4 Methods and Approaches to the Study of Emotion Expressions

4.1 Determining an Appropriate Research Method

The aim of this book is to analyze the impact that politicians' emotional expressions – particularly those of anger and indignation – have on the evaluation of politicians. This research endeavor depends on a causal claim, whereby the expression of anger precedes the evaluation of the politician. Quantitative approaches to the study of public opinion and attitude formation typically apply representative surveys and laboratory experiments (Frank et al. 2015: 21). Traditional observational studies and cross-sectional survey research, however, are not well suited to untangle cause and effects. This holds true in general as well as for the topic of candidate evaluation, as it has been previously noted:

> "It is hardly possible to decide whether a voter likes the politician because of his policy positions and party affiliation or whether the voter likes the politician because of the politician's character traits." (Huber 2014: 41)

Not much survey research has been conducted regarding politicians' expressions of emotions and the affective dispositions towards politicians. In the run-up to the German general election in 2013, the German Longitudinal Election Study (GLES) administered a pre-election survey in which participants were asked how strongly their positive and negative feelings were towards the two main contestants: Chancellor Merkel and Peer Steinbrück (Rattinger et al. 2017). The aim was to establish a measurement of ambivalence towards both contestants that could explain candidate preferences and candidate evaluations (Blumenstiel & Gavras 2015).[18]

In addition, the GLES online-tracking study has previously and consecutively asked participants in six panel waves, how often Angela Merkel and her (potential) contestants at the time – Frank-Walter Steinmeier, Peer Steinbrück, Sigmar Gabriel, and Martin Schulz – made them feel angry or

---

18 The items have since been regularly used in intermediate inquiry surveys (Roßteutscher et al. 2016a; Roßteutscher et al. 2016b); however, they have not been administered in the pre-election survey in 2017 (Roßteutscher et al. 2018).

enthusiastic. Participants could answer both items for each politician using a 7-point Likert scale ranging from "not at all" to "very much".[19] Since 1980, the American National Election Study (ANES) has asked participants how they describe their affective state towards the presidential candidates in terms of four discrete emotions: angry, afraid, hopeful or proud.[20] If participants indicate they have ever felt this way because of the presidential candidates, a follow-up question measures the frequency of those feelings. This measure of affect has been used to explain support for presidential candidates (Finn & Glaser 2010). Both surveys attempt to measure the emotional disposition towards candidates, especially during an election season, however, the measurements have severe drawbacks. It remains unclear which events participants recall when answering these survey questions. This problem is also noted by Redlawsk and Pierce in regard to the American National Election Study:

> "This suggests that questions such as 'has [candidate] ever made you feel angry?' – the typical survey question that informs AIT – are problematic; at best there will be a significant error in recall." (Redlawsk & Pierce 2017: 417)

Figure 3 shows the distribution of feelings of anger and enthusiasm towards Angela Merkel in six annual cross-sectional surveys from September 2010 to September 2017.[21] The figure shows the cumulative percentages of the participants who often felt very angry and angry or very enthusiastic and enthusiastic towards Angela Merkel. It can be seen that anger towards

---

19 GLES online-tracking studies t1 (2009) (Rattinger et al. 2011), t11 (2010) (Rattinger et al. 2015a), t15 (2011) (Rattinger et al. 2015b), t18 (2012) (Rattinger et al. 2014), t25 (2014) (Rattinger et al. 2015c), t29 (2015) (Roßteutscher et al. 2016a), t33 (Roßteutscher et al. 2016b), and t37 (Roßteutscher et al. 2017).
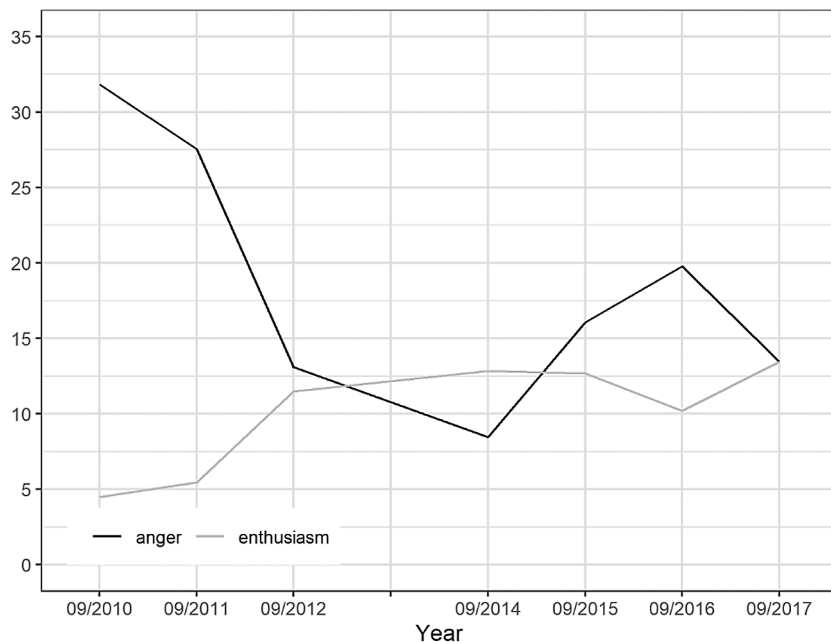
20 The question wording is reported as follows: "Now we would like to know something about the feelings you have toward [the democratic presidential candidate]. Has [the democratic presidential candidate] – because of the kind of person he is, or because of something he has done – made you feel: angry/ afraid/ hopeful/ proud?" (The American National Election Studies 2010).

21 The first GLES online tracking study t1 was conducted in April and May 2009 for some general tests, so the question was only administered to a subsample of the study resulting in missing data for three-quarters of the participants. In addition, the question wording has been changed since the first test. In this general test, participants could answer the question with a scale ranging from "never" feeling angry to "always" feeling angry. Due to the changes in question wording and the testing phase of the first online tracking study, the respective data have been omitted here.

102

her was high in 2010 and 2011 and has increased again since 2015. It is unclear as to why such a large percentage of participants felt angry towards Angela Merkel in 2010 and 2011. A sampling bias might have potentially caused higher values, since the sampling procedure changed in 2012, when the initial online access panel was replaced by a higher quality online access panel with offline recruiting.[22] Questions about feelings towards running candidates were not included in the online tracking study that was conducted during the election campaign season in September 2013. Since 2015, an increase in self-reported anger towards Merkel can be observed, which also corresponds to her decreasing popularity since August 2015 up until spring 2017 (see Figure 1). This trend is likely to be connected to the migration crisis, the emergence of Pegida, and a growing support for the AfD in the general population. The migration crisis has sparked a right-wing populism that has often targeted Merkel personally and deemed her policy decisions to be failures. Taking this longitudinal perspective, some insights can be drawn from the repetition of both questions in several surveys.

---

22 More information about the online access panels can be obtained from the code-books for each study, a short description is provided online at: http://gles.eu/wordpress/english/design/langfrist-online-tracking/ (last accessed: 05 June 2019).

Figure 3: Self-Reported Feelings Towards Angela Merkel

Similar to these self-reported feelings towards politicians, asking partici-pants about the perception of politicians' emotional expressions in ques-tionnaires is likely to suffer from problems regarding the recall of events. Such a measurement has not been implemented in any large-scale election study. One exception is a panel survey that was conducted before and shortly after the Iowa Caucus 2016 (Redlawsk et al. 2018). Caucus goers were asked whether they had ever heard or seen the political candidates ex-press visually or aurally four discrete emotions: hope, fear, anger, and con-tempt (Redlawsk et al. 2018: 177). Due to the preceding exposure to the candidates during a significant event, participants' recalls of events were

less problematic. This design is similar to a quasi-experimental design; a simple cross-sectional survey, however, can only indicate associations instead of causal mechanisms.

In the realm of observational studies, one might also investigate responses to real-life events via traditional content analysis of media outlets or more direct content analysis of social media and comment sections of online newspapers that follow any noticeable public outburst. These observations, however, can only be conducted as case studies that strongly depend on the specific context. Such an approach might complement any experimental approach to gain further insight, but it is not without drawbacks. Only a small number of the electorate actively engage on social media sites or online newspaper comments in the field of politics. According to a question from the GLES online tracking study in 2015 (Roßteutscher et al. 2016a), roughly 15 percent of the German internet population engaged in political debates online, either by writing or commenting on political posts on Facebook, Twitter and other social media sites.[23] Hence, any content analysis could only be interpreted as a reaction and potential evaluation of this small and potentially non-representative proportion of the general population.

In the light of the research question, an experiment that addresses how emotional expressions of anger affect the evaluation of politicians appears to be the most appropriate research method to study the subject systematically within the general population. Although cognitive attributions and affective responses were outlined as potential underlying causal mechanisms, the main aim of this endeavor is not to test which mechanism is at work, but rather to determine how emotional expressions affect the evaluation of a politician, while acknowledging the potential impact of both cognitive as well as affective responses.

Before it is outlined how the experiment was designed to test the impact of emotional expressions, designing an experiment begins with the careful consideration of any pitfalls; hence, the next Subchapter 4.2 starts with some general thoughts on the experimental method, highlighting the strength and potential drawbacks of experimental designs. The implement-

---

23 Different weights were applied to obtain this value: without weights (15.2 percent), weights adjusted to online users with speeders (15.2 percent), adjusted to online users without speeders (15.5 percent), adjusted to the micro census with speeders (14.0 percent), and without speeders (14.4 percent). Author's own calculations based on the GLES long-term online tracking study T29, 2015 (Roßteutscher et al. 2016a).

ed experimental design of this study is then thoroughly discussed, considering potential threats to its validity in the following Subchapter 4.3.

4.2 The Experimental Method

Experiments have been used in the social sciences and political science since the 1920s (Morton & Williams 2010: 3; McDermott 2002: 50). One of the first experiments in political science dealt with voting behavior, particularly voter turnout. In his study "An Experiment in the Stimulation of Voting", Gosnell (1926) sent out postcards with various contents in several randomly selected districts, for example cartoons displaying non-voters as "slackers", to measure the effect of those stimuli on the voter turnout in the following election.

Regardless of these early advances, experiments have only played a marginal role in political science up until the beginning of the 21st century (McDermott 2002). This was largely due to the high cost of experimental research and its limited feasibility in practical and ethical considerations in various subfields of political science in which random assignments could be considered unethical or infeasible. However, a continuous increase in experimental articles can be observed in the social sciences, beginning more slowly in the 1970s and amplifying in the 1990s (Morton & Williams 2010: 4). It was not until the turn of the millennium that experiments gained wider popularity as methodological tools in political science (McDermott 2002; Keuschnigg & Wolbring 2015). Especially in the last decade, studies that are built on experimental designs have increased in number. Further indications of this recent growth and academic interest in experimental studies stem from an increasing amount of literature on experimental method in political science, for example the "Cambridge Handbook of Experimental Political Science" (Druckman et al. 2011), and the launch of the "Journal of Experimental Political Science" in 2014.

