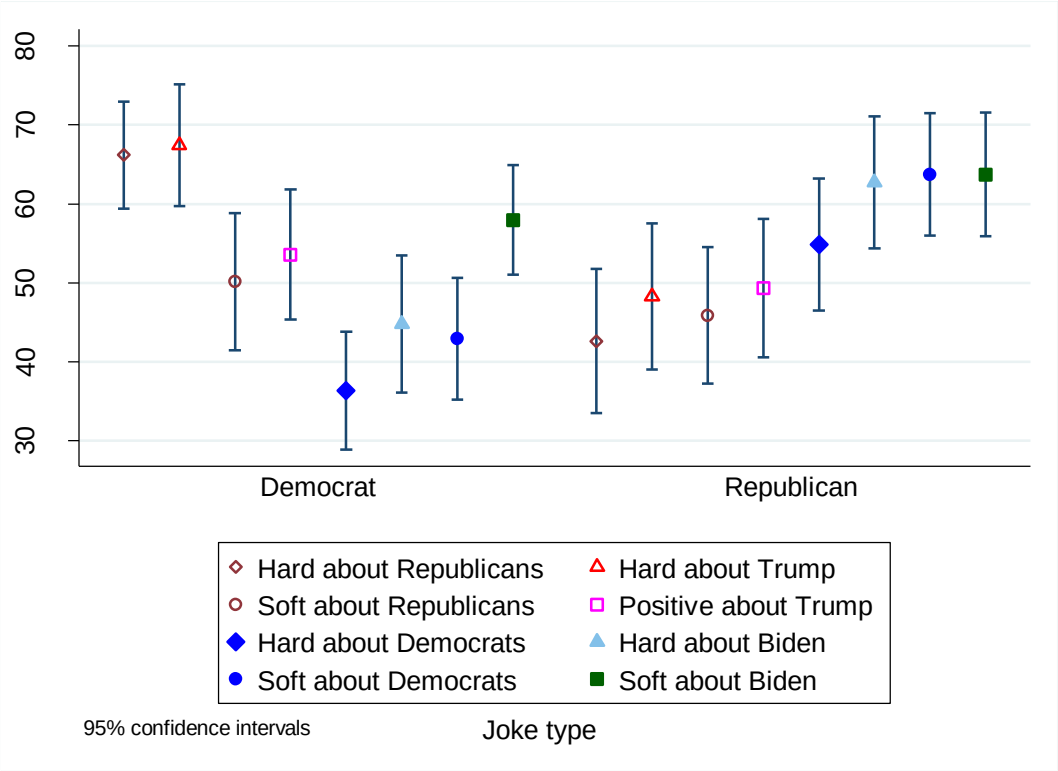

Figure 7: Appreciation indices of the jokes, by respondent's partisanship
Note: This is an index variable, which is mean of the answers to the questions "Are you comfortable with this joke?" and "Is it a good depiction of reality?", both measured on a 100 point scale (0 'Not at all', 50 'Somewhat', 100 'Extremely')

Table 2: OLS regressions, dependent variable: joke appreciation index of political jokes

	Joke appreciation index							
VARIABLES	'Hard Rep'	'Hard Dem'	'Soft Rep'	'Soft Dem'	'Hard Trump'	'Hard Biden'	'Soft Trump'	'Soft Biden'
Out-partisan joke source	-4.524	-4.346	-8.479	6.278	-1.574	-16.74**	-5.491	-2.864
	(6.820)	(7.199)	(7.163)	(6.890)	(8.482)	(7.508)	(7.381)	(6.583)
Republican	-36.20***	8.496	-14.80**	30.61***	-24.42***	4.214	-13.54*	-3.571
	(6.820)	(7.199)	(7.118)	(6.551)	(8.406)	(7.554)	(7.171)	(6.661)
Out-partisan joke source X Republican	13.62	15.79	9.480	-18.66*	11.03	20.76**	2.872	16.44*
	(9.545)	(10.08)	(10.03)	(9.632)	(12.06)	(10.50)	(10.32)	(9.197)
Constant	67.47***	38.21***	55.07***	38.81***	68.24***	52.84***	58.06***	58.44***
	(4.921)	(5.194)	(5.168)	(4.632)	(5.997)	(5.488)	(5.037)	(4.873)
Observations	149	149	149	149	107	149	149	149
R-squared	0.216	0.086	0.035	0.147	0.099	0.086	0.042	0.038

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: First day of data collection of hard Trump joke excluded because of error in the survey

Regarding the jokes about Democratic side, the effect of partisanship and joke source cue were heterogenous. In case of soft jokes about Democrats partisanship and the interaction term of partisanship and source cue were both significant at 1 percent. It means that Democrats appreciated substantially less this (by 31 points on average) than Republicans when the joke source was co-partisan. This difference between Republican and Democrat respondents was less than half of this size when the joke source was out-partisan.

In case of hard joke on Biden, source cue and interaction terms were both statistically significant at 5 percent. It means that Democrat respondents appreciated 25 points lower this joke when the source cue was out-partisan. The direction of this effect turns amongst Republican respondents and reduces to 4 points. Hard joke on Democrats showed similar effects in terms of direction and magnitude, though these effects were not statistically significant. In case of soft joke on Biden, only the interaction term was significant. It means that Democrats liked 13 points less this joke than Republicans when the joke source was out-partisan.

All in all, except the soft joke about Democrats, what these results tell about jokes about the Democratic side is that Democrat respondents appreciated them less than Republicans, especially compared to those Republican respondents who saw the meme with an out-partisan (Democrat) source cue.

Data was restructured for further analysis in the following way. In the previous form of the dataset, each rows identified separate respondents and the evaluation of different jokes were recorded as different variables. In the new format, each row identifies one respondent and one joke type (which is a new variable), and there are no different joke evaluation variables for different joke types. Accordingly, the number of observations were multiplied by 8 (the number of political jokes), 8 rows belong to each respondent. Every one of these rows has different *Joke type* values and evaluation variables regard the given joke types. This restructured data format enables integrated analysis of all jokes in single models. OLS regression were run on this pooled data with three different dependent variables (funniness, hostility and joke appreciation). These regressions were clustered by respondent ID to filter out the effects of the respondents' individual characteristics.

Only joke target congruence had a statistically significant effect in these models. When a meme was about the opposing political side, respondents evaluated it 10 points funnier on average (9 points funnier when partisanship and its interactions were included too) and its appreciation index was 16 points higher (15 points higher in the augmented model).

Table 3: Clustered OLS regressions, dependent variable: pilot joke evaluations

VARIABLES	(1) 'Funniness'	(2) 'Funniness'	(3) 'Hostility'	(4) 'Hostility'	(5) 'Joke appreciation'	(6) 'Joke appreciation'
Out-partisan joke source	-0.822	-4.910	3.426	2.342	-0.107	-3.878
	(3.117)	(3.964)	(3.197)	(4.408)	(2.841)	(3.493)
Out-partisan joke target	10.38***	9.049**	-3.172	-0.557	16.18***	15.15***
	(2.811)	(3.733)	(3.230)	(4.267)	(2.494)	(3.142)
Republican respondent		-3.003		6.100		-6.878
		(5.372)		(5.469)		(4.842)
Out-partisan joke source X Out-partisan joke target	1.063	1.644	-3.190	-1.781	-0.198	-1.288
	(4.546)	(6.163)	(4.559)	(6.508)	(4.036)	(5.431)
Out-partisan joke source X Republican		8.390		2.550		7.545
		(6.229)		(6.308)		(5.705)
Out-partisan joke target X Republican		2.673		-5.262		2.318
		(5.571)		(6.422)		(4.914)
Out-partisan source X Out-partisan target X Republican		-1.686		-3.158		1.683
		(9.142)		(9.148)		(8.083)
Constant	39.94***	41.44***	53.78***	50.72***	43.13***	46.58***
	(2.688)	(3.554)	(2.753)	(3.634)	(2.440)	(2.955)
Observations	1,150	1,150	1,150	1,150	1,150	1,150
R-squared	0.024	0.028	0.006	0.012	0.064	0.070

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: First day of data collection of hard Trump joke excluded because of error in the survey

The following lessons were built in the re-design of the second survey. Based on these results, it seems that the congruence of respondent and joke target drives solely the perceived funniness and appreciation of political jokes. Therefore, I decided that in the second survey, the partisan joke source effect were going to be tested only on non-political jokes.

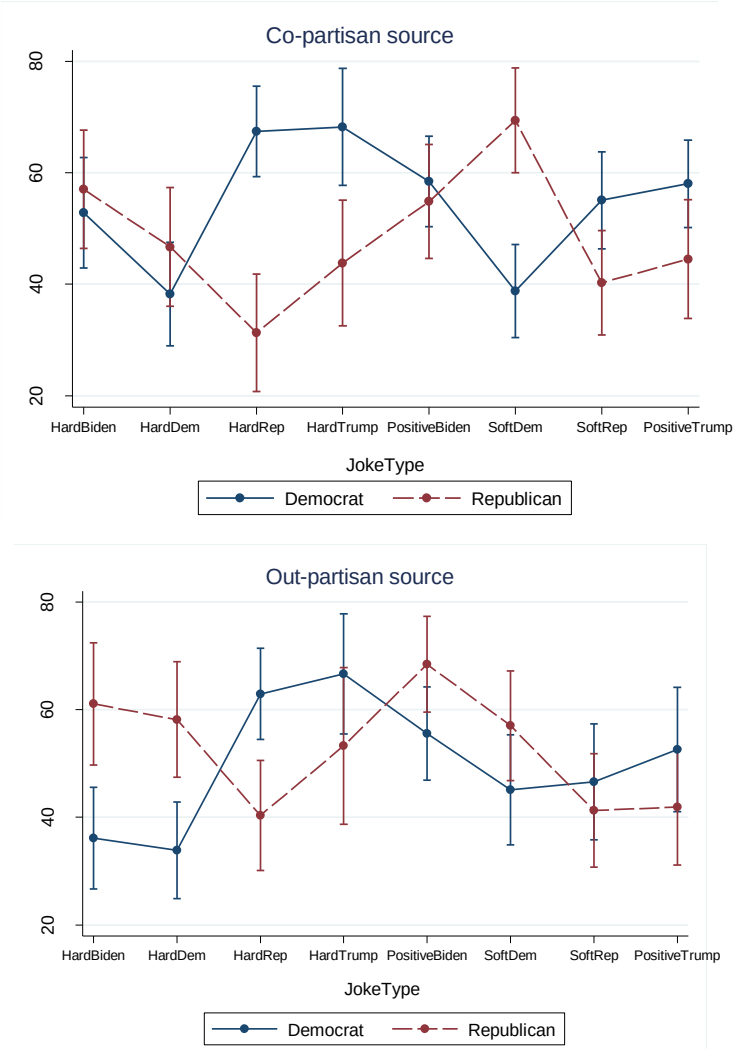

Figure 8: Predicted margins of the political jokes' appreciation from a clustered regression, by partisan source congruence
Note: Independent variables: partisanship, joke type, interaction, adjusted predictions with 95 percent confidence interval)

Furthermore, the idea of a "soft joke treatment" was rejected. This soft joke treatment was originally planned to be part of the humor effect experiment of the second survey, and it was expected to reduce affective polarization and strengthen trust in politicians. This treatment was planned to simulate a friendly, but political social media environment, but these jokes could not serve this purpose, since they were evaluated hostile especially when the target of the joke was on the same side as the respondent (except from the soft joke on Biden).

On the other hand, the Pilot showed that hard jokes can be used to simulate a politically intense, hostile social media environment, hence these hard jokes were used as vignettes in treatment group of the humor effect experiment of the second survey. Democrats appreciated these jokes the less

both with co-partisan and out-partisan source cue (see Figure 4). Republicans, however, evaluated soft jokes about their side worse than the hard joke about Trump, which was still selected into the second survey for the sake of creating a symmetrical treatment of the humor effect experiment**,** which attacks both partisans and party leaders on both sides. Surprisingly, Republicans appreciated more hostile jokes about their side from out-partisans (Democrats), though these are not significant differences. However, it was decided that all hostile jokes will be presented with partisan cues which are non-congruent with the joke's target (Republicans make fun of the Democratic side and vice versa) because it is more important to show authentic and believable "Tweet vignettes" and simulate inter-party hostility than maximizing the level of negative emotions triggered by the joke.

4.2 Humor effects experiment

First, the data from the humor effect experiment were statistically analyzed. OLS regressions were run to test the first two hypotheses. It was analyzed whether hostile political humor exposure, as our experimental manipulation, leads to increasing affective polarization and weakening trust in politicians.