The increase in experimental applications can be observed across all political science disciplines such as international relations, political economy, political psychology as well as public opinion and the social sciences more broadly. Experimental studies are frequently described as a gold standard for establishing causality across the sciences and serve "as a yardstick against which non-experimental research is assessed" (Bryman 2012: 50). In the natural sciences, such as physics and chemistry, as well as psychology, the experimental method is taught to scholars (McDermott 2002) and con-

sidered to be a leading methodological tool for generating new insights and creating knowledge (Lupia 2002: 319).

Due to lacking feasibility, or financial, and ethical constraints, observational studies might use several techniques to approximate an experimental design and the missing treatment or control group by using instrumental variables or statistical estimates of a counterfactual outcome (e.g., Abadie et al. 2010, Abadie et al. 2015, Horiuchi & Mayerson 2015).

True experimental designs can also vary largely across the subfields. Classic laboratory experiments manipulate one factor (or several factors systematically) – the stimulus or treatment – which is then administered randomly to participants of the experimental group in a controlled setting, while participants of the control group receive only a placebo treatment or none at all. If more than one factor is manipulated, it is desirable to test all possible factor combinations (e.g., a 2x2 design). Comparative politics might apply quasi-experimental designs in the absence of randomization. Field experiments are used to study the effects of canvassing on public opinion and voter intentions in the US (Gerber & Green 2000). Dictator games in their varieties, such as trust games, are commonly applied within the field of economic cooperation, social psychology, and international relations (e.g., Burns 2006). Studies that try to measure latent constructs, such as xenophobic attitudes, might consider implicit association tests as an experimental design (Greenwald et al. 1998). Implicit association tests that measure latent social attitudes can now be easily designed by computer software and later set up in laboratory or online settings (Greenwald et al. 1998).

One of the main reasons for the increase in experimental studies is the progress that has been made in computational technology (Morton & Williams 2010: 11). Creating and conducting experimental games in a laboratory setting can be time-consuming. Due to the advent of computer programs such as z-Tree (Zurich Toolbox for Readymade Economic Experiments), new software enables researchers to program economic experiments easily (Fischbacher 2007; for an overview of experimental games see Levitt & List 2007: 155). Hence, it has become less time-consuming and cost-intensive to run experiments.

The variations of experimental designs are manifold – offline and online, especially with regard to the possibilities of survey experiments (Mutz 2011). Survey experiments can incorporate an experimental treatment into the questionnaire and come in various forms and shapes (Gaines et al. 2007), such as list experiments which are suited to measure topics strongly suffering from social desirability biases, for example voter turnout, racial

attitudes, and drug usage (Rosenfeld et al. 2016). Vignette studies are also considered to be online survey experiments and have a wide application in market research and sociology (Wallander 2009). They have also been applied to measure political attitudes by embedding a treatment as frame, often as text messages (Goodwin et al. 2020; Hainmueller & Hiscox 2010). In addition, if they aim to set the participants in a certain mood or elicit emotions, other survey experiments might include scenario and induction techniques, such as emotion induction (Searles & Mattes 2015). In the case of political science, such emotion induction experiments have previously included campaign advertisements (Brader 2005; Searles & Mattes 2015). In this context, some studies have only changed additional wording as experimental treatment that generated the frames in which the negative campaign ads were presented (Mattes & Redlawsk 2015). Hence, experimental manipulations can be implemented in manifold ways. Previous studies have tested effects of candidate appearance and demeanor with varying tools, for example the use of photos on flyers or the ballot (Johns & Shephard 2011; Dumitrescu 2010), morphed faces (for an overview see: Todorov et al. 2015), altered voices (Tigue et al. 2012), video clips of various lengths (Masters et al. 1986; Stewart & Ford Dowe 2013; Stewart et al. 2009a; Dumitrescu et al. 2015). Especially the study of emotional expressions has been studied previously with video clips of politicians in a traditional laboratory setting (e.g., Masters et al. 1986; Stewart & Ford Dowe 2013); by doing so, the variance on terms of political party affiliation is often limited to predispositions of college students. In this regard, large-scale online samples offer an advantage by providing samples that are more heterogeneous.

These survey experiments are often embedded in online surveys and present a stark contrast to psychological and neurological experiments that use fMRI scans or physiological measurements, and are occasionally conducted in the field of political psychology to study underlying causal mechanisms (e.g., Schreiber et al. 2013; Hibbing et al. 2013). Since the aim of this research is to gauge the effect of emotional expressions on candidate evaluation, a population-based survey experiment is the most appropriate to measure the evaluation of politicians based on a heterogeneous sample. Online survey experiments are particularly useful if the experimental treatment consists of pictures or video clips that cannot be administered during a telephone interview. Furthermore, they are well-suited to minimize interviewer effects and maximize cost effectiveness compared to face-to-face and telephone surveys.

108

Experimental studies usually have a high internal validity (Bryman 2012: 50). However, their external validity is often limited, especially when they are merely based on student samples which are not representative of the larger population. When findings are solely based on students who grew up or live in "Western, Educated, Industrialized, Rich, and Democratic (WEIRD)" countries (Henrich et al. 2010: 61), the generalizability of their decisions made in artificial experimental settings is limited, because their decisions cannot be seen as representative of those from other living conditions (Henrich et al. 2010). The limited representativeness is amplified, especially when experimental findings are replicated merely with further student samples, and the findings are then considered as reliable evidence about universal human behavior (Heinrich et al. 2010). Hence, the external validity is typically a big concern in experimental studies and the lack of it has contributed to the current replication crisis in psychology (e.g., Shrout & Rodgers 2018). Nonetheless, the use of student samples can also be justified by interpreting the approach as theoretical sampling for specific questions (Druckman & Kam 2011): If the relationship can be found for this particular group in a given situation, it will likely apply to all other social groups, in the sense of a least-likely case sampling approach (e.g., Flyvbjerg 2006; King et al. 1994: 209).

One way to minimize the amount of student samples and to generate variation among the convenience samples is to approach other easily accessible groups, e.g., such as university employees. This specific group is accustomed to research studies and therefore potentially more likely to participate. For example, Stewart and Ford Dowe (2013) recruited 180 university employees of which 80 completed the study (Stewart & Ford Dowe 2013: 381).

Convenience samples can be easily recruited online through the use of the internet. Participants can be contacted via Facebook advertising, social media, Amazon Mechanical Turk (MTurk), or via university email lists (Buhrmester et al. 2011; Kapp et al. 2013; Ramo & Prochaska 2012). The application of experiments also benefits heavily from technological advances; computer software has been developed, such as Lime Survey and Unipark (QuestBack), which facilitates designing the online questionnaire, administering experimental treatments randomly and hosting the survey online. They are especially useful for pre-testing with student or convenience samples as well as for studies with limited budgets. Due to the cost-effectiveness, numerous American experimental research studies have recruited participants via MTurk (Buhrmester et al. 2011; Bohannon 2016). These convenience samples differ only slightly from the American internet

population regarding their overall characteristics; they seem to be younger, have lower income, less children, and are more likely to be female (Berinsky et al. 2012: 357). These concerns also apply to online access panels that frequently invite the same participants. This raises one concern regarding the familiarity of respondents with experimental designs, as they become more and more likely to have previously participated in experiments, even several times, and thereby might differ significantly from the average person who is rarely exposed to the logic of experimental designs. They can become experts in experimental studies, be aware of their participation in either a control or experimental group and thereby purposively try to resist any impact an experimental treatment might have. Such behavior caused by pre-testing could be described as a "reactive effect" (Bryman 2012: 55; Campbell & Stanley 1971: 179).

While some market research companies state they limit the amount of invitations they send to their panel members to discourage "professional usage", market research companies rarely share detailed information about their online access panels publicly. Hence, the concern about expert users remains with any form of opt-in panels which potentially enable heavy usage.

Furthermore, self-selection into studies is a problem in any study, if the respondents differ significantly from those who do not respond. However, this concern is not limited to online surveys and experiments but also poses an increasing issue with traditional telephone surveys, since the number of people who respond to survey invitations has decreased steadily (Curtin et al. 2005; Mutz 2011: 157). Experimental studies based on convenience samples that are recruited via opt-in panels, student samples, or university employees, often lead to valid results as a recent study has shown (Mullinix et al. 2015). By replicating studies using a population-based sample and a range of varying convenience samples, researchers have found that convenience samples do not automatically lead to false inferences (Mullinix et al. 2015: 111; see also Bieber & Bytzek 2012).[24] Therefore, empirical studies that focus on causal mechanisms and the internal validity can be conducted online with convenience samples at low cost and time expenditure. Such convenience samples might sometimes be the only option to gain some insight into the study of emotion research and emotional responses, for example to test the underlying causal assumptions of emotional contagion in studies that aim to elicit emotional responses within

---

24 With an exception of exit poll recruitments, however, further research is needed to explain the diverging effects for exit poll samples (Mullinix et al. 2015: 116).

the participants. Highly incentivized studies might be suited to test effects of emotional contagion by recording facial reactions of viewers or tracking eye movements. For such intrusive research as filming participants on camera, it might not be possible to draw from a representative study, but it could be possible to highly incentivize a small convenience sample.[25] Hence, to study certain topics and theories, (online) convenience samples might be the only way to examine underlying causal mechanisms – such as affective responses or appraisal – systematically.

Large population-based studies, when possible, are preferred because they allow estimating the effect sizes for specific sub-populations and enabling the analysis of heterogeneous treatment effects by providing variance in large samples (Mullinix et al. 2015: 111). This advantage has also been pointed out by Mutz (2011): "It may be well established that X causes Y from laboratory studies, but population-based experiments can tell us that X causes Y with certain kinds of people in certain kinds of situations" (Mutz 2011: 159).

Population-based survey experiments have recently found a widespread application within the social sciences, as they combine the advantages of both experimental and survey research (Mutz 2011), and are easily conducted on the internet. Based on online panels, population-based survey experiments implement an experimental component into a representative or at least large-scale survey. By administering large-scale surveys that incorporate an experimental treatment, the internal as well as the external validity of this approach are potentially high. While the potential benefits for survey experiments are high, researchers need to consider potential pitfalls: "Population-based experiments are not a panacea for all that ails observational and experimental methods" (Mutz 2011: 155). The limitations of experimental research designs need to be considered carefully (Gaines et al. 2007; McDermott 2002).

Several challenges occur when planning and designing an online survey experiment and a researcher needs to evaluate whether and how an online survey experiment might be conducted. Mutz outlines at least two crucial challenges when designing survey-experiments (Mutz 2011: 156–157). First, experimental treatments need to be effective in regard to the causal mechanism that is under scrutiny. In the context of public opinion re-

---

25 This is also common practice in market research companies in neuro marketing studies which test the affective impact of advertisements and communication strategies with physiological measurements and eye-tracking studies (e.g., Hamelin et al. 2017).

search, this implies that researchers need to design experiments that have a measurable impact on participants' attitudes. However, survey experiments also need to adhere to ethical standards when emotions are induced. This presents the second challenge, as researchers can not merely aim for the potentially largest effects, if those treatments present any harm to the participants. For example, in emotion induction techniques, researchers have to think critically about the pictures and video clips they present to participants when they aim to induce emotions such as fear or anger. Next to the potential harm that might result from showing disturbing pictures or putting participants in stressful situations, further ethical considerations have to be made regarding informed consent, privacy considerations, and deception (Bryman 2012: 129–155; McDermott 2002).

While it is technically feasible and cost-effective to administer pictures or video clips as experimental treatments in online survey experiments, the experimental design of stimulus and control conditions need careful consideration. When using video treatments, a placebo treatment is ideally administered for the control group. Based on the experimental paradigm, only the emotional expressions of politicians should differ between the treatment and the control group, while all other factors are equal across experimental groups. To estimate the effect of negative emotions such as anger, the emotional expression is best compared to a neutral expression of the same politician in the same contextual setting, conveying the same content except for the politician's verbal, paraverbal or non-verbal emotional expressions. However, this is hardly feasible with real world data and could only be approximated by using video clips taken from similar public appearances which lack any specific emotional expression. For this reason, previous research has often relied on hired actors who played putative politicians (e.g., Dumitrescu et al. 2015; Mutz & Reeves 2005; Mutz 2007).

Building on the potential pitfalls in experimental designs described by Campbell & Stanley (1971), Bryman (2012) identifies seven threats to the internal validity of an experiment as crucial: *history, testing, instrumentation, mortality, maturation, selection, and ambiguity about causality* (Bryman 2012: 52–53, see also: Behi & Nolan 1996; Campbell & Stanley 1971; Rubin & Babbie 2008). Bryman (2012) also points out the possible drawback of *reactive effects* which occur every time participants are aware of the fact that they are taking part in an experiment (Bryman 2012: 54). The results and external validity of the study could be affected if participants behave differently because of their knowledge of the situation. Since participants need to be informed about their involvement in any study due to ethical consideration, reactive effects in experiments cannot be eliminated completely.

112

While it is not necessarily needed to inform participants about the nature of the experimental study, only a slight amount of deception can be tolerated. Traditionally, studies in psychology and economics differ to the degree in which they allow for deception (McDermott 2002; Dickson 2011); psychological research usually has a higher tolerance for deception, whereas economists are generally more forthcoming with their research aim. In addition, experiments in economics and psychology typically also vary regarding the use of incentives and stylization (Dickson 2011). The following Subchapter 4.3 describes the experimental design of this study and discusses potential threats to the experimental design that could interfere with the study's objective.

4.3 The Experimental Design

This experiment was designed to measure the impact that emotional expressions of politicians have on viewers, their attitudes towards politicians, their evaluation of politicians and further aspects such as political parties and concepts such as political trust among voters.

   Due to the focus on attitudes among the general public, a large-scale online survey experiment that facilitates the use of video clips rather than laboratory studies seemed to be the best match regarding the research aim. Detecting the specific causal mechanisms at work was not the main interest of this study. Video clips can be easily incorporated into online surveys compared to telephone or paper-and-pencil questionnaires.

4.3.1 Pre-Test: The Structure of Politicians' Emotion Expressions

The media content analysis analyzed the emotional displays using the circumplex model as its theoretical underpinning. However, before experimental treatment was designed, a pre-test was conducted in July and August 2014 at the University of Koblenz and Landau and the University of Trier, in which eight student coders coded emotional expressions as discrete emotions in a sub-sample of 400 video sequences. This pre-test aimed at measuring the structure of emotions. Previous empirical studies have classified emotions by clustering emotion words (e.g., Schmidt-Atzert 2008; Shaver et al. 1987). Whereas those studies have often been concerned with how participants describe their own feelings, the novelty of the pre-

test can be seen in the new subject – televised emotional expressions by politicians instead of self-reported affective states.

A sample of 200 video sequences was taken from the news broadcasts as well as from the political talk shows of the media sample resulting in a total sample of 400 video sequences. The video sequences were drawn randomly from a pool of pre-coded sequences, which indicated that these clips potentially included an emotional expression.[26] After having received training in the recognition of emotions, each student coder rated half of the sequences according to a list of 17 potential discrete emotions (Ekman & Cordaro 2011) and the other half as open-ended items using their own words. The raters also indicated how certain they were about their rating as well as their perceived strength of the emotional expression.

The sequences that were rated with an open wording approach did not substantially differ from the 17 potential discrete emotions that were used for the other half of the items. Hence, the open-ended coding were later matched to the list of 17 discrete emotions, so that eight ratings were obtained for each video sequence. The prevalence of discrete emotions was determined by setting a criterion that at least four coders had to identify an emotional expression as being a specific discrete emotion within any given video sequence. Coders were instructed to rate more than one discrete emotion per video sequence if they recognized several distinct emotions or had the impression that one discrete emotion was not sufficient. As a result, two or even three discrete emotions frequently appeared within in one video sequence according to the raters (see Table 8).

---

26 The research assistants of the project rated the material previously to determine whether the sequence showed an emotional expression. This included two samples of news broadcasts and a random selection of video sequences from talk shows.

Table 8: The Percentages of Discrete Emotions in 400 Video Sequences

	All Video Clips		News		Talk Shows	
	N	Percent	N	Percent	N	Percent
Anger	103	25.75	27	13.5	76	38
Fear	17	4.25	13	6.5	4	2
Surprise	7	1.75	3	1.5	4	2
Sadness	14	3.5	13	6.5	1	0.5
Disgust	0	0	0	0	0	0
Contempt	31	7.7	9	4.5	22	11
Happiness	138	34.5	120	60	18	9
Pride	26	6.5	18	9	8	4
Amusement	64	16	19	9.5	45	22.5
Schadenfreude	15	3.75	6	3	9	4.5
Guilt	7	1.75	6	3	1	0.5
Hope	12	3	9	4.5	3	1.5
Embarrassment (Shame)	16	4	10	5	6	3
Compassion (Being Moved)	2	0.5	1	0.5	1	0.5
Relief	0	0	0	0	0	0
Rejoicing	10	2.5	9	4.5	1	0.5
Indignation	89	22.25	16	8	73	36.5

Note: Frequencies and percentages indicate that at least four out of eight coders recognized the discrete emotion as such within a video sequence. Ratings were observed for 400 video sequences.

As a next step, a cluster analysis was conducted to investigate which discrete emotions appeared together because they were coded within the same video sequence. Based on the circumplex model and the four categories along its two dimensions (valence and arousal), a four cluster solution could be expected, whereby negative-active, negative-passive, positive-active, and positive-passive emotions group together.

The k-means clustering approach is a non-hierarchical clustering approach that is suited to analyze the grouping of observations if the number of expected clusters is predetermined.[27] The following Table 9 shows the results of the cluster analysis for a four-cluster solution. All 17 discrete emotions were considered for each video sequence and had possible values ranging from "not recognized" (0) to "recognized by all coders" (8). The variables were z-transformed before the cluster analysis was conducted. Ta-

---

27 This approach uses a random starting point to form initial clusters. To ensure the reproducibility of the presented results, different starting points were chosen and cluster results with several different starting points were compared to the presented results in order to ensure a stable solution.

ble 9 presents the average values for each variable within the four clusters. The first cluster shows the highest average values for positive emotional displays such as happiness (1.01), pride (0.65), rejoicing (0.59) and hope (0.45). Because happiness, pride, and rejoicing co-occur with high arousal, they can be classified as positive-active emotions. Hope, on the other hand, could be classified as a positive-passive emotion due to its low arousal level (Russell 1980: 1167). The first cluster includes both states of arousal. The second cluster exhibits the highest average ratings for negative-active emotions such as anger (0.98), contempt (0.77), and indignation (1.00). According to the mean values of fear (1.56), sadness (1.97), guilt (2.01) and embarrassment (1.39), the third cluster can be interpreted as representing negative-passive emotional displays. The fourth cluster has the highest average ratings for emotional displays of schadenfreude (1.01) and amusement (1.47). These emotions are also positive-active emotions, but their point of reference is likely to be negative; for example, schadenfreude occurs at the misfortune of someone else and amusement could indicate the belittlement of someone else such as another politician in a talk show. Schadenfreude occurs only in 3.8 percent of the analyzed video sequences, while amusement occurs more frequently in 16.0 percent of the 400 video sequences. 76 observations group into this fourth cluster, which is distinct from the other three clusters and might indicate the special communication style of politicians in talk shows, including uncivil behavior in the sense of sneering comments, mockery, and ridicule, rather than another set of emotional expressions.

The cluster sizes also indicate that the majority of observations fell into either positive (131 observations) or negative-active emotional displays (147 observations). The positive emotional displays largely consist of positive-active emotions, such as happiness (34.5 percent), while positive emotions with a lower arousal, such as hope, cluster within this group but are rarely seen on television (3.0 percent). Negative-passive emotions are also less frequently displayed (46 observations) but appear fairly distinct from the other emotions; this impression is also amplified by a multidimensional scaling analysis of the cluster solution.

*Table 9: Mean Values of the K-Means Clustering for 17 Discrete Emotions*

	Cluster			
	1	2	3	4
Anger	-0.75	0.98	-0.29	-0.42
Fear	-0.24	-0.10	1.56	-0.33
Surprise	-0.20	-0.04	-0.25	0.56
Sadness	-0.33	-0.17	1.97	-0.29
Disgust	-0.30	0.40	-0.30	-0.07
Contempt	-0.65	0.77	-0.25	-0.22
Happiness	1.01	-0.77	-0.67	0.15
Pride	0.65	-0.25	-0.51	-0.33
Schadenfreude	-0.35	-0.09	-0.36	1.01
Guilt	-0.26	-0.26	2.01	-0.26
Rejoicing	0.59	-0.29	-0.30	-0.27
Embarrassment (Shame)	-0.21	-0.36	1.39	0.22
Relief	0.42	-0.32	-0.16	-0.02
Indignation	-0.66	1.00	-0.52	-0.49
Hope	0.45	-0.24	-0.20	-0.19
Compassion (Being Moved)	0.27	-0.25	0.43	-0.25
Amusement	-0.10	-0.47	-0.62	1.47

Note: Cell entries display mean values for each variable in each cluster after each variable was z-transformed.