4.2.1 Effects on affective polarization

First, the polarizing joke effect hypothesis was tested. Three different measurement of affective polarization were used as dependent variables in 6 different OLS regression models (see: Table 4). To test H1, the general affective polarization index was used in the first model where Humor exposure type was the only independent variable. The model showed a very surprising result. The hostile joke treatment turned out to have a moderately substantial effect which was also significant at 5 percent. However, the effect was opposite of the expected direction. The hostile political joke treatment, which aimed to simulate a politically intense social media environment turned out to reduce affective polarization by 6.56 points compared to the baseline group, whose members faced non-political jokes.

As an exploratory analysis, two other polarization variables, namely Candidate polarization and Party polarization were used too. These models were not registered in the pre-registration plan, however because of the surprising results of the first model, it was reasonable to check whether these results stand for the components of general polarization index too. Similarly, not pre-registered models involving partisanship and its interaction with humor exposure type were also run. These analyses were included to find out whether hostile humor exposure's type's effect is moderated by partisanship and whether this effect is driven by a partial effect on a partisan subgroup.

The negative effect of the hostile joke treatment remained approximately the same size in the simple models with Candidate polarization and Party polarization indices. This effect was only significant at 10 percent and at 5 percent in the Candidate polarization and Party polarization models accordingly.

Table 4: OLS regressions, Hostile political joke exposure's effect on affective polarization (baseline: non-political joke exposure)[3]

VARIABLES	Affective polarization		Candidate polarization		Party polarization	
Hostile political jokes	-6.560**	-5.869	-6.492*	-4.853	-6.627**	-6.884*
	(2.887)	(3.926)	(3.332)	(4.499)	(2.895)	(3.956)
Republican		-7.037*		-11.99**		-2.084
		(4.086)		(4.683)		(4.118)
Hostile political jokes X Republican		-0.961		-2.602		0.681
		(5.773)		(6.616)		(5.818)
Constant	49.86***	53.00***	51.02***	56.37***	48.69***	49.62***
	(2.039)	(2.729)	(2.353)	(3.128)	(2.044)	(2.750)
Observations	666	666	666	666	666	666
R-squared	0.008	0.018	0.006	0.030	0.008	0.008

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

[3] Models with partisanship interaction were not pre-registered

The interaction terms were not significant. It shows that the experimental effect was not moderated by partisanship. Partisanship alone, however, was significant in general polarization and candidate polarization models at 10 percent and at 5 percent. It means that Republicans respondents were less polarized regardless of the experimental manipulations on average and the difference in polarization between Democrat and Republican respondents was driven by lower level of Republican's candidate polarization. Experimental manipulation's effect completely lost its statistical significance in the models with interaction terms (except from the party polarization model). These additional variables only served to find potential moderations and there is no need for control because of the experimental design. Furthermore, the loss of significance is mainly because of the increased size of standard errors in these models, therefore the treatment's effect can be still considered causational.

Further analysis was conducted to find out whether the non-expected effect of the experimental manipulation is originated from changes in feelings towards the co-partisans, out-partisans or both. In these models, the dependent variables were respondents self-declared feelings toward co-partisans and out-partisans (see Table 5).

These models showed that the decreasing level of affective polarization is clearly origins from the growing sympathy toward out-partisans. Hostile political joke exposure changed feelings toward out-partisan by about 5 points in positive direction, and this effect was significant at 1 percent in the models of general and party-based feeling variables and at 5 percent in the model with the candidate-based feeling variable. Hostile jokes, on the other hand, did not have a significant effect on feelings toward co-partisans in any of the models.

Table 5: OLS regressions, Hostile political joke exposure's effect on feelings toward co-partisans and out-partisans (baseline: non-political joke exposure)[4]

	Feelings towards					
	General		Candidate-based		Party-based	
VARIABLES	Co-partisans	Out-partisans	Co-partisans	Out-partisans	Co-partisans	Out-partisans
Hostile political jokes	-1.287	5.272***	-1.431	5.062**	-1.144	5.483***
	(1.675)	(1.970)	(1.991)	(2.186)	(1.664)	(2.036)
Constant	73.80***	23.94***	72.79***	21.76***	74.82***	26.12***
	(1.183)	(1.391)	(1.406)	(1.544)	(1.175)	(1.437)
Observations	666	666	666	666	666	666
R-squared	0.001	0.011	0.001	0.008	0.001	0.011

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

These results show that the "polarizing joke hypothesis" can be rejected, however, the null hypothesis cannot be accepted. Since the coefficient of interest surpassed the pre-set inference criteria ($p<0.05$), an unexpected causal effect of hostile humor exposure was revealed.

4.2.2 Effects on trust variables

The experimental manipulation's effect on trust variables were analyzed similarly to the polarization effects. The primary point of interest was hostile joke exposure's effect on Trust in politicians, but it was also examined whether the manipulation affect trust in media, judiciary and police. These trust variables did not serve the purpose of hypothesis testing, but they were included to be benchmarks for potential effects on Trust in politicians.

Models including partisanship and its interaction with the experimental manipulation were included. These additional regressions were included to find out whether a partial effect is behind the main effect of the hostile joke exposure. It was necessary, since the results of these analyses were similarly surprising to the test of "polarizing joke hypothesis" (see Table 6). Hostile political joke treatment increased trust in politicians and judiciary to the same extent by around 3.5 points,

[4] These models were not pre-registered.

which was the opposite of the expected direction. However, this cannot be accepted as a causal effect opposing my hypothesis neither, as $p<0.5$ inference criteria was set in the preregistration plan of this research. These results are still notable and although the direction of these effects is contradictory to the expectations, it is coherent with the similarly surprising depolarizing effect of the hostile jokes.

Table 6: OLS regressions, Hostile political joke exposure's effect on trust variables (baseline: non-political joke exposure)

	How much do you trust… ?			
VARIABLES	'politicians'	'judiciary'	'police'	'media'
Hostile political jokes	3.514*	3.688*	2.585	3.544
	-1.929	-1.888	-2.212	-2.199
Constant	31.35***	49.71***	54.04***	39.37***
	-1.368	-1.339	-1.568	-1.559
Observations	796	796	796	796
R-squared	0.004	0.005	0.002	0.003

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

None of the interaction terms were significant showing that these limited effects were not heterogenous based on partisanship. There were some significant differences between respondents of different partisanship. Independent trusted less the judiciary by 8.9 points, and the media by 11.9 points, while Republicans trusted 21.3 points more the police and 19.9 points less the media compared to Democrats. This remarkable difference regarding trust in police could occur probably because the issue of racist police violence have been politicized since the beginning of protests because of the recent death of George Floyd (Epstein 2020). The New York Times reports that 55 percent of Joe Biden supporters back the idea of defunding law enforcement agencies while 66 percent of registered voters opposes the same idea according to newspapers' national poll conducted in cooperation with Sienna College (ibid). The small and insignificant effect size of

hostile jokes on trust in politicians could suggest that social attitudes, which are related to salient issues, are harder to be influenced by political jokes.

4.2.3 Further exploratory analyses

It was also tested how hostile political meme exposure affects political activity. Besides the index of political activity, its components were also included in separate models as dependent variables (see: Table 7). The experimental treatment's coefficient was insignificant in all models. It means that hostile humor exposure was not able to mobilize or demobilize respondents politically in any direction.

Table 7: OLS regressions, Hostile political joke exposure's effect on political activity (baseline: non-political joke exposure)[5]

	Political activity				
VARIABLES	General (index)	Volunteer in campaign	Vote for president	Protest	Debate online
Hostile political jokes	2.505	2.717	1.878	2.677	2.747
	(1.662)	(2.265)	(2.019)	(2.319)	(2.355)
Constant	52.26***	38.60***	79.44***	50.27***	40.76***
	(1.178)	(1.606)	(1.431)	(1.644)	(1.669)
Observations	796	796	796	796	796
R-squared	0.003	0.002	0.001	0.002	0.002

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As an exploratory analysis, it was tested whether political jokes had any direct effect on Apology variables. It aimed to show that hostile jokes may influence respondents' opinion about issued which are the topics of the joke. The hostile humor exposure involved one joke about Vice President Biden's alleged inappropriate behavior with women, one joke about President Trump's

[5] This analysis was not registered, as political activity questions were added to the survey after the submission of preregistration plan.

comments about injecting disinfectant. It was analyzed whether the hostile joke exposure (containing these jokes) affected respondent's opinion about these actions/accusations.

Chi-squared tests were used because the dependent variables are dichotomous (Yes or No). First, two tests were conducted separately for the Trump apology and the Biden apology variables. The ratio of respondents who thought that President Trump should apologize for his comments about injecting disinfectant was 5.6 percentage points higher in the treatment group, this difference was not significant. The difference of ratios between the treatment and baseline group regarding opinions about whether Biden should apologize for hist past behavior towards women was neither significant nor substantial.

Table 8: Chi-squared tests, Hostile political joke exposure's effect on Apology variables

Should Trump apologize… ?				Should Biden apologize… ?			
	Baseline	Hostile jokes	Total		Baseline	Hostile jokes	Total
No	34.09	29.5	31.78	No	23.8	22	22.89
Yes	65.91	70.5	68.22	Yes	76.2	78	77.11
Total	100	100	100	Total	100	100	100
	Pearson chi2(1) = 1.9344		Pr = 0.164		Pearson chi2(1) = 0.3638		Pr = 0.546

Further chi-squared tests were run on Democrat and Republican subsamples of the respondents to find out whether hostile political jokes' potential effects are conditional on partisanship (see Table A5 in the Appendix). It could be very likely that partisan respondents may be already convinced about the moral hazard of the out-party's candidate (therefore there is no need for any humorous argumentation about it). On the other hand, they may be defensive toward their party's candidate by default, and this defensive standpoint may be changed by a morally destructive hostile joke.

The difference between the treatment and the baseline group was indeed larger among Republicans (8.1 percentage points) than among Democrats (4.7 percentage points) regarding the ratio of the

Trump apology variable (see: Table A5 in the Appendix). Similarly, the difference was also larger amongst Democrats (5.1 percentage points) than Republicans (0.8 percentage points) regarding the rations of the Biden apology variable (see: Table A6 in the Appendix). Although, these differences are larger regarding the opinions about the co-partisan candidate in accordance with the argumentation above, these still did not pass any standard inference criteria of significance.

4.3 Partisan humor bias experiment

Most of the humor bias hypotheses were tested with the statistical analysis of the data from the partisan humor bias experiment. In the following, mainly OLS regression models were run clustered by respondent ID. For this purpose the data was restructured the same way as described at the analysis of the pilot study.

4.3.2 Joke target effects

The first joke target hypothesis was tested in a clustered OLS regression, in which the dependent variable was the Joke appreciation index. This variable was used only in this part of the survey to measure the joke target's effect on a more complex evaluation index than funniness. It consists more general emotions (comfortability with the joke) and opinions (validity of the joke). The second joke target hypothesis was tested with a similar clustered OLS regression but using simple perceived funniness as dependent variable.

The analyzed subsample consists the respondents assigned to the treatment group of the humor effect experiment, as only this group evaluated political jokes. Independent respondents are excluded from these analyses, as joke target congruence (the main independent variable) is not applicable to them. Models with partisanship interaction terms were included here too. The results show that congruence of joke target had a substantial and statistically significant effect at 1 percent on both dependent variables in all models (see: Table 9). The appreciation index of the jokes was 15.7 points higher and perceived funniness was 13.1 points higher when the joke was about the out-party compared to the jokes about the in-party. The joke target congruence variable's

coefficient passed the pre-set inference criteria ($p < 0.05$) and its direction was the same as expected. Therefore, both joke target hypotheses can be considered accepted.

Table 9: Clustered OLS regressions, joke target congruence's effect on joke evaluation (baseline: co-partisan joke target)

VARIABLES	Appreciation	Appreciation	Funniness	Funniness
Out-partisan joke target	15.68***	18.40***	13.06***	12.69***
	(1.569)	(2.267)	(1.592)	(2.197)
Republican		1.654		3.286
		(3.207)		(3.443)
Out-partisan joke target X Republican		-5.686*		0.775
		(3.115)		(3.189)
Constant	37.20***	36.41***	35.22***	33.64***
	(1.596)	(2.131)	(1.712)	(2.219)
Observations	1,328	1,328	1,328	1,328
R-squared	0.056	0.059	0.033	0.035

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

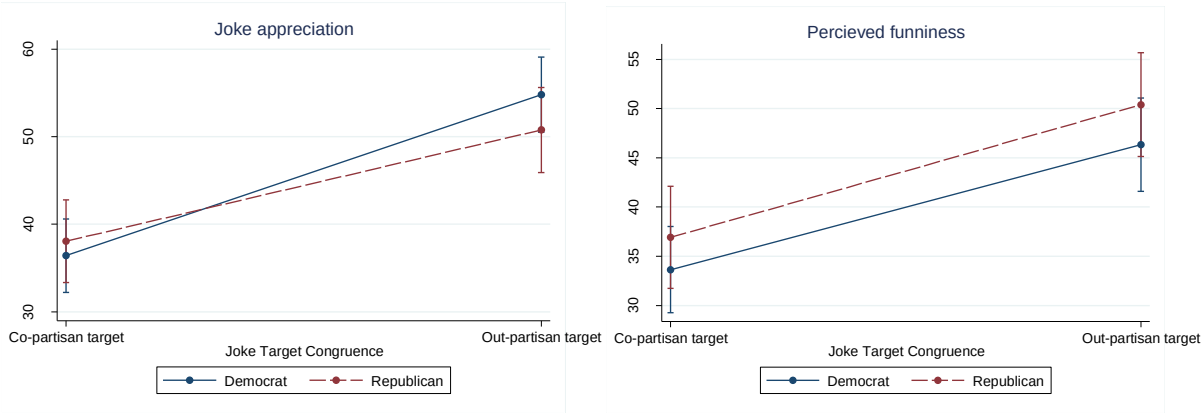

Figure 9: Joke target congruence effect on joke appreciation index, by partisanship (adjusted predictions with 95 percent confidence interval)

The effect on Joke appreciation index was heterogenous, Democrats evaluated jokes about out-partisans 18.4 points funnier than jokes about co-partisans. Meanwhile, the size of the same effect was 12.7 points among Republicans. This heterogeneity roots from that Republicans evaluated slightly higher jokes about their side, and moderately lower jokes about Democratic side (see Figure 9). It is hard to draw conclusions from this significant moderation effect. It can mean that more

emphatic toward the out-partisans than Democrats. Another explanation is that jokes about Republican side used in the survey were simply less hostile and more spot on than jokes about the Democratic side. Taking into consideration that Republicans respondents were on average less polarized (regardless of which group they were assigned to in the humor effect experiment), it makes sense that they are moderately more emphatic toward Democrats.

Another significant moderator of joke target effect on both dependent variables was the strength of partisanship (see Table A7 in the Appendix and Figure 10 bellow)[6]. Respondents with strong partisanship appreciated the jokes about the opposite party 19 points better and found these jokes 15.7 points funnier. These effects were 7.3 and 5.8 points smaller amongst respondents with weak partisanship.

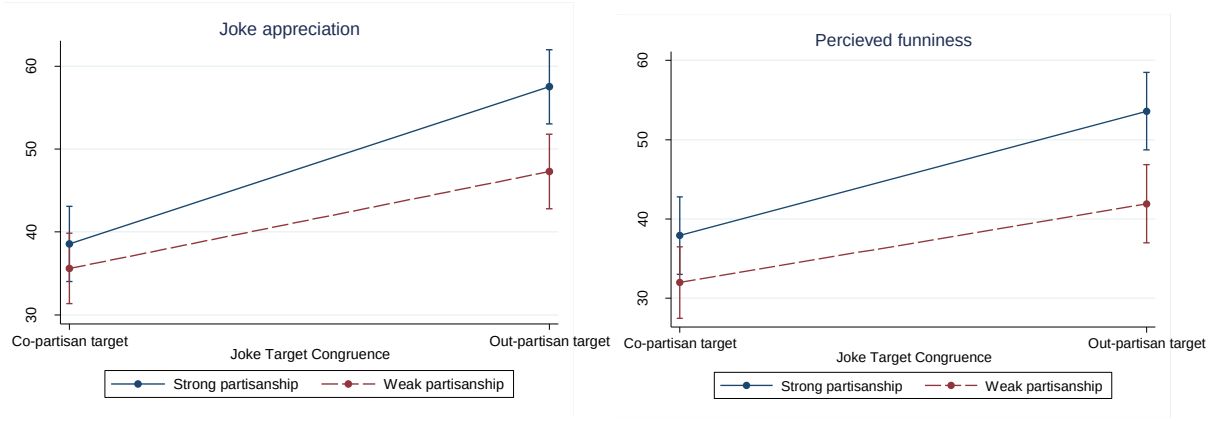

Figure 10: Joke target congruence effect on joke appreciation index, by strength of partisanship (adjusted predictions with 95 percent confidence interval)

4.3.2 Joke source effects

The first joke source effect hypothesis (H4.1, baseline is co-partisan joke source) was tested on the baseline group of the humor effect experiment. These respondents faced four non-political jokes with randomized partisan joke source cues (Republican or Democrat, which was coded as co-partisan or out-partisan for this analysis). The hypothesis was tested with a clustered OLS

[6] These analyses were not pre-registered.

regression, in which the dependent variable is perceived funniness and the independent variable is joke source congruence.

Joke source congruence's coefficient turned out to be insignificant (see Table A8 in the Appendix and Figure 11 bellow). It means that there is no significant difference between the evaluation of jokes with co-partisan and out-partisan source cues. Hence the null hypothesis of the first joke source effect hypothesis (H4.1) was accepted. I also run models interacting joke source congruence with partisanship and partisanship strength[7]. None of the coefficients were significant in these models. Neither partisanship nor strength of partisanship was a moderator of joke source effect.

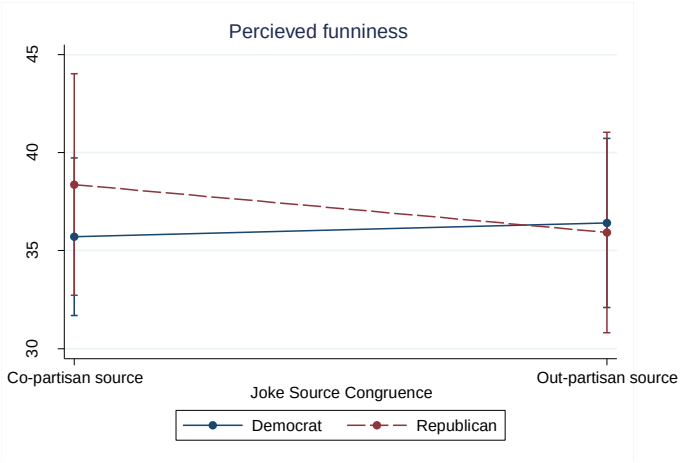

Figure 11: Joke source congruence effect on joke appreciation index, by partisanship (adjusted predictions with 95 percent confidence interval)

4.4 Partisan and gender humor bias experiment

4.4.1 Joke source effects

The remaining partisan source effect hypotheses (H4.2 and H4.3) and the gender source hypothesis (H6) was tested on all respondents as everybody faced two additional non-political jokes with randomized cues after the follow-up questions of the humor effect experiment. These hypotheses were tested with a clustered OLS regression. Perceived funniness is the dependent variable and

[7] These analyses were not pre-registered.

partisan congruence and gender cue is the independent variable. None of the experimental manipulations turned out to be statistically significant (see: Table 10). Therefore, the null hypotheses were accepted. To find out whether there are any partial effects, another regression with partisanship interactions was run. This model showed that partisanship did not moderate joke source effect.

Table 10: Clustered OLS regressions, Joke source partisanship's congruence and joke source's gender effects on perceived funniness (baseline partisanship: Democrat)

VARIABLES	Funniness	Funniness	Funniness	Funniness
Co-partisan source	-0.239	2.138	0.208	3.941
	(2.364)	(3.070)	(2.394)	(3.236)
Out-partisan source	0.901	1.681	1.349	3.484
	(2.311)	(3.068)	(2.340)	(3.148)
Female source	-1.070	-3.293		
	(2.378)	(3.355)		
Opposite gender source			-0.165	0.489
			(2.377)	(3.339)
Baseline joke source	Non-partisan, male source		Non-partisan, congruent gender source	
Republican		2.588		5.273
		(3.483)		(3.679)
Co-partisan source X Republican		-5.417		-8.102*
		(4.804)		(4.820)
Out-partisan source X Republican		-1.756		-4.441
		(4.664)		(4.704)
Female source X Republican		4.220		
		(4.763)		
Opposite gender source X Republican				-1.302
				(4.752)
Constant	59.83***	58.65***	59.38***	56.85***
	(1.724)	(2.277)	(1.843)	(2.577)
Observations	1,332	1,332	1,332	1,332
R-squared	0.001	0.004	0.000	0.004

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

These models were run again with an important modification. In the first two regressions the baseline was the group of respondents' answers which evaluated one of the two non-political jokes

of this experiment with a non-partisan, male joke source. It was in accordance with the gender effect hypothesis which predicted that female joke source has a negative effect on perceived funniness (regardless of the gender of the respondents). However, similar analysis in the pilot showed some notable though not significant differences between males and females (see: Figure 1). Therefore, as an alternative coding of the experimental manipulations, congruent and opponent gender categories were used in the rerun regressions[8] instead of the original male and female joke source categories. It makes sense because maybe physical attraction drives differences between the perceived funniness of a joke with female and male sources and not social status associated with gender. The models with the recalibrated independent variable were not preregistered neither.

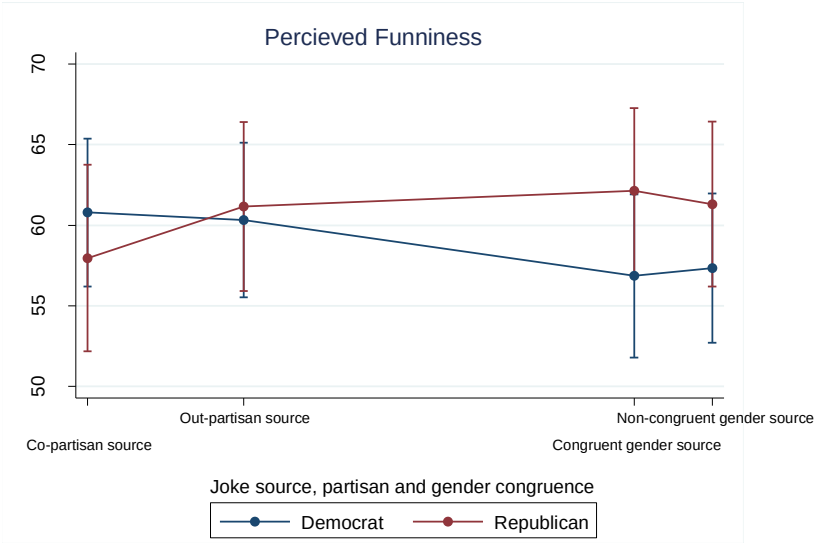

Figure 12: Joke source effect on perceived funniness, (recalibrated joke source variables)
Note: adjusted predictions with 95 percent confidence interval)

The only significant coefficient was the interaction of co-partisan joke source and Republican partisanship, and its direction was opposite as expected. Republicans evaluated the jokes 4.16 points less funny when it came from a co-partisan male source compared to a non-political, congruent gender source (see Figure 12). This partial effect cannot be considered causal either, as it is only significant at 10 percent, therefore it does not pass the preset inference criteria. These

[8] These analyses were not pre-registered.

regressions were run on male and female subgroups of the sample, the coefficients were not significant in those models neither (see Table A9 in the Appendix)[9].

4.4.2 Joke source moderation

Finally, it was tested whether affective polarization moderates joke source effects. In a clustered OLS regression, partisan congruence and gender cue was interacted with affective polarization. The same regression was run with the recalibrated independent variable here too. These regressions were run then on the subsamples of Democrat and Republican respondents. Affective polarization's effect was significant in most models (see: Table 11). Interpreting its effect in the first model, it means that a 10 point increase in affective polarization decreased perceived funniness by 1.4 point on average. Although this effect is still quite small, it shows that individual affective polarization and mood in general could be related to each other.

Only one interaction term was significant (at 10 percent) in the model with the original independent variable analyzing the subset of Democrat respondents. It was the interaction of co-partisan source and affective polarization. but the effect size was marginal. It can be interpreted as a 10 point increase of affective polarization led to 0.01 point higher perceived funniness among respondents who faced the given joke with co-partisan source. In other words, the interaction basically neutralized the otherwise negative effect of affective polarization on perceived funniness.

[9] These analyses were not pre-registered.