The following Figure 4 presents a multidimensional scaling analysis based on a Euclidean distance matrix for the 400 video sequences. The group membership in this figure is determined by the previously described four-cluster solution of the k-means clustering approach (Table 9). By visually inspecting this graph, it can be seen that the third cluster of negative-passive emotions forms a distinct cluster that is the furthest apart from the other clusters. The vertical dimension divides the clusters along the dimension of arousal and thereby clearly separates the negative-passive cluster from the three other clusters, which include predominantly active or dominant emotions. The horizontal dimension distinguishes positive and negative emotions, as positive emotions have lower values and negative emotions higher values, particularly the negative-active emotions. The Clusters 1 and 2 of positive emotions and negative emotions are well separated. The fourth cluster, which shows the highest average values of schadenfreude and amusement, is located between positive and negative emotions. In many cases it overlaps with the first cluster of positive-active emotions. A few observations also indicate the expression of negative emotions and discrete emotions such as schadenfreude and amusement. Hence, this visual inspection is further indication that the fourth cluster does not represent a category which is completely distinct from positive-active emotions according the circumplex model of emotions, but could also emphasize the im-

portance of communication styles, uncivil behavior, and context. These findings are in line with the three-factor model of emotions (Morgan & Heise 1988; Mehrabian 1980; Steenbergen and Ellis 2006), which is proposed as a good empirical fit (2006): "Rather, three dimensions (one of positive affect and two of negative affect) are required in many circumstances" (Steenbergen & Ellis 2006: 111). This also seems to be a fitting model in the context of broadcasted emotional expressions by politicians on television. A mere distinction between positive and negative emotions is not deemed sufficient: "In this view, the two-dimensional model of affect may be insufficient: instead, we may need a three-dimensional model of emotions, encompassing positive affect, and two distinct types of negative affect" (Steenbergen & Ellis 2006: 110).

Figure 4: Multidimensional Scaling of Discrete Emotions
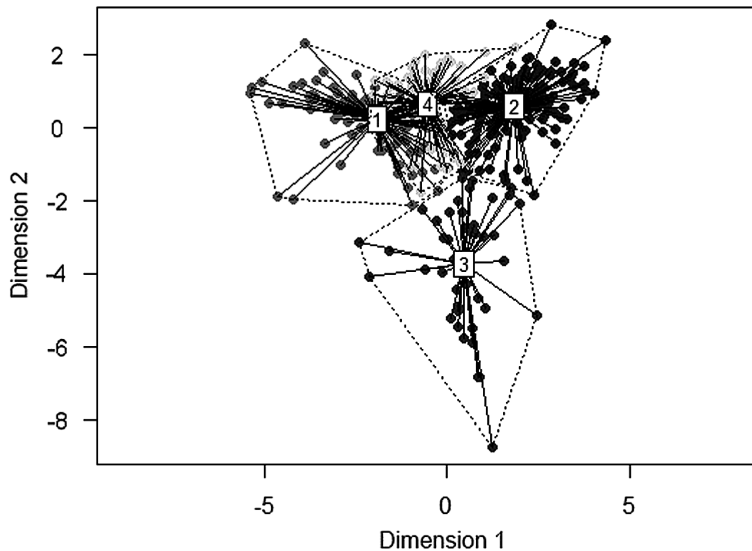
Note: Author's own illustration. The figure shows a multidimensional scaling analysis based on the Euclidean distance matrix. The first cluster shows positive-active emotions, the second negative-active, the third negative-passive and the fourth such emotions as amusement and schadenfreude.

Based on the observed frequencies within this pre-test, the experimental treatment was designed to largely follow the proposed three-factor model (Steenbergen & Ellis 2006) to measure effects of positive-active emotions

118

(positive), negative-active emotions (aversion) and negative-passive emotions (anxiety) for politicians as social groups.

In addition, the analysis searched for a fourth cluster of positive-passive emotions. However, a distinctive fourth cluster did not occur. The estimated fourth cluster is positioned between positive-active and negative-active emotions because its video sequences included both expressions. These sequences can be described by the concept of incivility – dominant expressions that show negative-active and positive-active emotions such as schadenfreude or amusement to belittle the political opponent. Therefore, the experiment also tests effects of this specific behavior, though it might not be a distinct emotion in itself.

Based on the frequencies within this pre-test, it can also be said that the negative-active emotions predominantly consist of anger and indignation, whereby two-thirds of the political talk show sequences showed anger and indignation. These discrete emotions are commonly displayed in political talk shows. This finding of three distinct clusters is in line with the three-factor model, which emphasizes that it captures the functions of emotions more strongly than the circumplex model (Steenbergen & Ellis 2006: 111):

> "The three-dimensional model brings some of the nuances of functional theories into the circumplex model. It agrees with the circumplex model that emotions can be represented with reasonable accuracy in a small number of dimensions. It also agrees with functional theories, however, in arguing that two dimensions are insufficient to capture the qualities of different emotions." (Steenbergen & Ellis 2006: 111)

Hence, the three-factor model is applied to the experimental design of the study,[28] as it has also found a wide recognition within the field of political communication research (e.g., Sullivan & Masters 1988; Bucy & Grabe 2008; Stewart et al. 2009b; Stewart & Ford Dowe 2013). In this study, the three-factor model of emotional expressions can be implemented with regard to the evaluation of politicians as a social group, while the emotional expressivity of the key political leaders only enables testing specific factors for each politician. Displays of fear/evasion or negative-passive emotions were hardly shown by any of the three political leaders during the time of

---

28 When video sequences were selected for the experimental treatment additional attention was paid to the discrete nature of the emotional expressions.

the media content analysis;[29] hence, displays of fear/evasion are not tested for specific political leaders in the following analysis. Display of happiness/ reassurance and anger/threat were more frequently displayed by the political leaders. In addition, Sigmar Gabriel is the key political leader whose displays of anger/threat include expressions of incivility. The inclusion of his expressions allows for testing the specific impact of anger/threat displays in conjunction with incivility. In comparison to his expressions, Angela Merkel and Gregor Gysi hardly show any uncivil behavior. In contrast, Gregor Gysi's speech acts often contain concerns of social injustices – regardless of his emotional expressions.

4.3.2 The Procedure

The online survey experiment was conducted over a period of six weeks in Germany, from March 19th to May 4th, 2015. In the span of six weeks, a pre-test was administered, followed by a first post-test at least two weeks later and a second post-test one week later. The participants were invited to the second post-test seven days after they had taken part in the second wave. Due to high response rates, it was not necessary to send out reminders for the second post-test. Participants could answer each questionnaire during a time period of ten days. The first wave was in the field from March 19th to March 29th, the second wave from April 16th to April 26th, and the third wave from April 23rd to May 4th.

Unlike a vast amount of studies in political psychology, this analysis is based on a quota sample from an online panel of the former polling agency Link (now Forsa), which is representative for German internet users aged 18 to 68. Participants for this panel were solely recruited offline, for example participants who were randomly contacted as part of a computer-assisted telephone interview were asked whether they might be interested in joining an online panel to participate in future online surveys at the end of the survey. Only previously contacted participants could have registered into the panel at this stage, not everyone could self-select into the panel via online advertisement. Therefore, self-selection biases are mitigated compared to other online panels to which participants can sign up online.

---

29 One exception was a eulogy to Nelson Mandela that Angela Merkel delivered in light of his passing. While such a delivery requires emotional expressions of sadness, the delivery can be seen as act of statesmanship outside the realm of daily politics, and should not induce negative evaluations by viewers.

However, those who respond to the survey invitation might still differ systematically from those who do not. The access panel provider claimed that panel members were only invited to a limited number of studies per year in order to keep response rates and data quality high. By doing so, participants are also less likely to change their response patterns due to frequent participation in research studies.

Nevertheless, online studies are at best only representative for internet users and not the population as a whole. Online samples are typically skewed towards a younger demographic; in regard to political attitudes, supporters of mainstream parties are underrepresented within online samples, while smaller parties such as the Greens usually have a higher number of supporters who use the internet frequently (Forschungsgruppe Wahlen 2015b). To counteract a severe bias in the sample due to heavy internet users and early responders, it was necessary to use a quota for gender and age. Reminders were sent out to participants for each of the three survey waves in order to increase participation rates.

Furthermore, the completion of the first survey was incentivized with an immediately available Amazon voucher of €2. The incentive to participate in first post-test questionnaire was higher to reduce the amount of panel attrition. Participants received an additional €3 voucher for completing the second questionnaire. Participation in the third questionnaire was also incentivized with an Amazon voucher of €2.50.

To ensure the experimental groups were sufficiently large as well as to allow subgroup analysis and the test of moderating variables, the study was outlined to survey 350–400 participants per experimental group. The experimental design consists of a total of nine experimental and five control groups that can be divided into three sub-experiments (*Type 1, Type 2,* and *Type 3*).

In the first wave, 36.8 percent of the invited panelists completed the questionnaire, resulting in 6011 initial responses without correcting for speeders. About 85 percent of the initial respondents also took part in the second wave. After data cleaning, 4840 participants remained who took part in both waves of the survey experiment (80.5 percent). The attrition of participants did not result in major changes regarding participants' demographic characteristics. An additional third wave was conducted to test the longevity of experimental effects. For this third wave, only half of the participants of the second wave were randomly invited to take part; therefore, the sample size was significantly smaller with 2120 initial participants and 2031 participants after data cleaning.

The experimental treatment was administered at the beginning of the second wave of the survey and was specific to the experimental groups. Before the video clips were shown, participants were notified about video clips in the subsequent questionnaire and encouraged to turn on their audio settings. The information was followed up with a multiple-choice question in order to test participants' recognition of a short sound bite among other sounds.[30] The experimental treatment was only administered if participants identified the short sound correctly, and the video clips were played one time without offering a repeated playback. An overview of all experimental groups is presented in Table 10.

Table 10: Overview of the Experimental Design

Group	Type	Video Treatment (Post-Test)	Politicians (N)	Video Clips (N)
		Experimental Groups		
EG-1	Type 1	1 Video: Male Politicians, Negative-Active	12	17
EG-2	Type 1	1 Video: Male Politicians, Negative-Passive	11	15
EG-3	Type 1	1 Video: Male Politicians, Positive-Active	22	22
EG-4	Type 1	1 Video: Female Politicians, Negative-Active	18	18
EG-5	Type 1	1 Video: Female Politicians, Negative-Passive	10	15
EG-6	Type 1	1 Video: Female Politicians, Positive-Active	12	19
EG-7	Type 2	2 Videos: Merkel / Gysi, Negative-Active	1/1	9/13
EG-8	Type 2	2 Videos: Merkel / Gysi, Positive-Active	1/1	12/10
EG-9	Type 3	2 Videos: Merkel / Gabriel, Negative-Active / Incivility	1/1	9/8
		Control Groups		
CG-1	Type 1	1 Video: Male Politicians, Neutral Expressions	15	15
CG-2	Type 1	1 Video: Female Politicians, Neutral Expressions	9	16
CG-3	Type 2	2 Videos: Merkel / Gysi, Neutral Expressions	1/1	9/12
CG-4	Type 3	2 Videos: Merkel / Gabriel, Neutral Expressions	1/1	9/8
CG-5	Type 1/2/3	No Video		

Note: The table provides an overview of all experimental groups including the number of displayed videos, politicians, and video clips.

The allocation to the experimental groups was conducted randomly after participants passed the initial quota. Each experimental group had 429 or 430 participants in the first wave; as a result of panel attrition, the number of participants dropped to 346 to 392 per group in the second wave. After data cleaning and deleting speeders from the analysis, the number of par-

---

30 In this multiple choice question, a brief sound bite of a bicycle bell was incorporated.

ticipants for both time points decreased slightly further by approximately 20 participants per group, which results in a total number of 328 to 373 participants per group. Only completed interviews were considered for the following analysis. Speeders were deleted from the analysis because they can signal low data quality (Greszki et al. 2015: 472). Participants were identified as speeders if they took less time to complete the questionnaire than a cut-off point that was determined as being 2.5 standard deviations smaller than the median value across all variables which measured the speed.[31] An overview of the number of participants per experimental group can be found in Table 11.