Table 11: Clustered OLS regressions, Joke source partisanship's congruence and joke source gender effects on perceived funniness and moderation by affective polarization[10]

VARIABLES	All respondents		Democrats		Republicans	
Co-partisan source	-2.805	-0.537	-5.877	-1.384	-0.641	0.0430
	(3.276)	(3.330)	(5.031)	(5.415)	(4.370)	(4.077)
Out-partisan source	-1.596	0.672	0.291	4.784	-2.896	-2.211
	(3.386)	(3.599)	(5.402)	(5.937)	(4.338)	(4.427)
Female source	-4.205		-8.873		-1.208	
	(3.652)		(5.886)		(4.626)	
Opposite gender source		0.435		1.305		0.0745
		(3.677)		(5.723)		(4.763)
Baseline joke source	Non-partisan male source	Non-partisan, congruent gender source	Non-partisan male source	Non-partisan, congruent gender source	Non-partisan male source	Non-partisan, congruent gender source
Affective polarization	-0.140***	-0.100*	-0.159**	-0.0954	-0.119*	-0.0888
	(0.0496)	(0.0518)	(0.0746)	(0.0787)	(0.0666)	(0.0702)
Co-partisan source X Affective polarization	0.0607	0.0205	0.168*	0.105	-0.0533	-0.0831
	(0.0631)	(0.0628)	(0.0895)	(0.0901)	(0.0919)	(0.0907)
Out-partisan source X Affective polarization	0.0552	0.0150	0.0378	-0.0255	0.0609	0.0311
	(0.0607)	(0.0622)	(0.0946)	(0.0956)	(0.0799)	(0.0849)
Female source X Affective polarization	0.0706		0.122		0.0454	
	(0.0680)		(0.103)		(0.0933)	
Opposite gender source X Affective polarization		-0.0108		-0.0200		-0.0124
		(0.0685)		(0.100)		(0.0948)
Constant	66.22***	63.95***	66.18***	61.68***	66.39***	65.71***
	(2.634)	(2.807)	(4.141)	(4.695)	(3.367)	(3.356)
Observations	1,332	1,332	716	716	616	616
R-squared	0.014	0.013	0.018	0.014	0.023	0.022

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

[10] Models on Republican and Democrat subsamples, and models with recalibrated gender source variable were not pre-registered.

4.4.3 Further exploratory analyses

The last statistical analyses aimed to explore the mechanism of the hostile joke treatment and whether it has behavioral consequences. This exploratory analysis integrates humor effect experiment and humor and gender bias experiment. With clustered OLS regressions and instrumental technique (2SLS regressions), I aimed to find out whether hostile joke exposure (treatment of the Humor effect experiment) leads to stronger joke source effects (effects of the experimental manipulation of the humor and gender bias experiment). The two different models were designed to test whether any potential effect of hostile humor exposure on joke source effect is direct (moderation) or indirect through its effect on affective polarization (mediation).

To test whether hostile humor exposure moderates joke source effect, a clustered regression was run. Perceived funniness of the non.political jokes of the partisan and gender humor bias experiment was the dependent variable. The independent variables were Partisan congruence and gender cue, Humor exposure type (of the humor effect experiment) and their interaction. These regressions were also run on partisan subsamples. In the first model, there was no significant moderation effect, but the hostile humor exposure had a significant and substantial effect alone on perceived funniness (see: Table 12 and Figure 13). It is a further support for the claim that hostile humor exposure's effect may be mediated by a general elevation of mood. The only significant interaction term was between out-partisan joke source and humor exposure in the model of the Republican subsample. This interaction term canceled out the otherwise significant and large effect of hostile humor exposure.

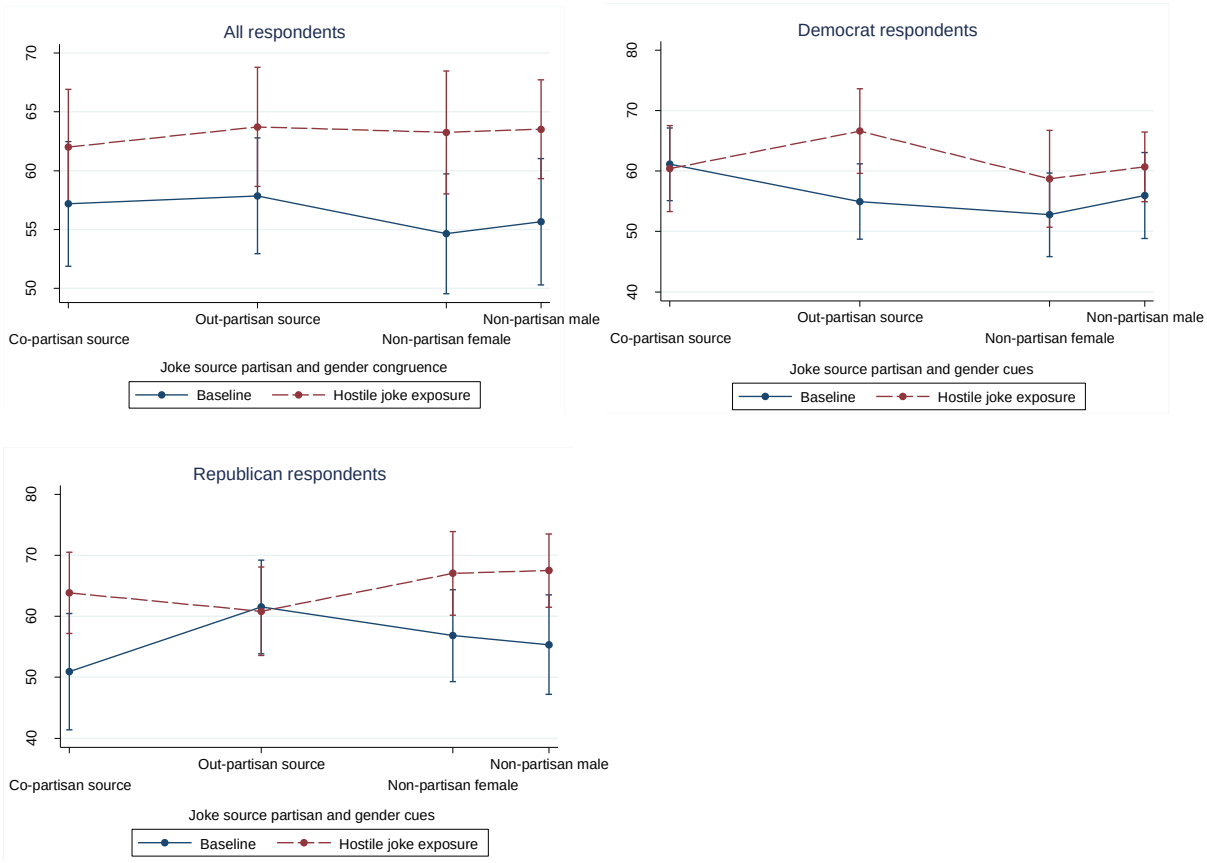

Figure 13: Hostile joke exposure's moderation effect on joke source effect, by partisanship
Note: adjusted predictions with 95 percent confidence interval

Table 12: Clustered OLS regressions, Hostile joke exposure's moderation effect on partisan and gender source effect

	Perceived funniness		
VARIABLES	All	Democrats	Republicans
Co-partisan joke source	1.521	5.170	-4.419
	(3.625)	(4.392)	(6.106)
Out-partisan joke source	2.209	-1.014	6.184
	(3.385)	(4.287)	(5.286)
Female joke source	-1.019	-3.175	1.494
	(3.538)	(4.906)	(5.100)
Hostile humor exposure	7.865**	4.742	12.17**
	(3.473)	(4.644)	(5.153)
Co-partisan joke source X Hostile humor exposure	-3.040	-5.475	0.743
	(4.747)	(6.233)	(7.341)
Out-partisan joke source X Hostile humor exposure	-2.014	6.928	-12.89*
	(4.649)	(6.116)	(7.000)
Female joke source X Hostile humor exposure	0.753	1.193	-1.963
	(4.757)	(6.780)	(6.736)
Constant	55.66***	55.96***	55.35***
	(2.738)	(3.609)	(4.151)
Observations	1,332	716	616
R-squared	0.013	0.017	0.028

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In accordance with the "polarizing joke hypothesis", it was expected originally that the hostile joke exposure would increase affective polarization, which could indirectly influence joke source effect of the partisan and gender bias experiment. Although, this effect turned out to be opposite, humor exposure type can be still a good instrument as it is a statistically significant predictor of affective polarization. Therefore, 2-staged least squared regression models were run. In the first staged, individual level of affective polarization was predicted by Joke exposure type. In the second stage, this predicted variable was interacted with Partisan congruence and gender cue variable.

Hostile political joke exposure was a strong instrument of affective polarization, and the predicted polarization variable had a significant and substantial effect on perceived funniness of the non-political jokes of the humor and gender bias experiment. However, none of the interaction terms were significant which showed that the treatment did not moderate indirectly the joke source effects.

Table 13: 2SLS regression, Hostile joke exposure's indirect effect on funniness (mediated by affective polarization) and indirect moderation of partisan and gender source effect (by instrumented affective polarization)

	First stage	**Second stage**
VARIABLES	DV: Affective Polarization	DV: Funniness
Hostile joke exposure	-6.560***	
	(0.912)	
Co-partisan source		-21.58
		(33.40)
Out-partisan source		-13.10
		(32.95)
Non-partisan female source		4.706
		(33.61)
Predicted affective polarization		-1.199**
		(0.529)
Co-partisan source X Predicted affective polarization		0.463
		(0.724)
Out-partisan source X Predicted affective polarization		0.307
		(0.709)
Non-partisan female source X Predicted affective polarization		-0.115
		(0.725)
Constant	49.86***	115.4***
	(0.644)	(24.30)
Observations	6,660	1,332
R-squared	0.008	0.013

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The significant effect of predicted polarization, on the other hand, show that the treatment may have effected the general level of perceived funniness through affective polarization at first glance. However, it seems a less valid causal model. What seems more plausible is that affective polarization correlates strongly with the general mood of respondents, which lies behind the significance and effect size of the predicted polarization's coefficient. It is an added value of this analysis that these lasts exploratory analysis (see. Table 12 and Table 13) is that affective polarization's correlation with funniness (first shown in Table 11) is not originated from some pre-existing composition effect.

Chapter 5: Discussion of the results

The pilot study showed promising results, but its lessons also pushed toward the redesign of the second survey. Both positive and negative gender biases were revealed toward females amongst Democrat and Republican male respondents accordingly. Partisan joke source effect was not revealed but joke target effect was significant. It was assumed that target effect dominates the evaluation of political jokes, therefore joke source effect was only analyzed regarding non-political jokes in the second survey. The pilot also showed that even soft and positive jokes are considered hostile (except the positive joke about Biden). It turned out that both Democrat and Republican respondents were surprisingly emphatic towards the out-party. Since the jokes created a general perception of hostility, I turned toward the investigation of a symmetrical hostile joke exposure's theorized negative effects in the second survey.

Most of the hypotheses of these thesis were rejected (for an overview of the hypothesis tests, see Table 14.) However, the statistical analysis of the results showed very surprising patterns, which need additional explanations. Only the joke target hypotheses were accepted. The humor effect hypothesis showed theory contradicting results, although only the test of the polarizing joke hypothesis surpassed the pre-set level of inference criteria.

These theory contradicting results deserve special attention because they are very counter-intuitive. Hostile humor exposure, as an experimental treatment was expected to increase polarization and decrease trust in politicians, but the opposite happened. Some of the exploratory analysis suggest that a general mood elevation might be the driver of these effects. Hostile humor exposure also increased the evaluation of the non-political jokes showed afterwards. It is possible that the non-political jokes shown to the baseline group was simply worse at cheering up respondents. Simple t-tests partially supported this explanation, as the average level of perceived funniness of the jokes was significantly higher in the treatment group, but only in the subgroup of strong partisans (see

Table A10 in the Appendix)[11]. Hence, I tested whether perceived funniness of the jokes, as a proxy of respondents' mood influenced polarization and trust.

Table 14: Summary of hypothesis tests (second survey)

Humor effects hypotheses	
(H1) ***Polarizing jokes hypothesis*****:** Hostile political humor exposure leads to higher level of affective polarization	Contradicting effect
(H2) ***Trust eroding jokes hypothesis:*** Hostile political humor exposure leads to lower level of trust in politicians	Rejected (Contradicting effect, but only significant at 10 percent)
Humor bias hypotheses	
Joke target effect hypotheses:	
(H3.1) People appreciate hostile political jokes less, if it is about their preferred political side compared to a joke about the opposite political side	Accepted
(H3.2) People find hostile political jokes less funny, if it is about their preferred political side compared to a joke about the opposite political side	Accepted
Partisan joke source hypotheses:	
(H4.1) People find non-political jokes less funny, if the joke source is out-partisan compared to co-partisan joke source	Rejected
(H4.2) People find non-political jokes less funny, if the joke source is out-partisan compared to non-partisan joke source	Rejected
(H4.3) People find non-political jokes more funny, if the joke source is co-partisan compared to non-partisan joke source	Rejected
(H5) ***Partisan source moderation hypothesis****:* Joke source effect is moderated by affective polarization – the higher individual level of affective polarization is, the stronger partisan joke source effect is.	Rejected
(H6) ***Gender effect hypothesis:*** People find non-political jokes less funny, if the joke source is (non-partisan) female compared to a (non-partisan) male joke source. (gender effect - baseline)	Rejected

[11] These analyses were not pre-registered.

The OLS models showed that the average level of the jokes' perceived funniness indeed decreased significantly and substantially affective polarization and its component indices (see Table A11 in the Appendix)[12]. However, hostile humor exposure effect did not disappear, on the contrary, its magnitude increased to 9.2 points in the affective polarization model and 11.7 in the candidate polarization model. (Due to the increased standard error, the effects were only significant at 10 percent and 5 percent accordingly). It means that mood elevation is very likely to be the part of the causal mechanism. However, filtering out this affective channel, somehow revealed that hostile humor exposure's depolarizing effect is even more important.

Looking for other explanations, it is worth to return to the theorized mechanisms of the polarizing joke effect and trust eroding effect (see Figure1 and Figure 2 in Chapter 2). The joke target effect proves that jokes about co-partisans were more welcomed, especially amongst strong partisans. It means that there was no conceptual problem with the manipulation. The lack of increase of positive partisanship shows that despite mockeries of out-party were evaluated better, these jokes were not strong enough to trigger a feeling of moral superiority or convince the respondents about the negative qualities of the other political camp. Taking a closer look at how the strong and weak partisan subgroups of Republicans and Democrats evaluated these jokes, it makes sense (see Figure 14). Only strong partisans found the jokes about the out-party funny. Even they evaluated these jokes relatively low.

[12] These analyses were not pre-registered.

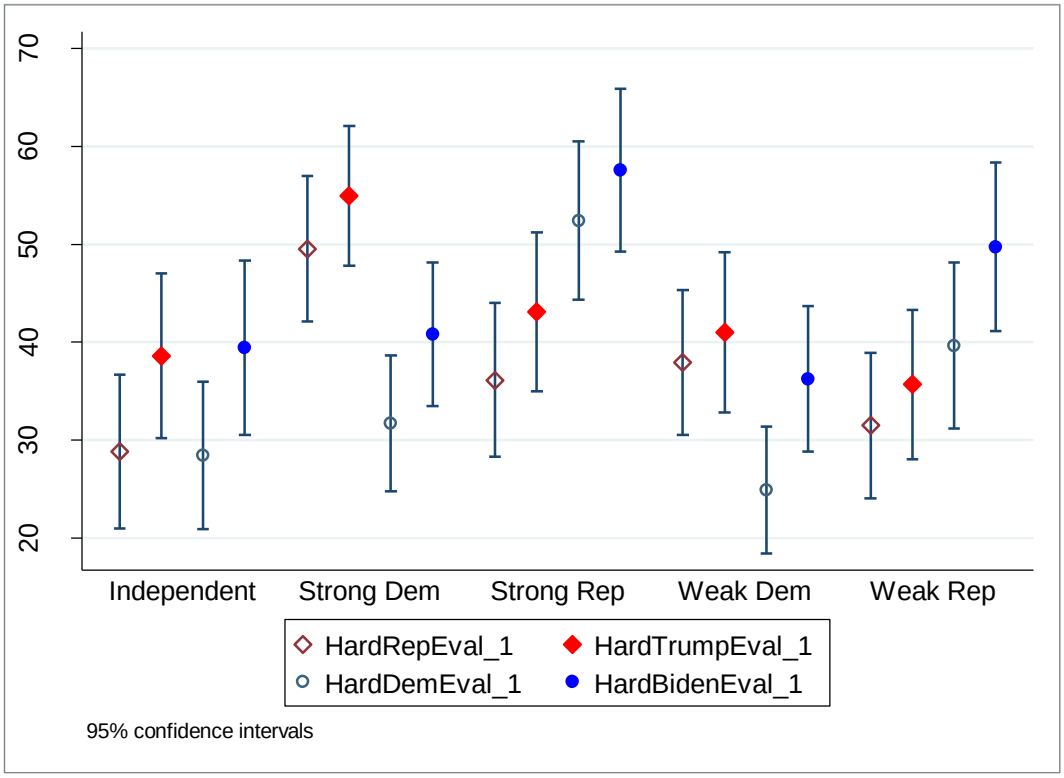

Figure 14: Perceived funniness of partisan jokes, by partisanship, strength of partisanship

However, it only explains lack of increase of positive partisanship, and why trust in politicians did not decrease. It does not explain the increased sympathy toward the out-party and the somewhat increased level of trust in politicians. It is possible that the jokes about the out-party were not evaluated especially well because they were too extreme for the respondents. Sympathy for the out-party may be the result of a counter effect originated from respondents' resistance toward these extreme messages. A study of the 2001 British parliamentary elections showed that negative advertising could result in increased sympathy toward the target of the attack (Sanders and Norris 2005). This counter effect was only revealed amongst non-partisan voters. A meta-analysis of negative campaigning shows that several studies revealed similar backfire effects, but the majority of studies showed that negative campaigns worsen the evaluation of their targets (Lau, Sigelman, and Rovner 2007).

This sympathy-triggering effects of an attack toward the target seems to be very context dependent. Maybe political humor, more precisely user generated hostile political memes are especially likely to stimulate such counter effects. Billboard and social media campaigns may be similar, as they work with similar visual materials with simplistic messages. On the other hand, the literature suggests that smarter and more sophisticated political jokes would work differently. Previous political humor experiments, which showed that critical humor exposure reduced the target's evaluation (Becker 2012, Becker and Haller 2014) and increased polarization (Skoroda 2018), used scenes from satirical political television shows.

Now, it is clearer why pro-attitudinal jokes did not affect polarization. It is less obvious why the humorous attacks on co-partisans did not resulted in an increase of negative attitudes towards the out-party. It is likely that the joke-teller was seen aggressive. However, maybe it did not translate into a general dislike of the other side, as respondents could saw the joke source as an exception, not as the typical figure of the other side. This could be another explanation for the increased sympathy. Seeing the extremes may have triggered a counter effect making the respondents think that "they are better than this".

The symmetrical character of the joke exposure strengthens this explanation. Respondents encountered hostile jokes from their own party as well. It may have led them to the understand that there are extremities on both sides, but average folks are normal across the aisle. Furthermore, receiving hostile about their own party could made them more emphatic toward the out-party, as they realize that both sides experience similar insults.

I originally theorized that symmetrical exposure can affect affective polarization in another symmetrical channel, through increasing the perceived level of polarization. It was based on evidence that media coverage of polarization increases affective polarization (Levendusky & Malhotry 2016a in Iyengar et al. 2019). Looking at the results, this expectation seems to be

ungrounded. It is also possible that the opposite happened. People aim to get rid of their antipathy toward out-partisans consciously when they experience how deep partisan divides are.

The null results of the partisan bias hypothesis are unfortunate for my theory, but good news for American society. It means that polarization did not spoil humor, at least political jokes. Respondents evaluated the funniness of jokes regardless of the partisanship of the joke-teller. One can argue that partisan stickers are not strong enough cues to influence joke evaluation, but it simulates well how people identify the partisanship of their acquaintances on social media platforms. For further research, it would be worth to investigate whether other manipulations of joke sources, such as politicians or partisan outlets, influence the evaluation of jokes.

The experimental design gives the advantage of a relatively high certainty about internal validity. However, the external validity of this research is not guaranteed, especially because original experiments were conducted on an online sample. It would be worth to replicate this study incorporating the lessons of this thesis.

Conclusion

Humor in general and political humor may have many beneficial consequences, such as resolution of conflicts, political education this thesis focused on the adverse effects of the widespread usage of political humor. This thesis investigated those aspects of political humor which have not yet been discussed to my knowledge (or very rarely) in the scholarly literature. It was theorized that (1) hostile political humor exposure would have potential polarizing and trust eroding effects and that (2) there is a partisan bias in the appreciation of jokes. To test these theories, online survey experiments were conducted with US participants. 169 and 759 respondents took my pilot and my second surveys accordingly.

The pilot survey was aimed to test how respondents evaluate the selected hostile, soft critical and positive memes. The results showed that even soft critical and positive jokes were considered moderately hostile. Therefore, I focused on the hypothesized effects of hostile political jokes in the "humor effect experiment" of my second survey. This experiment involved a symmetrical hostile political humor exposure targeting both political sides, as an experimental treatment. It aimed to test how it affects affective polarization, trust in politicians and political activity. Tweet-like vignettes were used containing hostile political memes. Such mockeries are assumed to be an integral part of a politically intense social media environment.

The statistical analysis of the responses led to theory contradicting results. It was hypothesized that hostile political joke exposure would increase affective polarization and reduce trust in politicians, but the opposite happened. Affective polarization was reduced by the treatment, which was driven by the increased sympathy toward the out-party. Trust in politicians increased, although this effect was only significant at 10 percent not passing the pre-set criteria of inference. Looking for explanations, it turned out that the hostile joke exposure potentially elevated mood.

Although mood elevation could play a part, since the average level of perceived funniness also decreased polarization. However, this "mood proxy" control did not neutralize the treatment's depolarizing effect, but it magnified that. There are no straightforward answers, but it is possible that respondents perceived the hostile jokes' sources as the extremes and refused to accept stereotypes. Maybe, the symmetrical feature of the hostile joke exposure raised awareness to the fact that mockeries are hurtful for both sides. Anyway, respondents' resilience to memes which were considered to be polarizing gives reasons for optimism.

The second pillar of this thesis was the test of the partisan humor bias theory. Can partisan appreciate jokes about their own side? Did polarization in the US reached a level where people cannot laugh at funny jokes told by out-partisans? I carried out two experiments in the second survey to answer these questions. I expected that people appreciate a political joke less when it targets their preferred party. Furthermore, it was also expected that in such a highly polarized country as the US, people would appreciate even non-political jokes less when it comes from the opposite party. To test the joke source's effects, I used partisan cues (Republican or Democrat) in the "Tweet vignettes" as experimental manipulations in the "partisan humor bias experiment" of the second survey. Joke source effect was tested in one more experiment. In the "partisan and gender humor bias experiment" of the second survey, a gender manipulation of the joke source was used besides the partisan source cues. This gender manipulation served as a benchmark for the co-partisan/out-partisan joke source effect. Based on psychology literature, it was expected that jokes told by women are less appreciated.

The pilot results showed signs of partial gender bias – Republican males appreciated less, while Democrat males appreciated more non-partisan female joke sources compared to male joke sources. However, this result did not replicate on the much larger sample of the second survey. Only the joke target hypothesis was accepted, but it is less surprising. People appreciated those jokes more which attacked the opposite party. My results imply that humor bias is partial and based

on the defense of the in-party. Co-partisans joke-tellers are not liked better and out-partisans are not discriminated. Non-political humor may be an escape from heated partisan fights.

On one hand, the theories built in this thesis did not get empirical support. On the other hand, both my null results and theory contradicting results give reasons for hope about our political processes and the role of political humor in the current media environment. Getting closer to the 2020 US presidential campaign, the public discourse are going to get more and more intense. Based on the lessons of the last presidential elections, I expect that both sides are eager to use online trolling, sarcastic shows, meme warfare and other forms of political humor as an integral part of their campaign strategy. My results show that these processes are not necessarily as consequential as it is commonly believed and as I expected. However, it is not advisable to draw universal conclusions based on these results because online experiments carry external validity limitations. Nevertheless, this thesis carries some surprising and useful lessons. Political humor may have adverse political and social effects, it can be also an antidote to troubling political processes. It would be worth to analyze these positive effects of humor in future research with experimental methods.

Bibliography

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. 'Exposure to Ideologically Diverse News and Opinion on Facebook'. *Science* 348 (6239): 1130–32.

Baum, Matthew. 2002. 'Sex, Lies and War: How Soft News Brings Foreign Policy to the Inattentive Public'. *American Political Science Review* 96 (3): 91–109.

Baumgartner, Jody, and Jonathan S. Morris. 2006. 'The Daily Show Effect: Candidate Evaluations, Efficacy, and American Youth'. *American Politics Research* 34 (3): 341–67.

Becker, Amy 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, Amy B., and Beth A. Haller. 2014. 'When Political Comedy Turns Personal: Humor Types, Audience Evaluations, and Attitudes'. *Howard Journal of Communications* 25 (1): 34–55.

Benton, Gregor. 1988. 'The Origins of the Political Joke'. In *Humour in Society: Resistance and Control*, edited by Chris Powell and George E. C. Paton, 33–55. London: Palgrave Macmillan UK.

Bowyer, Benjamin T, Joseph E Kahne, and Ellen Middaugh. 2017. 'Youth Comprehension of Political Messages in YouTube Videos'. *New Media & Society* 19 (4): 522–41.

Cammaerts, Bart, Michael Bruter, Shakuntala Banaji, Sarah Harrison, and Nick Anstead. 2014. 'The Myth of Youth Apathy: Young Europeans' Critical Attitudes Toward Democratic Life'. *American Behavioral Scientist* 58 (5): 645–64.