Table 11: Number of Participants in Each Panel Wave (Waves 1 – 3)

Group	W1	W1_WS	W2	W2_WS	W3	W3_WS
EG-1	430	404	373	350	162	156
EG-2	429	408	365	347	166	163
EG-3	429	406	361	340	149	145
EG-4	429	405	356	334	151	141
EG-5	430	412	376	361	162	155
EG-6	430	409	378	358	158	150
EG-7	429	410	359	340	146	139
EG-8	429	407	346	328	130	126
EG-9	429	414	349	336	139	135
CG-1	430	408	372	354	164	160
CG-2	429	405	362	342	152	145
CG-3	429	407	354	333	135	129
CG-4	430	403	366	344	146	137
CG-5	429	410	392	373	160	150
Total	6011	5708	5109	4840	2120	2031

Note: The column names stand for each panel wave (W1 = Wave 1, W2 = Wave 2, W3 = Wave 3). The additional "_WS" indicates the number of participants in each wave without speeders.

The full experiment consisted of three sub-experiments called *Type 1*, *Type 2* and *Type 3* (see Table 10). This study hypothesizes that emotional expressions of politicians have an impact on several dependent variables such as candidate evaluations or trust in the political system. The first type of experimental setup presented video clips of a group of several well-known and lesser well-known male or female politicians to capture effects on attitudes towards politicians in general (see *Type 1* in Table 10). Three different emotional expressions of female and male politicians were shown to

---

31 The speeders were identified as such by the access panel provider Link (now Forsa), which hosted the online access panel.

participants: negative-active emotions, negative-passive emotions and positive emotions based on the results of the pretest and the three-factor model (Steenbergen & Ellis 2006). Judging from the frequency distributions it can be said that anger is the predominant discrete emotion within the negative-active material, while joy (happiness) is the predominant emotion within the positive-active material. The negative-passive emotion consists mainly of guilt and sadness. In addition, two control groups that showed neutral appearances of female or male politicians were included in the experimental design.

These three emotional dimensions were chosen based on a pre-test that focused on a subset of video clips in the media content analysis and analyzed these clips according to the discrete emotions that were expressed by politicians (see Subchapter 4.3.1 for more details). This analysis showed that emotional expressions of politicians fell distinctively into one of the three distinct categories: positive-active, negative-active and negative-passive emotions. These clusters could also be described as happiness/reassurance, anger/threat, and fear/evasion or sadness. A fourth cluster emerged in which some dominant positive emotions co-occur with dominant negative emotions, such as amusement and anger. This fourth cluster is likely the result of an uncivil communication style that was present in several political talk shows in the media sample.

The empirical findings of the pre-test largely confirmed the three-factor model of emotions as the underlying theoretical model of the experimental design. In real life situations, discrete emotions are often in a state of flux or accompanied by other emotions; even the concept of compound emotions has been introduced in the literature on emotion research (Du et al. 2014). While some discrete emotions are certainly dominant within each of the experimental treatments, the three-factor model was chosen because the co-occurrence of emotions cannot be eliminated. At the same time, it relates to the social functions of emotions. In these experimental groups, participants saw a number of different politicians and so enough video material was collected for each emotional expression according to the three-factor solution. Besides this practical consideration of implementing the three-factor model with emotional expressions of several male and female politicians, the design enables testing the impact of emotional expressions on the evaluation of politicians in general. The separate experimental groups for male and female politicians allow the testing of gender effects that might occur due to different expectations for each group of male and female politicians regarding their emotional expressions.

The second experiment included emotional expressions of two specific political leaders within the German political system: Chancellor Angela Merkel and a key figure of the opposition Gregor Gysi (former parliamentary leader of the Left in the German Bundestag). By doing so, this design is more similar to previous research that focuses on single politicians and evaluations of them after seeing emotional expressions than the experiment *Type 1* (e.g., McHugo et al. 1985; Grabe & Bucy 2009; Stewart & Ford Dowe 2013). The effects for specific politicians can be expected to be stronger than the effects in experiment *Type 1*, because the evaluation of a specific political leader as a dependent variable is directly related to the experimental stimulus.

The experiment *Type 2* that deals with Angela Merkel and Gregor Gysi consists of experimental conditions that featured positive-active and negative-active emotional expressions by both politicians. In addition, a control group was included that showed neutral emotional expressions of Merkel and Gysi. The experimental condition of negative-passive emotions was not implemented for these two politicians, because neither politician displayed a sufficient amount of negative-passive emotional expressions during the time of media content analysis.[32] Negative-passive emotional expressions are generally expressed less frequently than negative-active and positive-active emotional expressions. Hence, negative-passive emotional expressions by these two politicians were not analyzed.

Further, it has to be noted that within this *Type 2* experiment, participants saw two video clips in total in each experimental group. The order in which participants saw Angela Merkel or Gregor Gysi was randomized. After the participants saw the first video, they were asked whether the politician in the first video clip made an emotional impression on them. Only if they indicated to have recognized any form of emotional expressions, were they further asked which of six discrete emotions they perceived.[33] After that, the second video was administered and only after this second video did the rest of the questionnaire continue, in which participants were asked to evaluate both candidates in more detail. Participants were not asked to evaluate the first politician after the first video, so that the evalua-

---

32 The only negative-passive emotional expressions by Angela Merkel occurred when she displayed sadness during a eulogy she delivered shortly after Nelson Mandela's death. Negative-passive emotional displays by Gregor Gysi were not recorded during the media content analysis.

33 The participants could choose one or several emotional expressions from six discrete emotions: happy, sad, outraged, angry, proud, and amused.

tion of the second politician would not be affected by the potential discovery of the research intention. However, because the evaluation of both politicians was only derived at the end of both clips, the effect cannot be clearly attributed to either clip, and it is possible that emotional expressions of one politician affect the ratings of the other politician. Both politicians could be evaluated in contrast to each other.

Finally, to analyze the distinct effect of incivility, a third experimental type was conducted. Participants saw video clips with emotional expressions by Angela Merkel and Sigmar Gabriel. Sigmar Gabriel, then leader of the Social Democrats and Minister for Economic Affairs and Energy, displayed anger and incivility, while the same video clips were used for Angela Merkel as in the experiment *Type 2*. This third type within the experiment aimed at measuring the effects of a specific anger, one that occurs together with uncivil behavior. Next to anger, other discrete emotions such as contempt are also more likely to be part of these clips due to the hostile nature of incivility. This third experiment for Sigmar Gabriel consisted of only two groups: the incivility condition and one neutral control group for comparison. The effects of incivility can then be compared to the negative-active emotions of the two other politicians. Further, the effects for Angela Merkel are replicated within a different context, which can provide additional insights into how much the context of other politicians' emotional expressions shape the impact of her anger. The comparison between Sigmar Gabriel's incivility and Gregor Gysi's anger is particularly interesting, as Gabriel's incivility has a higher amount of contempt due to the verbal attacks directed at other politicians, while the experimental treatment for Gregor Gysi does not include such personal verbal attacks on others and rather, next to anger, features indignation and outrage about policies and social injustices.

4.3.3 The Material

This study uses real-life video material that was aired on TV during the time of the media content analysis. These politicians – Angela Merkel, Gregor Gysi, and Sigmar Gabriel – are well known by the German public; it can be assumed that the participants of the study had already formed an opinion about them and as a result held certain predispositions towards them. Political science research within the context of U.S. politics has previously studied the impact of emotional expressions by former U.S. presidents such as Ronald Reagan (Masters & Sullivan 1989a), George W. Bush

(Stewart et al. 2009a), Bill Clinton (Grabe & Bucy 2009) and Barack Obama (Stewart & Ford Dowe 2013). In contrast to the majority of previous studies, this design does not focus only on incumbents and therefore allows for testing the effects of anger displays when politicians hold different offices within the political system.

While the largest proportion of all video sequences was collected as part of the media content analysis, a few exceptions had to be made for the specific political leaders of the experiment *Type 2* and *Type 3*. For those politicians, only video clips that aired on TV during May 2013 and April 2014 were selected as video sequences. Additional material from the public news channel "Phoenix" was collected on Angela Merkel as well as Gregor Gysi. This was done, to ensure a sufficient number of video clips in each emotional category and could not have been ensured otherwise, as Angela Merkel does not generally appear as a guest in political talk shows. Further video sequences of her were taken from televised party conventions and speeches at the German Bundestag. In contrast to Angela Merkel, appearances of Gregor Gysi and Sigmar Gabriel were hardly selected into the random sample of the media content analysis. Although both appeared as guests during the time of the media content analysis on several occasions (see Table A.1 in the online appendix),[34] none of these talk shows were part of the randomly selected sample; therefore, material from these talk shows was additionally screened for emotional expressions. This turned out to be highly sufficient for both politicians' negative-active emotions. For Gysi's positive emotional displays, coverage of parliamentary debates and interviews that was broadcast on Phoenix was also screened to ensure enough suitable material for the experimental conditions.

Each video clip had a length of approximately 1:30 minutes and consisted of several shorter video sequences that were sampled. Because participants saw two video clips in the sub-experiment *Type 2* and *Type 3*, the video segment lasted in total about three minutes, while the experimental treatments in the sub-experiment *Type 1* were only half as long. The control group that received no video treatment, had a longer questionnaire,

---

34 With eight appearances, Gysi qualifies as one of the most frequent talk show guests during May 2013 and April 2014. Only a few politicians appeared in talk shows more often during this time. Sahra Wagenknecht (10 appearances), Katrin Göring-Eckhardt (9 appearances), and four male politicians who belong to parties of the government coalition (SPD and CDU) or are as less well-known politicians (Thomas Oppermann (10 appearances), Wolfgang Bosbach (10 appearances), Armin Laschet (9 appearances), and Wolfgang Kubicki (FDP) (9 appearances) (see Table A.1 in the online appendix).

because it served as a baseline comparison and included questions regarding all political leaders and treatment conditions. The length of the video clips was deemed as being sufficiently long to display the emotion and to induce an effect. As the clips consisted of several shorter video sequences, the specific context of each clip might be less important. However, the anger displays by Sigmar Gabriel and Gregor Gysi each have an overall theme. Gregor Gysi's anger mainly targeted social injustices, while Sigmar Gabriel's emotional expressions can be classified as uncivil behavior.