Campbell, David. 1992. 'Writing Security (Chapter 6)'. In *Writing Security: United States Foreign Policy and the Politics of Identity*. Minneapolis: University of Minnesota Press.

Chen, Hsuan-Ting, Chen Gan, and Ping Sun. 2017. 'How Does Political Satire Influence Political Participation? Examining the Role of Counter- and Pro-Attitudinal Exposure, Anger, and Personal Issue Importance'. *International Journal of Communication* 11 (0): 19.

Dassonneville, Ruth, and Marc Hooghe. 2017. 'Voter Turnout Decline and Stratification: Quasi-Experimental and Comparative Evidence of a Growing Educational Gap'. *Politics* 37 (2): 184–200.

Encyclopaedia Britannica. 2020. 'Boris Johnson. Biography, Facts, & Role in Brexit'. In *Encyclopaedia Britannica*. https://www.britannica.com/biography/Boris-Johnson (accessed May 30, 2020).

Epstein, Reid J. 2020. 'These Top Democrats Go Further Than Biden on Diverting Police Funds'. *The New York Times*, 26 June 2020, sec. U.S. https://www.nytimes.com/2020/06/26/us/politics/defund-police-protests-democrats.html (accessed June 30, 2020).

Galles, Gary M. 2018. 'Will Rogers on Government: Remembering Cowboy Political Philosophy'. FEE.Org. 4 November 2018. https://fee.org/articles/will-rogers-on-government-remembering-cowboy-political-philosophy/ (accessed July 19, 2020).

Garrett, R. Kelly, Shira Dvir Gvirsman, Benjamin K. Johnson, Yariv Tsfati, Rachel Neo, and Aysenur Dal. 2014. 'Implications of Pro- and Counterattitudinal Information Exposure for Affective Polarization'. *Human Communication Research* 40 (3): 309–32.

Heintz, Sonja, Willibald Ruch, Tracey Platt, Dandan Pang, Hugo Carretero-Dios, Alberto Dionigi, Catalina Argüello Gutiérrez, et al. 2018. 'Psychometric Comparisons of Benevolent and Corrective Humor across 22 Countries: The Virtue Gap in Humor Goes International'. *Frontiers in Psychology* 9.

Holbert, R. Lance, Jennifer L. Lambe, Anthony D. Dudo, and Kristin A. Carlton. 2007. 'Primacy Effects of The Daily Show and National TV News Viewing: Young Viewers, Political Gratifications, and Internal Political Self-Efficacy'. *Journal of Broadcasting & Electronic Media* 51 (1): 20–38.

Huddy, Leonie, Alexa Bankert, and Caitlin Davies. 2018. 'Expressive Versus Instrumental Partisanship in Multiparty European Systems'. *Political Psychology* 39 (S1): 173–99.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. 2019. 'The Origins and Consequences of Affective Polarization in the United States'. *Annual Review of Political Science* 22 (1): 129–46.

Lau, Richard R., Lee Sigelman, and Ivy Brown Rovner. 2007. 'The Effects of Negative Political Campaigns: A Meta-Analytic Reassessment'. *Journal of Politics* 69 (4): 1176–1209.

Levendusky, Matthew S. 2013. 'Why Do Partisan Media Polarize Viewers?' *American Journal of Political Science* 57 (3): 611–23.

Mann, Sheilah. 1999. 'What the Survey of American College Freshmen Tells Us about Their Interest in Politics and Political Science'. *PS: Political Science and Politics* 32 (2): 263–68.

Martin, Rod A. 2007a. 'Introduction to the Psychology of Humor'. In *The Psychology of Humor: An Integrative Approach*. Burlington, MA: Elsevier Academic Press.

———. 2007b. 'The Cognitive Psychology of Humor'. In *The Psychology of Humor: An Integrative Approach*. Burlington, MA: Elsevier Academic Press.

———. 2007c. 'The Social Psychology of Humor'. In *The Psychology of Humor: An Integrative Approach*. Burlington, MA: Elsevier Academic Press.

Mattson, Kevin. 2003. *Engaging Youth: Combating the Apathy of Young Americans toward Politics*. New York: The Century Foundation Press.

McGrane, Sally. 2010. 'Icelander's Campaign Is a Joke, Until He's Elected'. *The New York Times*, 25 June 2010. https://www.nytimes.com/2010/06/26/world/europe/26iceland.html (accessed May 5, 2019).

Mencimer, Stephanie, and Mark Helenowski. 2019. '"The Left Can't Meme": How Right-Wing Groups Are Training the next Generation of Social Media Warriors'. *Mother Jones*, 2 April 2019, sec. Politics. https://www.motherjones.com/politics/2019/04/right-wing-groups-are-training-young-conservatives-to-win-the-next-meme-war/ (accessed July 20, 2020).

Plevriti, Vasiliki. 2014. 'Satirical User-Generated Memes as an Effective Source of Political Criticism, Extending Debate and Enhancing Civic Engagement'. MA Global Media and Communication, The University of Warwick.

Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Ray, Michael. 2020. 'Volodymyr Zelensky. Biography, Facts, & Presidency'. In *Encyclopaedia Britannica*. https://www.britannica.com/biography/Volodymyr-Zelensky (accessed May 19, 2020).

Roberts, Steven V., and Special To the New York Times. 1987. 'WASHINGTON TALK; Reagan and the Russians: The Joke's on Them'. *The New York Times*, 21 August 1987, sec. U.S. https://www.nytimes.com/1987/08/21/us/washington-talk-reagan-and-the-russians-the-joke-s-on-them.html (accessed July 20, 2020).

Ruch, Willibald, Sonja Heintz, Tracey Platt, Lisa Wagner, and René T. Proyer. 2018. 'Broadening Humor: Comic Styles Differentially Tap into Temperament, Character, and Ability'. *Frontiers in Psychology* 9.

Sanders, David, and Pippa Norris. 2005. 'The Impact of Political Advertising in the 2001 U.K. General Election'. *Political Research Quarterly* 58 (4): 525–36.

Sandvig, Christian. 2015. 'The Facebook "It's Not Our Fault" Study'. *Multicast, Harvard.Blog.Edu* (blog). 7 May 2015. http://blogs.harvard.edu/niftyc/archives/1062 (accessed April 20, 2020).

Skoroda, Jeffrey. 2018. 'Real Fake News: The Colbert Report and Affective Polarization'. *Senior Independent Study Theses*, no. Paper 8283 (January). https://openworks.wooster.edu/independentstudy/8283 (accesses May 20, 2020).

Szászi, Áron József. 2020. 'Political Humor, Partisan Humor Bias and Polarizing Effects: An Experimental Study in the United States (OSF Preregistration Plan)'. Center for Open Science. https://osf.io/a5v8u (accessed June 20, 2020).

Tenbarge, Kat. 2020. 'Bernie Sanders Is Winning 2020 Meme Race, but Will It Get Him Elected? - Business Insider'. *Business Insider*, 21 February 2020. https://www.businessinsider.com/bernie-sanders-democratic-frontrunner-memes-meme-economy-once-again-asking-2020-2 (accessed July 15, 2020).

Tosun, Sümeyra, Nafiseh Faghihi, and Jyotsna Vaid. 2018. 'Is an Ideal Sense of Humor Gendered? A Cross-National Study'. *Frontiers in Psychology* 9 (February).

Urban Dictionary. 2018. 'The Left Can't Meme'. Urban Dictionary. 2018. https://www.urbandictionary.com/define.php?term=The%20Left%20can%27t%20Meme (accessed July 20, 2020).

Williams, Anthony. 2016. 'Biden and Obama Jokes on Trump Imagined'. *BBC News*, 14 November 2016, sec. US Election 2016. https://www.bbc.com/news/election-us-2016-37972626 (accessed July 20, 2020).

Xenos, Michael A., and Amy B. Becker. 2009. 'Moments of Zen: Effects of The Daily Show on Information Seeking and Political Learning'. *Political Communication* 26 (3): 317–32.

Young, Dannagal G. 2018. 'Theories and Effects of Political Humorr: Discounting Cues, Gateways, and the Impact of Incongruities'. In *The Oxford Handbook of Political Communication*. Oxford University Press.

———. 2020. *Irony and Outrage: The Polarized Landscape of Rage, Fear, and Laughter in the United States*. Oxford University Press.

Young, Dannagal G., Benjamin E. Bagozzi, Abigail Goldring, Shannon Poulsen, and Erin Drouin. 2019. 'Psychology, Political Ideology, and Humor Appreciation: Why Is Satire so Liberal?' *Psychology of Popular Media Culture* 8 (2): 134–47.

Zannettou, Savvas, Tristan Caulfield, Jeremy Blackburn, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Guillermo Suarez-Tangil. 2018. 'On the Origins of Memes by Means of Fringe Web Communities'. *ArXiv:1805.12512 [Cs]*, May. http://arxiv.org/abs/1805.12512 (accessed July 10, 2020).

Appendix

The full surveys, the used vignettes and analyzed data are uploaded to the online Appendix, available at the following link:

https://drive.google.com/drive/folders/1AXZFkkJlKwSV2d43Ge13kbTTtw_CIox7?usp=sharing

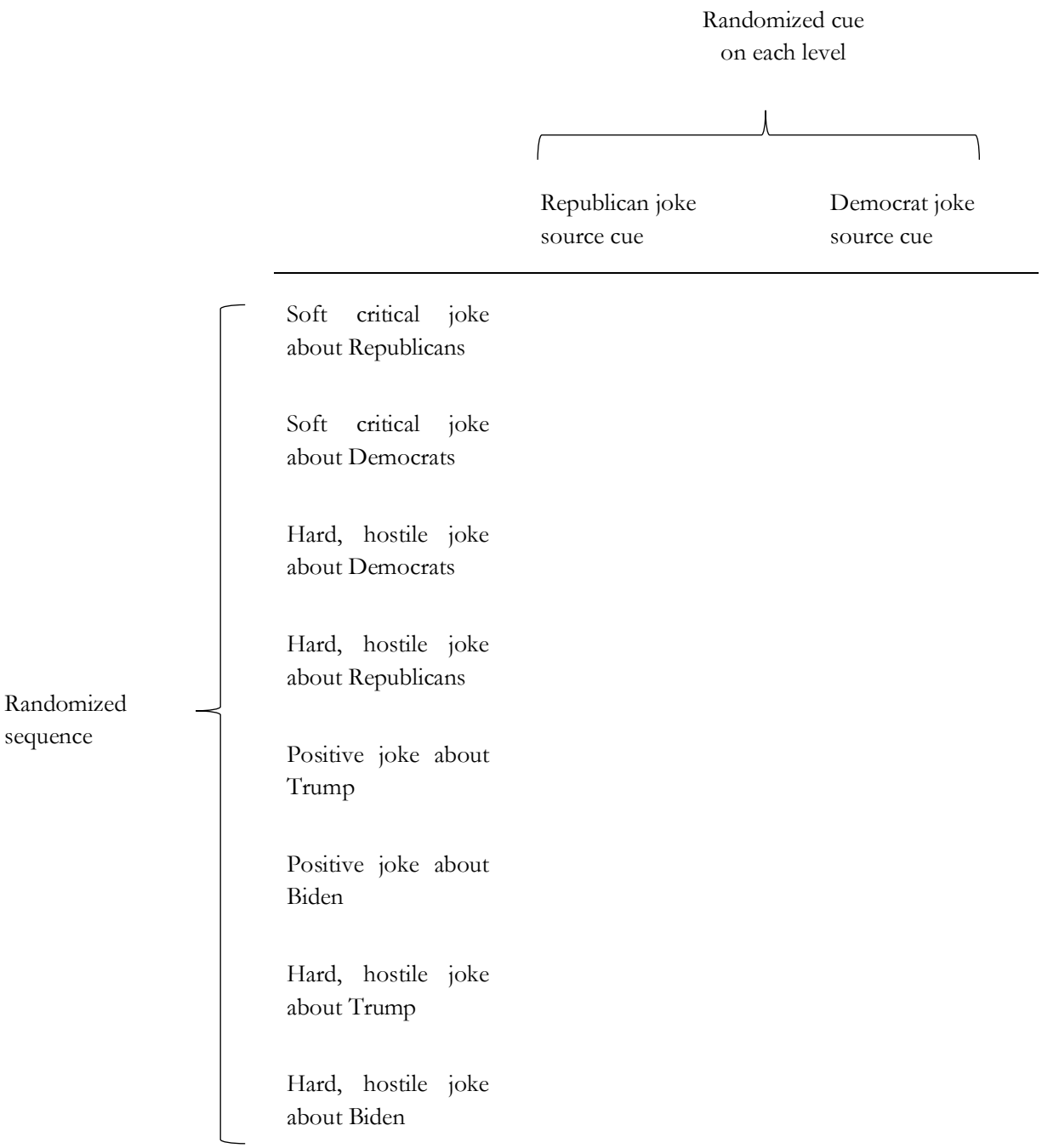

Figure A1: Research design of the second part of the pilot survey

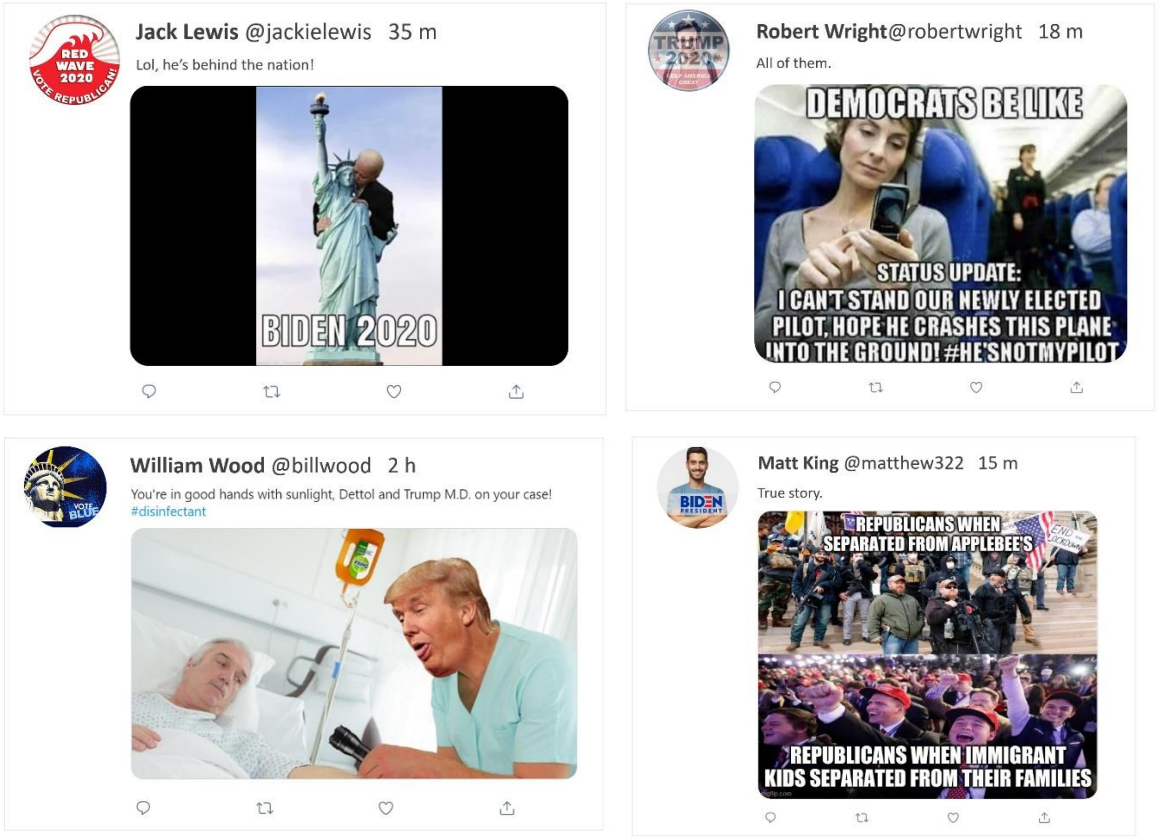

Figure A2: Vignettes of the hostile political joke exposure (2nd survey, humor effect experiment, treatment group)

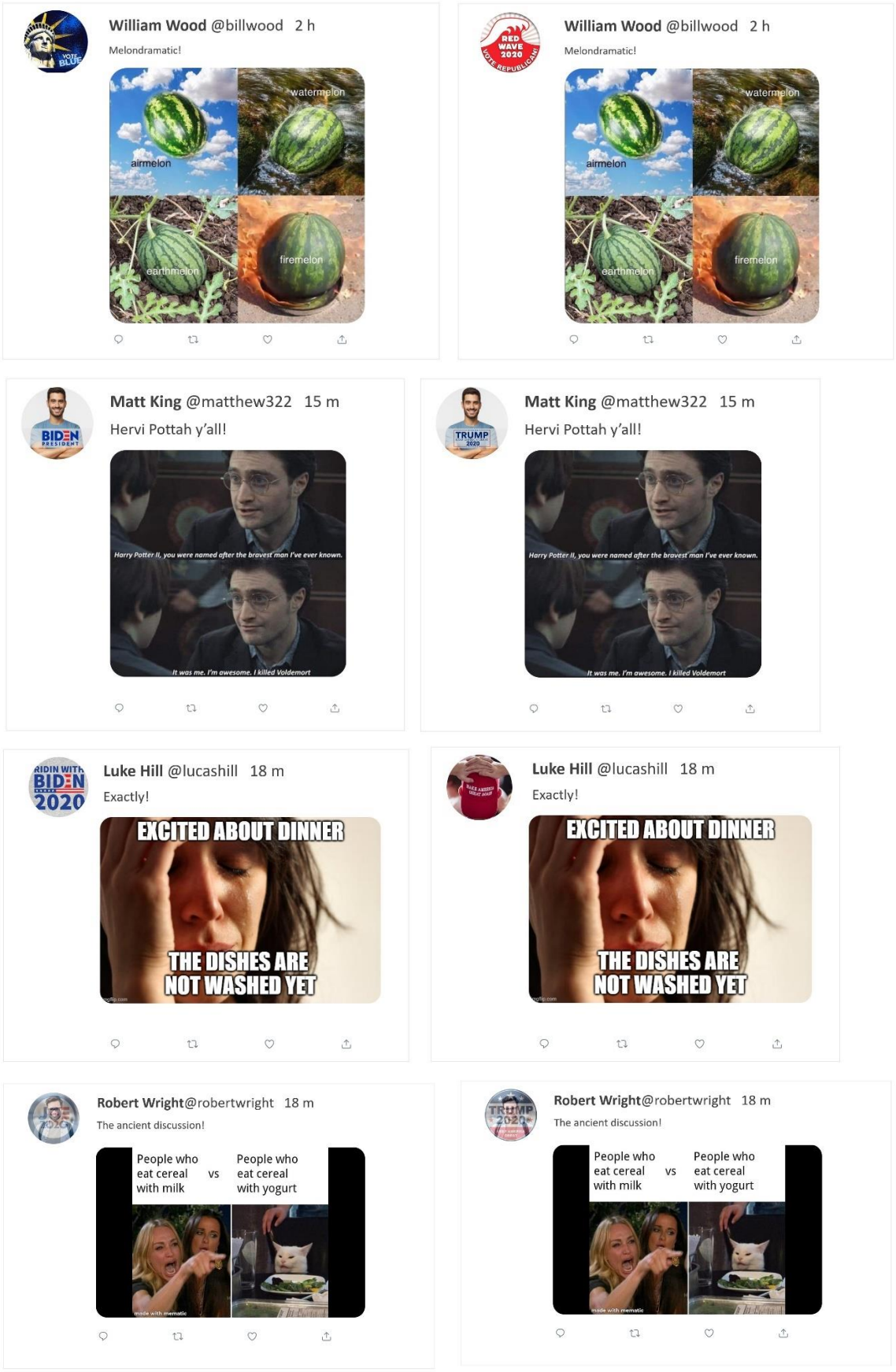

Figure A3: Vignettes of the non-political joke exposure, with different partisan joke source cues (2nd survey, humor effect experiment, baseline group)

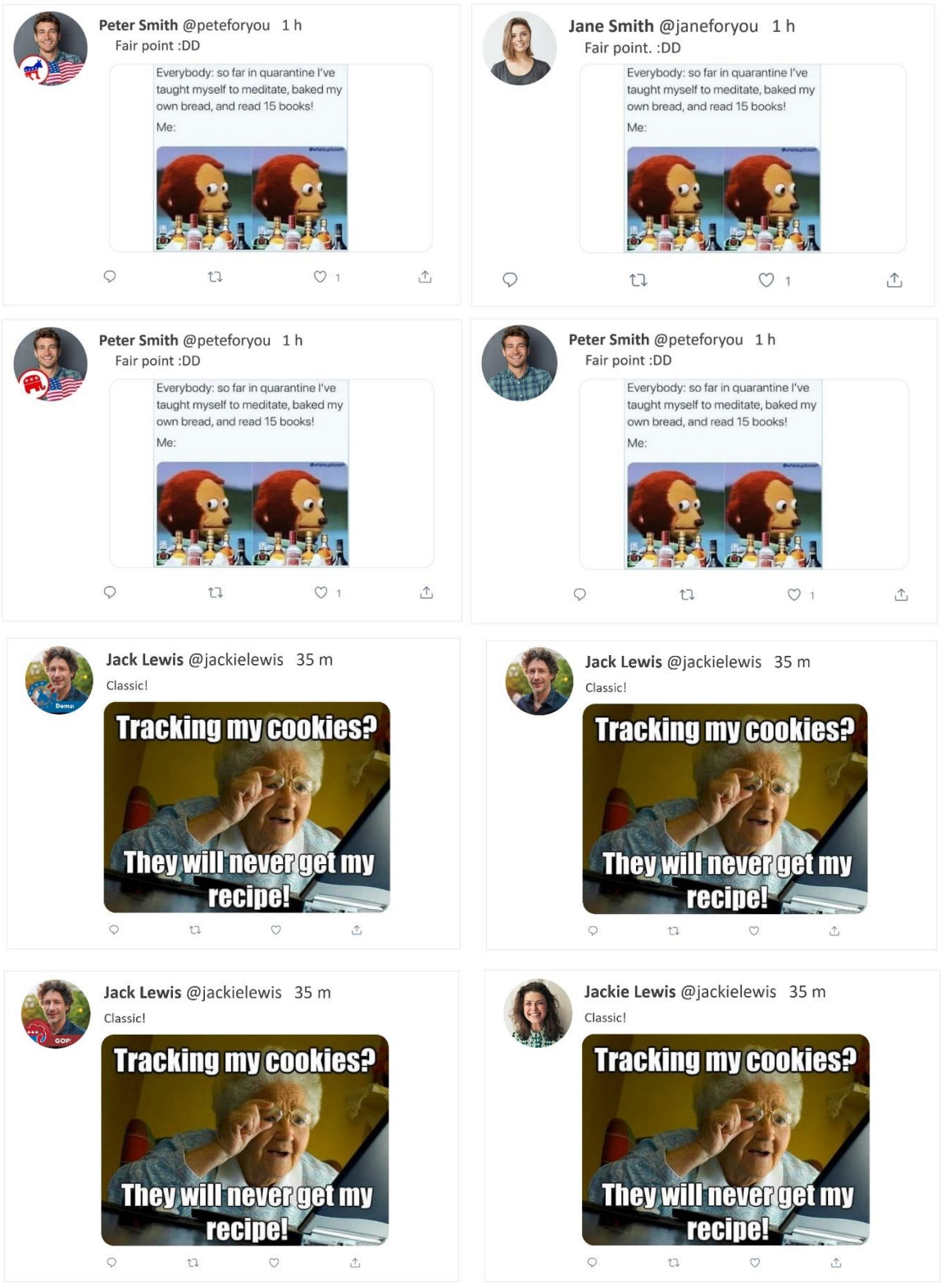

Figure A4: Vignettes of the non-political joke exposure, with different partisan and gender joke source cues (partisan and gender humor bias experiment)

Table A1: OLS regressions, partisan and gender joke source effects on the appreciation index of non-political jokes, by gender subgroups (pilot study, baseline: non-political male joke source)

	All respondents		Male respondents		Female respondents	
VARIABLES						
Co-partisan source	-5.392	-6.520	0.425	4.028	-13.14	-18
	(6.035)	(9.048)	(8.277)	(12.63)	(9.206)	(12.85)
Non partisan female source	-0.360	0.0610	1.592	15.64	-3.694	-12.40
	(6.115)	(8.620)	(8.504)	(12.63)	(9.206)	(11.73)
Out-partisan source	2.004	5.472	3.968	8.188	-1.721	2.653
	(5.962)	(8.814)	(8.003)	(13.00)	(9.323)	(11.94)
Republican		-2.528		1.375		-5.104
		(8.814)		(11.87)		(13.88)
RepublicanXCo-partisan source		2.085		-6.426		9.917
		(12.23)		(16.65)		(18.48)
RepublicanXNon partisan female source		-1.777		-27.82		24.65
		(12.44)		(17.06)		(18.92)
RepublicanXOut-partisan source		-6.693		-6.579		-13.49
		(12.06)		(16.44)		(19.05)
Constant	65.47***	66.84***	64.58***	63.75***	67.75***	69.94***
	(4.354)	(6.494)	(5.853)	(9.191)	(6.904)	(9.088)
Observations	149	149	81	81	67	67
R-squared	0.012	0.023	0.004	0.069	0.040	0.110