Experimental studies that use real-world material are rare and it is a challenge to find a suitable control group with a placebo treatment of neutral emotions and a similar context. For that reason, several studies within the context of political psychology and political communication deploy actors instead of real-world politicians to analyze appearance or communication style effects and their impact on political trust and candidate evaluations (Dumitrescu et al. 2015; Mutz & Reeves 2005). This approach offers the advantage that an actual placebo treatment can be designed for the control groups that is identical to the experimental treatment except for the factor of interest. In this case, for an ideal placebo to exist, it would consist of identical video clips in the same talk show, with the same verbal content except the emotional expression itself. However, using real-world data as such, a placebo treatment is not feasible and can only be approximated by existing video clips with neutral expressions that are taken from the same talk shows with different wording, or similar wording from different talk show appearances.

Having established that due to heuristic inferences appearances – visually and aurally – can have an automatic impact on the evaluation of politicians, it is advisable to control for the physical appearance and demeanor of politicians by using neutral displays as a baseline comparison. Social psychologists have previously stressed the necessity of including a control group with neutral expressions to avoid any form of misattribution (Hareli et al. 2009: 38). Such a control group with neutral expressions also becomes a necessity when studying attitudes towards real-world politicians. Participants are likely to have already formed opinions about well-known politicians, especially those with a high interest in politics. If they are less familiar with appearances of politicians and their policy preferences, even neutral expressions could cause changes in the attitudes towards them due to mere exposure (Zajonc 1968).

128

4.3.4 The Sample

Before the study was conducted it was also considered whether the sample size would be sufficient to detect significant effects. The statistical power of an analysis indicates the probability of detecting an effect statistically. Typically, a power of at least 0.8 is required for research studies, which translates to an 80 percent change of detecting an effect. Three factors influence the power of an experimental design: the sample size, the effect size, and the alpha level of significance. A power analysis was conducted which estimated the sample size that is needed to detect a significant effect in a one-way analysis of variance or regression analysis (F-test) at the 5-percent level with a probability of 80 percent and small, medium and large effect sizes. As an indication of previous effect sizes in comparable experiments, the experimental study by Stewart, Waller and Schubert (Stewart et al. 2009a) was considered, which focused on emotional responses of viewers based on former U.S. president George H.W. Bush's facial expressions. Stewart et al. (2009a) reported an effect size that can be considered as a small effect; when Bush's micro-expressions of anger and happiness were removed from a video treatment, participants felt angrier than the control group (Stewart et al.2009a: 129).[35, 36] The experiment is sufficiently large to detect small effect sizes (f = 0.1), as 240 participants per experimental group are needed in experiment *Type 1*, 273 in experiment *Type 2*, and 322 participants in experiment *Type 3*. To detect medium effect sizes between experimental groups and the overall mean, the three sub-experiments need 39, 45, and 52 participants per experimental group depending on the number of experimental groups (five, four or three groups). Hence, the statistical power to detect significant effects for the average treatment effects is not a major concern of this study in regard to determining the main effects of the experimental treatment (F-tests).

---

[35] The formula to determine the effect size "f" is the following: $f = \sqrt{\frac{\eta^2}{1 - \eta^2}}$ (Cohen 1988: 284). Cohen (1988) further classifies a small effect size as f = 0.1, a medium effect size f = 0.25 and a large effect size f = 0.4 (Cohen 1988: 284–288). Taking the η^2 of the treatment effect of removed anger expressions on viewers' feelings of anger (Stewart et al. 2009a: 129), the effect size is determined as a small to medium effect size: $f = \sqrt{\frac{0.023}{1 - 0.023}} = 0.15$.

[36] Other studies with a focus on the evaluation of politicians as a dependent variable (e.g., Tiedens 2001; Sullivan & Masters 1988) did not report effect sizes, therefore the study of Stewart et al. (2009a) was chosen as an approximation of the effect size that might be observed in the experimental study.

Further, the heterogeneous sample prevents the results from being mainly based on student samples whose participants form a specific sub-population and their response behavior might not be universal (Henrich et al. 2010). Although the representativeness of online samples for the general population can be questioned, the study is based on a panel that is representative for German internet users. Previous studies have shown that the German internet population differs from the general population and that online studies commonly show certain biases as a result. While most people under the age of 60 use the internet, only one in two elderly citizens use the internet, and elderly women use the internet even less frequently than elderly men – 40 percent vs. 59 percent (Forschungsgruppe Wahlen 2015b: 1).

Regarding their level of education, German citizens above the age of 35 who have obtained only a basic level of education (Hauptschulabschluss) use the internet less frequently (Forschungsgruppe Wahlen 2015b: 2). In addition, citizens who work as blue-collar workers use the internet less frequently (Forschungsgruppe Wahlen 2015b: 2). In terms of party preferences, supporters who regularly vote for the CDU/CSU or SPD use the internet less frequently, while voters of smaller parties, particularly those who vote for the Greens and the AfD, use the internet more often than the average citizen (Forschungsgruppe Wahlen 2015b: 2).

The descriptive statistics in terms of age, gender, education and party preference for this study is displayed in Table 12. Quotas were applied for age and gender. However, slightly more men participated than women. The study provides no information about the provenience or residency of participants on a state level or a regional level, such as East and West Germany (East/West). Hence, no weighting can be applied in this regard, which is an otherwise typical standard procedure in traditional telephone surveys.

Table 12: Descriptive Statistics of the Participants in Wave 1

	Wave 1	Wave 1_WS
Age	44.40 (45)	44.82 (45)
Gender		
Male	3155 (52.5)	2992 (52.4)
Female	2856 (47.5)	2716 (47.6)
Education		
Left School	15 (0.2)	14 (0.2)
Basic Education	765 (12.7)	739 (12.9)
Medium Education	2909 (48.4)	2776 (48.6)
Qualification for University Entrance (Applied Sciences)	553 (9.2)	532 (9.3)
General Qualification for University Entrance	1649 (27.4)	1531 (26.8)
Other	104 (1.7)	100 (1.8)
Pupils	15 (0.2)	15 (0.3)
Party Identification (PID)		
CDU/CSU	1899 (31.6)	1830 (32.1)
SPD	1397 (23.3)	1336 (23.4)
Greens	657 (10.9)	622 (10.9)
The Left	464 (7.7)	440 (7.7)
Other Party	376 (6.3)	350 (6.1)
No Party	1209 (20.1)	1122 (19.7)
Scalometer		
Merkel	6.60 (7)	6.65 (8)
Gysi	4.60 (5)	4.59 (5)
Gabriel	4.75 (5)	4.76 (5)
Ideological Self-Positioning		
Left/Right Cultural	5.57 (5)	5.57 (5)
Left/Right Socioeconomic	4.63 (5)	4.63 (5)
Observations	6011	5708

Note: Cells display mean values and median values in parentheses for metric variables, frequencies and percentages for categorical variables. The first column shows frequencies before data cleaning, while the second column shows frequencies after the data cleaning process. Both columns only consider completed interviews for participants who took part in both surveys. The measurements refer to the questionnaire during the first wave.

Compared with the general population, the sample shows lower percentages for those who have obtained a basic level of schooling (*Hauptschule* or *Volksschule*), larger percentages for those with a medium level of schooling as well as participants who held university entrance level qualifications (specific for applied sciences or general) (see for comparison: Statistisches Bundesamt 2015). Regarding party preferences, a direct comparison to the general public is a challenging endeavor. In telephone surveys that were conducted around the time of the study, participants indicated whether they typically vote for one party. In those weekly surveys, participants are

commonly asked about their hypothetical voting intention if the general election was held next Sunday.[37] Since the vote intention was not measured during this study, only the party identification of the participants can be inspected for comparison. The concept of long-term party identification potentially differs from any short-term vote intention for a party. According to Ohr and Quandt (2012), only 60 percent of the population identifies with a political party (Ohr & Quandt 2012: 189). The participants of this study are more likely to be attached to a political party than the general public. In addition, they might be slightly more likely to vote for the Greens and less likely to vote CDU/CSU compared to the general population (for comparison Forschungsgruppe Wahlen 2015b).

Furthermore, the scalometer ratings in the study can be compared to scalometer ratings that were obtained for the Angela Merkel, Gregor Gysi, and Sigmar Gabriel around the time of the study by the Politbarometer (Forschungsgruppe Wahlen 2017). It has to be noted that with the previous vote intention for political parties, these evaluations are estimates based on telephone surveys and liable to the margin of error. As a result, these estimates can deviate from the true value; nonetheless, a comparison can indicate how strongly the online sample could potentially differ from weighted estimates based on telephone surveys. If the results of the online study are recoded to the scalometer ratings ranging from -5 to +5, the participants of this study hold on average lower opinions of all three politicians (Angela Merkel 1.6 on average vs. 2.8, Gregor Gysi -0.4 vs. 0.3, and Sigmar Gabriel -0.25 vs. 1.3). Looking at the median of Merkel's scalometer rating, the deviance is less severe since her overall evaluation is negatively skewed in this sample. In addition to sampling biases, the differences between both samples could be the result of mode effects. In contrast to a telephone survey, the online survey provided a visual representation of the scale on a screen which could have affected the response behavior.

4.3.5 Potential Threats to the Internal Validity of the Study

Although experiments are well suited to test causal effects, they are not a panacea (McDermott 2002: 340), and can still suffer from poor design choices that are a threat to the causal claim of the experiment. Hence, the

---

37 German federal elections are typically held every four years on the next to last or last Sunday in September.

next section focuses on possible threats to the causal claims of this experimental design.

Bryman (2012) considers *testing* to be a potential source of error in experimental designs. If participants are primed by the pre-test, they might be more aware of the experimental treatment as a result and consequently alter their post-test responses (Bryman 2012: 52). Since the first and crucial post-test was conducted a full two weeks after the initial pre-test was conducted, the time span seems sufficiently long enough to ensure that participants will most likely not remember their initial answers to the pre-test. For this reason, severe effects due to the pre-test are not likely to occur. The Solomon four-group design is often advocated as the most suitable experimental design to control for effects of testing (Spector 1981: 60). However, its drawback is that a larger sample size is needed; as a result, simpler designs are often used instead of a Solomon four-group design (Campbell & Stanley 1971: 196). The effects of testing do not seem to be severe in this design due to the two weeks between the pre- and post-test; as a result, a Solomon four-group design has not been implemented. Rather testing several emotional expressions according to the three-factor model were employed instead.

Large parts of the study questionnaire were pre-tested at the University of Stuttgart during the summer term of 2012 with a student sample of 238 students. For this pre-test, the emotional expressions were measured according to the valence model of negative and positive emotions.[38] Pre-tests are an important step in any original large-scale study. They show whether and in which respects the research design might need to be revised. In this case, the underlying structure of emotions was based on the valence model which might have obscured the effects of discrete emotions. In addition, the treatment material could be specified to show stronger emotions and

---

38 These expressions were compared to a control group without video treatment. The sample had a mean age of 24.1 years of age (median 23 years of age). Participants saw videos of Sigmar Gabriel (SPD), Angela Merkel (CDU), Claudia Roth (The Greens) and Guido Westerwelle (FDP). The comparison was conducted between and within groups by comparing candidate evaluations before and after the video treatment. Only a few significant group differences could be determined. Participants rated Claudia Roth as less likeable after seeing her negative emotions compared to the group without video treatment ($F = 4.16$, $df = 208$, $p < 0.05$; post hoc test (Tukey HSD Mean Difference = -0.4 (-0.76; -0.03, $p < 0.03$). They also rated Sigmar Gabriel's leadership skills as lower after being exposed to his positive emotional expressions ($F = 4.34$, $df = 181$, $p < 0.01$; post hoc test (Tukey HSD Mean Difference = -0.52 (-0.94; -0.10, $p < 0.01$).

include longer video clips which are sufficient to induce responses. Since the focus of this study lies on the specific impact of negative-active emotions, the results of the pre-tests are not considered any further at this point.