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2: OLS regressions, Partisan and gender joke source effects on perceived funniness of political jokes, by gender subgroups (pilot study, baseline: non-political male joke source)

VARIABLES	All respondents		Male respondents		Female respondents	
Co-partisan source	0.679	1.154	5.850	8.056	-7.341	-5.625
	(7.537)	(11.32)	(10.22)	(15.16)	(11.61)	(16.23)
Non partisan female source	-6.817	-1.185	-2.178	27.50*	-13.84	-24.67
	(7.636)	(10.79)	(10.50)	(15.16)	(11.61)	(14.81)
Out-partisan source	4.975	10.68	9.383	19.38	-3.021	2.205
	(7.446)	(11.03)	(9.881)	(15.60)	(11.76)	(15.08)
Republican		5.572		16.50		-3.417
		(11.03)		(14.24)		(17.53)
RepublicanXCo-partisan source		-0.959		-2.510		-2.308
		(15.30)		(19.99)		(23.33)
RepublicanXNon partisan female source		-11.83		-56.06***		31.50
		(15.57)		(20.48)		(23.88)
RepublicanXOut-partisan source		-10.67		-16.64		-15.54
		(15.09)		(19.74)		(24.05)
Constant	54.40***	51.38***	53.40***	43.50***	57.79***	59.25***
	(5.438)	(8.126)	(7.226)	(11.03)	(8.707)	(11.47)
Observations	149	149	81	81	67	67
R-squared	0.018	0.025	0.021	0.134	0.026	0.095

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A3: OLS regressions, hostile joke exposure's effects, mediated by strength of partisanship (baselines: non-political joke exposure, strong partisanship, strong Democrat)

VARIABLES	'General polarization'		'Candidate polarization'		'Party polarization'	
Hostile joke exposure	-8.228**	-11.78**	-8.334*	-11.57*	-8.122**	-11.99**
	(3.755)	(5.113)	(4.430)	(5.989)	(3.711)	(5.085)
Weak partisanship	-24.24***		-21.43***		-27.05***	
	(3.894)		(4.594)		(3.849)	
Strong Republican		-12.08**		-18.73***		-5.427
		(5.340)		(6.255)		(5.311)
Weak Democrat		-29.01***		-27.82***		-30.21***
		(5.203)		(6.094)		(5.174)
Weak Republican		-30.39***		-32.23***		-28.55***
		(5.527)		(6.474)		(5.497)
Interactions						
Hostile joke exposure X Weak partisanship	2.856		3.330		2.382	
	(5.521)		(6.514)		(5.457)	
Hostile joke exposure X Strong Republican		8.296		8.186		8.406
		(7.490)		(8.773)		(7.450)
Hostile joke exposure X Weak Democrat		12.29		14.05		10.53
		(7.488)		(8.771)		(7.448)
Hostile joke exposure X Weak Republican		-0.150		-1.640		1.341
		(7.844)		(9.188)		(7.802)
Constant	61.25***	66.64***	61.10***	69.45***	61.41***	63.83***
	(2.670)	(3.568)	(3.150)	(4.179)	(2.639)	(3.548)
Observations	666	666	666	666	666	666
R-squared	0.101	0.116	0.058	0.088	0.127	0.131

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A4: OLS regressions, Hostile political joke exposure's effect on trust variables, partisanship interaction included (baseline: non-political joke exposure)

	How much do you trust… ?			
VARIABLES	'politicians'	'judiciary'	'police'	'media'
Hostile political jokes	4.563	2.617	1.651	5.165*
	-2.871	-2.779	-3.1	-3.133
Independent	2.39	-8.945**	-6.097	-11.89***
	-3.983	-3.856	-4.301	-4.347
Republican	4.283	0.795	21.33***	-19.88***
	-2.988	-2.892	-3.226	-3.26
Hostile political jokes X Independent	-7.468	-2.885	4.36	-4.164
	-5.564	-5.386	-6.008	-6.072
Hostile political jokes X Republicans	0.245	4.187	-0.45	-0.833
	-4.221	-4.086	-4.558	-4.607
Constant	29.37***	50.82***	46.97***	48.71***
	-1.995	-1.932	-2.155	-2.178
Observations	796	796	796	796
R-squared	0.015	0.037	0.123	0.096

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A5: Chi-squared tests on subsamples, by partisanship, Hostile political joke exposure's effect on Trump apology variable

Republican respondents				Democrat respondents			
Should Trump apologize…?	Baseline	Hostile jokes	Total	Should Trump apologize…?	Baseline	Hostile jokes	Total
No	59.06	50.94	54.87	No	15.68	10.98	13.41
Yes	40.94	49.06	45.13	Yes	84.32	89.02	86.59
Total	100	100	100	Total	100	100	100
	Pearson chi2(1) = 2.0466		Pr = 0.153		Pearson chi2(1) = 1.6959		Pr = 0.193

Table A6: Chi-squared tests on subsamples, by partisanship, Hostile political joke exposure's effect on Biden apology variable

Republican respondents				Democrat respondents			
Should Biden apologize…?	Baseline	Hostile jokes	Total	Should Biden apologize…?	Baseline	Hostile jokes	Total
No	16.78	17.61	17.21	No	28.8	23.7	26.33
Yes	83.22	82.39	82.79	Yes	71.2	76.3	73.67
Total	100	100	100	Total	100	100	100
	Pearson chi2(1) =0.0373		Pr = 0.847		Pearson chi2(1) = 1.1979		Pr = 0.274

Table A7: Clustered OLS regressions, joke target congruence's effect on political jokes' evaluation, mediated by strength of partisanship (Baseline: Co-partisan joke target)

VARIABLES	Appreciation	Appreciation	Funniness	Funniness
Out-partisan target	18.98***	21.70***	15.68***	15.97***
	(2.154)	(3.103)	(2.240)	(3.070)
Weak partisanship	-2.950		-5.916*	
	(3.159)		(3.372)	
Strong Republican		-0.0557		3.345
		(4.642)		(4.983)
Weak Democrat		-4.673		-5.705
		(4.255)		(4.402)
Weak Republican		-1.066		-2.695
		(4.217)		(4.653)
Weak partisanship X Out-partisan target	-7.266**		-5.755*	
	(3.116)		(3.155)	
Strong Republican X Out-partisan target		-5.582		-0.598
		(4.287)		(4.492)
Weak Democrat X Out-partisan target		-7.127		-7.092
		(4.515)		(4.356)
Weak Republican X Out-partisan target		-13.20***		-4.867
		(4.323)		(4.435)
Constant	38.54***	38.57***	37.91***	36.28***
	(2.306)	(2.946)	(2.478)	(3.154)
Observations	1,328	1,328	1,328	1,328
R-squared	0.069	0.072	0.049	0.051

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A8: Clustered OLS regressions, joke source congruence's effect on non-political jokes' evaluation (Baseline: co-partisan joke target)

VARIABLES	Funniness	Funniness	Funniness	Funniness
Out-partisan source	-0.694	0.710	-1.101	0.663
	(1.774)	(2.222)	(2.358)	(3.007)
Republican		2.662		
		(3.530)		
Weak partisanship			-0.644	
			(3.421)	
Strong Republican				6.242
				(5.061)
Weak Democrat				2.671
				(4.084)
Weak Republican				1.450
				(4.937)
Interactions				
Republican X Out-partisan source		-3.149		
		(3.629)		
Weak partisanship X Out-partisan source			0.858	
			(3.571)	
Strong Republican X Out-partisan source				-4.005
				(4.830)
Weak Democrat X Out-partisan source				0.177
				(4.467)
Weak Republican X Out-partisan source				-2.274
				(5.309)
Constant	36.90***	35.71***	37.21***	34.44***
	(1.713)	(2.049)	(2.462)	(2.923)
Observations	1,336	1,336	1,336	1,336
R-squared	0.000	0.001	0.000	0.004

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A9: Clustered OLS regressions, partisan and gender joke source effects on non-political joke evaluation, by gender subsamples

VARIABLES	Male respondents		Female respondents	
Co-partisan male	-1.824	1.402	1.290	2.776
	(3.405)	(4.725)	(3.286)	(4.031)
Out-partisan male	0.718	1.967	1.105	1.663
	(3.223)	(4.365)	(3.317)	(4.218)
Non-partisan female	-1.205	-2.979	-0.926	-3.545
	(3.482)	(5.040)	(3.243)	(4.529)
Republican		1.196		4.044
		(5.075)		(4.818)
Co-partisan male X Republican		-5.957		-3.780
		(6.776)		(6.958)
Out-partisan male X Republican		-2.293		-1.309
		(6.388)		(6.817)
Non-partisan female X Republican		3.120		5.436
		(6.987)		(6.415)
Constant	60.13***	59.51***	59.51***	57.94***
	(2.533)	(3.547)	(2.340)	(2.969)
Observations	664	664	668	668
R-squared	0.001	0.004	0.001	0.008

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A10: Two-sample t-tests with equal variances, difference between the average funniness of the jokes in the baseline and the treatment group of humor effect experiment

All respondents

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
baseline	396	37.57071	1.340358	26.67278	34.93558	40.20583
treatment	400	40.39875	1.422797	28.45594	37.60163	43.19587
combined	796	38.99183	0.978334	27.6022	37.07141	40.91226
diff		-2.828043	1.95535		-6.66631	1.010224

diff = mean(baseline) - mean(treatment) t = -1.4463
Ho: diff = 0 degrees of freedom = 794

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0742 Pr(T > t) = 0.1485 Pr(T > t) = 0.9258

Subgroup of strong partisans

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
baseline	177	36.64124	2.118349	28.1828	32.46061	40.82188
treatment	181	45.74724	2.212359	29.76425	41.38174	50.11273
combined	358	41.24511	1.549045	29.30931	38.19871	44.29151
diff		-9.105995	3.064867		-15.13352	-3.078474

diff = mean(baseline) - mean(treatment) t = -2.9711
Ho: diff = 0 degrees of freedom = 356

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0016 Pr(T > t) = 0.0032 Pr(T > t) = 0.9984

Subgroup of weak partisans

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
baseline	157	36.44108	1.942044	24.33374	32.60499	40.27718
treatment	151	36.95364	2.132112	26.19983	32.74079	41.16649
combined	308	36.69237	1.437376	25.22584	33.86402	39.52072
diff		-0.51256	2.879842		-6.179359	5.15424

diff = mean(baseline) - mean(treatment) t = -0.178
Ho: diff = 0 degrees of freedom = 306

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.4294 Pr(T > t) = 0.8589 Pr(T > t) = 0.5706

Table A11: OLS regressions, the effect of perceived funniness of jokes on polarization

VARIABLES	'General polarization'		'Candidate polarization'		'Party polarization'	
Average level of perceived funniness	-0.288***	-0.341***	-0.328***	-0.415***	-0.247***	-0.268***
	(0.0514)	(0.0757)	(0.0593)	(0.0873)	(0.0519)	(0.0764)
Hostile joke exposure		-9.628*		-11.69**		-7.560
		(4.927)		(5.684)		(4.975)
Average level of perceived funniness		0.116		0.176		0.0557
X Hostile joke exposure		(0.103)		(0.119)		(0.104)
Constant	57.85***	62.33***	60.64***	66.18***	55.06***	58.48***
	(2.460)	(3.412)	(2.838)	(3.936)	(2.483)	(3.445)
Observations	666	666	666	666	666	666
R-squared	0.045	0.051	0.044	0.050	0.033	0.039

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A12: OLS regressions, the effect of perceived funniness of jokes about co-partisans and out-partisans on subcomponents of polarization

VARIABLES	Feelings towards General		Feelings towards Candidate-based		Feelings towards Party-based	
	Co-partisans	Out-partisans	Co-partisans	Out-partisans	Co-partisans	Out-partisans
Average perceived level of funniness						
Jokes on co-partisans	-0.377***	0.301***	-0.460***	0.288***	-0.294***	0.314***
	(0.0874)	(0.0961)	(0.106)	(0.111)	(0.0888)	(0.101)
Jokes on out-partisans	0.113**	-0.276***	0.147**	-0.295***	0.0797	-0.257***
	(0.0506)	(0.0557)	(0.0616)	(0.0642)	(0.0514)	(0.0582)
Jokes on co-partisans X	0.00423***	0.00389***	0.00471***	0.00435***	0.00375***	0.00342**
Jokes on out-partisans	(0.00118)	(0.00130)	(0.00144)	(0.00150)	(0.00120)	(0.00136)
Constant	70.58***	23.00***	69.63***	20.89***	71.53***	25.11***
	(2.404)	(2.642)	(2.926)	(3.048)	(2.441)	(2.765)
Observations	332	332	332	332	332	332
R-squared	0.113	0.321	0.105	0.276	0.082	0.287

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1