The assignment of participants to experimental groups was conducted randomly. In that sense, true randomization occurred, and the control and experimental groups should not differ from each other at the beginning of the study. Nonetheless like any online survey which intends to reach a certain population and is left with those participants who respond to the invitation, the study potentially suffers from a selection bias. As long as the participants are missing at random and did not systematically refuse to participate in the study, concerns regarding the selection processes are not extremely severe. Furthermore, the design is unbalanced due to the panel attrition which led to slightly different numbers of observation for each experimental group. However, these differences can also be considered as missing at random and to not pose a systematic bias to the study.

Furthermore, even experimental designs can suffer from an "ambiguity about the direction of causal influence" (Bryman 2012: 53). Problems of endogeneity are a common concern in observational studies (for example: Allison 1999: 53; Lewis-Beck 2006: 211). They can even occur in experimental studies if the sequential order of the treatment and dependent variable are not clearly untangled and the direction of the causal effect remains unclear (Bryman 2012: 53). While the experimental design prevents problems regarding the temporal order of cause and effect (the treatment clearly precedes the post-test evaluation), the *Type 2* and *Type 3* sub-experiments in fact suffer from an attribution problem. Participants in these experimental groups were exposed to two video clips of political leaders and answered only after seeing both video clips questions about the politicians regarding their specific character traits. Because of this design choice, it cannot be determined whether the evaluation of a politician is solely the result of the emotional expressions displayed by that politician or in fact a result of the comparative evaluation of emotional expressions of both politicians. Strictly speaking, the experimental treatment can only be interpreted as a result of being exposed to video clips with emotional expressions of both politicians. One politician might look better or worse in contrast to the emotional expressions of the other, which is known in survey research as a contrast effect (Kuklinski et al. 1997: 328).

Some ethical concerns are addressed as it is common practice in experimental research to do so. Those who responded to the survey invitation participated voluntarily and they were aware about their participation in a

study regarding "current politics". Therefore, informed consent did exist, although a slight amount of deception was used since participants were not aware of the experimental nature of the study and its aim. This slight deception was justified by the possible impact this information might have had on the participants' behavior. Previous research has shown that participants behave differently when they are aware of the experimental manipulation. Furthermore, the experiment was conducted online and can be considered fairly non-intrusive; the respondents had the opportunity to opt-out at any given time. Nonetheless, video treatments need to be considered more carefully than a mere survey questionnaire, because their content could potentially affect participants' emotional states. The video clips showed emotional expressions of politicians and could potentially induce a certain emotional response. These clips had no disturbing content and therefore should not have induced highly unpleasant emotions for participants. The video clips have been broadcast on television and due to their status as public figures, no privacy rights of the politicians were violated.

After data cleaning, the sample has a sample size of 4840 participants across all 15 experimental and control groups. Hence, the study can still be considered a large-scale study that allows for subgroup analysis due to its size. Heterogeneous treatment effects can be tested for various subgroups of respondents, for example the treatment effects can be analyzed for various party identifiers and non-identifiers. Studies that merely use student samples often lack variance regarding variables such as party identification and political attitudes since college students are likely to identify with liberal parties and liberal attitudes (Henrich et al. 2010; Mutz 2011: 144–146). Therefore, this study offers a rare advantage to test the experimental effects for different party identifiers and political attitudes.

Furthermore, the experimental treatment has to be administered in an exact manner to each participant of the treatment unit, as biases could be introduced into the study if slight variations occur when administering the treatment. With automated online survey experiments, concerns regarding the *instrumentation* can be well tested before the experiment is conducted, and the reliability between iterations is maximized and interviewer effects are minimized (Bryman 2012: 53).

Selection can be a concern if the units of analysis are not truly randomly allocated to treatment and control group (Bryman 2012: 53). In the field of online survey experiments, the allocation to the experimental groups is automated and based on random number algorithms, therefore, the random allocation is not a major concern for this study.

135

Lastly, *history*, *mortality*, and *maturation* can be identified as potential threats to experiments (Bryman 2012: 52–53). History is a threat to any experiment with pre-test and post-test conditions. *History* subsumes all potential external factors that could influence the participants during the period of the investigation and interfere with the treatment. The threat of *history* increases the longer the study runs. One way to control for possible events that could influence participants' responses to repeated measures is by including control groups without any treatment. Significant changes in these control groups should not occur. The threats of *mortality* and *maturation* are also connected to the duration of an experimental study. *Mortality* refers to the panel attrition and increasing likelihood of participants dropping out of the study. The use of a control group ensures that dropout rates for experimental and control groups will likely be similar, avoiding a systematic bias of the study. The concept of *maturation* refers to the individual changes – improvements or refinements – of personal skills and attitudes. Hence, people can have a change in attitudes simply due to the fact that someone matured during the period of investigation; however, changes that are due to *maturation* should occur at an equal rate within the control group.

While this approach does eliminate direct threats of history, mortality, and maturation, it is also worthwhile to focus on real-life events during the time of the data collection that might affect the response to the stimuli, and in this case to three specific politicians. Any scandal that might have occurred during the time of the experiment could have had a severe impact on the evaluation of the politicians.

Ideally, a media content analysis could have been conducted during the time of the experiment. Data from *Google Trends* can be analyzed as a feasible cost- and time-effective approximation. Google Trends provides access to an (unbiased) sample of google searches;[39] it classifies the volume of searches for any given search term on a hit index, ranging from 0 to 100. For this experiment, the Google Trend data was accessed for the whole year of 2015 when searches from Germany were made for "Merkel", "Gysi", and "Gabriel". Additionally, the full names "Angela Merkel", "Sigmar Gabriel", and "Gregor Gysi" were accessed; however, the results did

---

[39] For more information on *Google Trends Data* see a description of Google Search Data by Google, available online at: https://medium.com/google-news-lab/what-is -google-trends-data-and-what-does-it-mean-b48f07342ee8 (last accessed: 05 June 2019).

136

not differ substantially, and it seems likely that most people merely google politicians by their last names.

Hence, Figure 5 shows the interest over time for each politician according to Google Trends data. The interest was not unusually high during the time of the data collection, which indicates the absence of any major scandal. One spike occurred for Sigmar Gabriel in early 2015; this spike could have occurred due to the fact that he attended a discussion with some Pegida attendees in Dresden – he was widely criticized for this and the incident gained a lot of media attention.[40] The search index for Gregor Gysi shows a spike in early June 2015, when he decided to step down as chairman of the parliamentary group on June 7, 2015,[41] while keeping his mandate as Member of Parliament. This event was then followed by his highest favorability rating in the history of the Politbarometer until then (Forschungsgruppe Wahlen 2017; see also Figure 1),[42] indicating that it was widely received by the public and not just according to Google Trends data. Fortunately, his withdrawal took place a month after the data collection had already been completed. Lastly, the search index is overall higher for Angela Merkel, which reflects her status as the most important German politician. It can also be seen that spikes occur during the second half of 2015, which are likely related to the refugee crisis and her role in handling the crisis.

---

40 See for example an article in the *Süddeutsche Zeitung* (06 February 2015), available online at: http://www.sueddeutsche.de/politik/spd-chef-ueber-pegida-gabriel-es-gibt-ein-recht-deutschnational-zu-sein-1.2336763 (last accessed: 05 June 2019).

41 See for example an article in *Der Tagesspiegel* (07 June 2015), available online at: http://www.tagesspiegel.de/politik/gregor-gysi-tritt-ab-ende-einer-aera-was-wird-aus-der-linkspartei/11881594.html (last accessed: 05 June 2019).

42 For comparison, see the Politbarometer newsletter from June 2015, available online at: https://www.forschungsgruppe.de/Umfragen/Politbarometer/Archiv/Politbarometer_2015/Juni_2015/(last accessed: 05 June 2019).

Figure 5: Interest Over Time in Key German Politicians
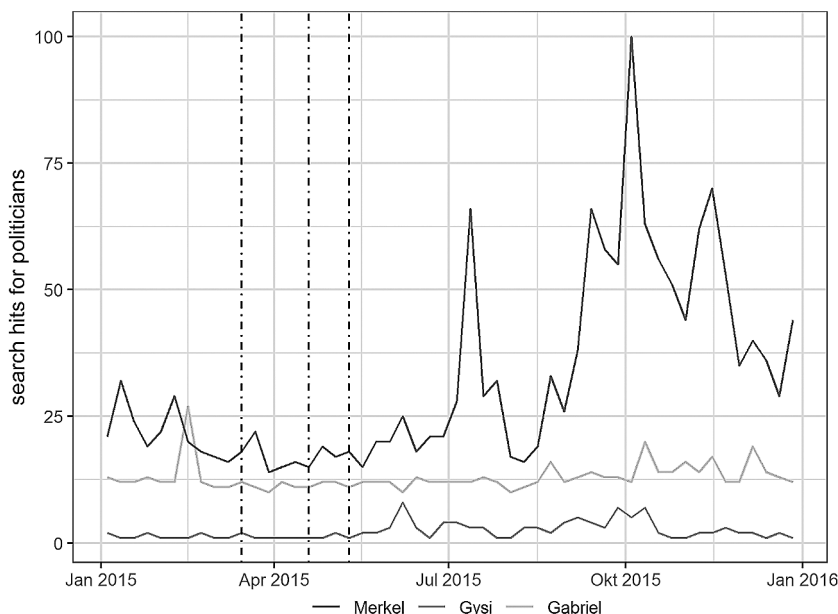
Note: Author's own figure based on Google Trends data accessed via "gtrendsR" (Massicotte & Eddelbuettel 2017).The first two vertical lines indicate the beginning of the first and second panel wave, respectively, while the third vertical line marks the end point of the data collection for the third and final wave.

The next section describes the measurement of key concepts for this study. The independent variables are the experimental treatment groups that have previously been described. Therefore, measurements of the dependent variable, moderating factors and potentially confounding effects are discussed in the next section.

4.4 The Measurement of Key Concepts

4.4.1 Measurements of the Dependent Variable: The Structure of Leadership Evaluations

Several trait assessments have been included in this study and have been measured in all three panel waves. According to the concept of *candidate orientations,* traits such as "likeability" and "trustworthiness" are often clas-

sified as personal factors, while "leadership skills" and "problem-solving skills" are classified as performance-related factors (for example see Gabriel et al. 2009). Hence, the latter variables can be seen as representing a dimension of competence, while likeability and trustworthiness potentially represent one dimension of warmth – or more broadly friendliness and integrity. The wording of the four items for politicians in general was as follows:

> "We have gathered some opinions about politicians in general. How much do you agree with the following opinions? - (A) Most politicians have strong leadership skills. (B) Most politicians are likeable people. (C) Most politicians are able to solve political problems. (D) Most politicians are trustworthy."

Besides the four items of candidate orientation, two items of external efficacy were also included in the questionnaire matrix for politicians in general (Beierlein et al. 2012). The two items generally measure how caring politicians are perceived to be, by asking participants on a 5-point Likert scale how caring and in close contact with the people politicians are. Therefore, these items are of similar quality to the items of candidate orientation, and often similar items are used to study candidate qualities and relate more specifically to the integrity of politicians (Hayes 2005: 921; Brettschneider & Gabriel 2002: 145). These items could also be interpreted as political empathy (Kinder 1986: 241). In this study the items were measured according to the Political Efficacy Short Scale (Beierlein et al. 2012):

> "(E) Politicians care about what ordinary people think. (F) Politicians seek a close contact with the people."

And for the three leading politicians items of candidate orientation were measured as follows:

> "And to what extent do the following characteristics apply to [Angela Merkel/ Gregor Gysi/ Sigmar Gabriel]? - (A) [He/She] has strong leadership skills. (B) [He/She] is a likeable person. (C) [He/She] is able to solve political problems. (D) [He/She] is trustworthy."

In addition to these four frequently used items of candidate orientation, which were measured with a 5-point Likert scale ranging from totally disagree (0) to totally agree (4), this study also included a set of character traits on semantic differential scales. The semantic differential scales included five pairs of character traits for each of the leading politicians, ranging from emotional to rational, polite to impolite, relaxed to agitated (tense, excited), aggressive to peaceful (amicable), and arrogant to modest

(humble).[43] For the politicians in general two additional sets of items were included in the questionnaire. Participants also rated politicians in general from decisive to unassertive and from resilient to overstrained. These items are linked to gender stereotypical expectations and were adjusted from the German Extended Personal Attributes Questionnaire (GEPAQ) to measure gender stereotypical evaluations (Runge et al. 1981; Renner 2019). "Decisive" and "resilient" can be attributed to male stereotypes, while being "unassertive" and "overstrained" are stereotypically linked to female behavior (Renner 2019).

These semantic differentials were introduced with the following question:

> "How typical are the following characteristics for [name of politician/ most politicians]. Please use the following word pairs to answer. The middle categories can be used to specify your opinion gradually. - [Most politicians/ Angela Merkel/ Gregor Gysi/ Sigmar Gabriel] are... [emotional... rational / polite... impolite / relaxed... agitated / aggressive... peaceful / arrogant... modest / decisive... unassertive / resilient.... overstrained]."

Lastly, a measurement of an overall summary score was also included in the study. This measurement used an 11-point Likert scale ranging from -5 to +5 in the questionnaire. This summary score for personality impressions of a politician adheres to the common phrasing of this score in German surveys on politics and politicians and has been occasionally applied by Gallup in the context of U.S. politics (e.g., Forschungsgruppe Wahlen 2017; Rattinger et al. 2017; Saad 2012). For politicians in general and the three political leaders the questions were administered as follows:

> "All in all, what do you think about politicians in Germany? - I have a very low opinion of them (-5); I have a very high opinion of them (+5)." / "Please, tell us, what you think of some leading politicians [Name of Politician]. - I have a very low opinion of her/him (-5); I have a very high opinion of her/him (+5)."

---

43 In addition, one item measured evaluations ranging from attractive to unattractive. Since this last pair of personal characteristics focuses on the appearance of political leaders rather than their personality traits, the item is excluded from the following analysis, which focuses on the structural relationship of personality traits rather than physical attractiveness.

The use of these questions can be interpreted as a favorability measure comparable to the use of a feeling thermometer, which is more commonly applied in English language surveys and has been used in the study of emotional expressions and evaluations of politicians (e.g., ANES 2010; Stewart & Ford Dowe 2013); however, due to the well-tested phrasing of the item in German questionnaires, the scalometer was chosen for this study. Such an overall score is often interpreted as a summary score, which is frequently updated by voters as a running online-tally and stored in their memory, while the event that had led to the evaluation is forgotten. Such a running tally of negative and positive affect simply describes information processing as keeping count of positive and negative evaluations of a politician (Laustsen & Petersen 2017; Lodge & Taber 2013) and represents an overall impression which most strongly reflects the warmth-dimension instead of the competence-dimension (Laustsen & Bor 2017).

The next section investigates the underlying dimensions of the politicians' trait evaluations based on exploratory factor analyses. Exploratory factor analyses aim at discovering underlying dimensions of latent variables or factors. In contrast to a confirmatory factor analysis, the number of factors and specific items belonging to an underlying factor are not predetermined by theoretical assumptions. Exploratory factor analysis is more suitable for finding a factor structure and particularly so before any confirmatory factor analysis is conducted (Patil et al. 2008: 166). Hence, an exploratory factor analysis appears to be a good starting point to examine the underlying structure of candidate evaluations.[44] The following analysis empirically tests whether an underlying structure between the items can be identified for each of the three key politicians as well as politicians in general; to do this, the variables were taken from the pre-test condition (Wave 1) before any experimental manipulation had occurred.

---

44 Based on the strongest theoretical evidence a two-factor solution can be expected with two latent factors of warmth and competence or integrity and leadership. A confirmatory factor analysis, however, requires at least three items per factor, since the model is otherwise underspecified. If the overall summary score of the scalometer is counted towards the dimension of warmth, then this requirement is only fulfilled for the dimension of warmth and not for the dimension of competence. Furthermore, focusing merely on the four or five items regarding the concept of candidate orientation, a confirmatory factor analysis for the four items of candidate orientation is not feasible. Only if at least the summary score and potentially the semantic differentials are included in the analysis is a confirmatory factor analysis statistically feasible.

One concern lies in the measurement of the available items. The four items measuring the concept of candidate orientation differ structurally from the semantic differentials. The item wording of the semantic differentials did not include statements that were made up of complete sentences, instead only two opposing adjectives were displayed. Therefore, not only did the items differ syntactically, the semantic differentials could have also evoked a stronger link towards specific personality traits compared to broader concepts of warmth and competence or even leadership and problem-solving skills. Leadership and problem-solving skills seem to be more complex evaluations compared to specific character trait evaluations such as being emotional or rational.

Even if character traits might not be related to the competence dimension at first sight, more character traits could be interpreted as well as being important for both dimensions. A competent actor has to be rational in decision-making tasks and demonstrate social skills to succeed in social interactions. This can be done in more than one way. While the dimension of competence might also be interpreted as dominance dimension – for example a leader who demonstrates his dominance in negotiations – leadership can be practiced very differently and show more characteristics of compromise, consensus, and therefore agreeableness rather than dominance. So, even problem-solving skills could be interpreted as one dominant decision made by a political leader, or it could be interpreted as skill that includes thoughtful, tactful, and consensual behavior.

It is also worth mentioning that these semantic differentials are potentially stronger related to the experimental treatment – they were designed with the experimental treatment in mind and adapted from previous measures of impression formations in regard to politicians' negativity and incivility (Mutz & Reeves 2005: 14). Mutz and Reeves' (2005) main focus lies on the effects incivility has on political trust; however, they use semantic differentials as manipulation check for "perceived levels of incivility" (Mutz & Reeves 2005: 14).[45] In that sense, these semantic differentials show candidate impressions more directly compared to the more abstract evaluations of certain characteristics such as the candidate orientation.

───────────────

45 The manipulation check by Mutz & Reeves (2005) was measured by the following semantic differentials: "emotional–unemotional", "quarrelsome–cooperative", "friendly–hostile", "rude–polite", "calm–agitated" (Mutz & Reeves 2005: 14).

Before a factor analysis is considered a meaningful endeavor, sufficient bivariate correlations should occur between the variables.[46] Hence, correlation matrices are inspected as a first attempt to gain some insight into the underlying factor structure (see Figure 6; Figure 7; Figure 8; Figure 9). Analyzing the correlation matrix, it can be seen that moderate to strong correlations occur for the overall rating (scalometer) and the four items of candidate orientation. Leadership skills are slightly less correlated with overall ratings and likeability scores for the three politicians as well as politicians in general, for example the correlation between Merkel's likeability ratings and leadership assessment is only moderate (r = 0.47). The semantic differentials are noticeably less correlated with each other as well as the overall ratings and the candidate orientation items. The item "emotional–rational" is not correlated with the other items for each politician and politicians in general. The item "calm–agitated" shows a weak negative correlation with the other variables. For the subsequent factor analysis, the items "calm–agitated" and "emotional–rational" were reverse coded for all models. Also reverse coded were the additional semantic differentials of being "decisive–unassertive" and "resilient–overstrained" that were measured only for politicians in general, which show negative correlations with the other variables.

---

46 A rule of thumb is generally that metric and quasi-metric variables should correlate at least moderately with each other (Pearson's r = 0.3 to 0.4).

Figure 6: Correlation Matrix of Leadership Trait Evaluations for Angela Merkel
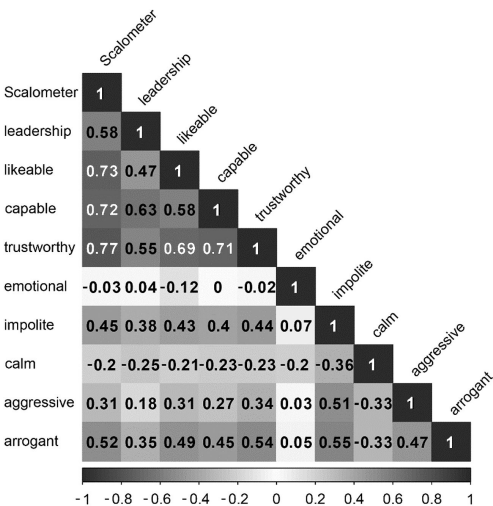
	Scalometer	leadership	likeable	capable	trustworthy	emotional	impolite	calm	aggressive	arrogant
Scalometer	1									
leadership	0.58	1								
likeable	0.73	0.47	1							
capable	0.72	0.63	0.58	1						
trustworthy	0.77	0.55	0.69	0.71	1					
emotional	-0.03	0.04	-0.12	0	-0.02	1				
impolite	0.45	0.38	0.43	0.4	0.44	0.07	1			
calm	-0.2	-0.25	-0.21	-0.23	-0.23	-0.2	-0.36	1		
aggressive	0.31	0.18	0.31	0.27	0.34	0.03	0.51	-0.33	1	
arrogant	0.52	0.35	0.49	0.45	0.54	0.05	0.55	-0.33	0.47	1

-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1

Note: The figure displays the lower half of a full correlation matrix. Cells show correlation coefficients. The diagonal line displays eigenvalues.

Figure 7: Correlation Matrix of Leadership Trait Evaluations for Gregor Gysi
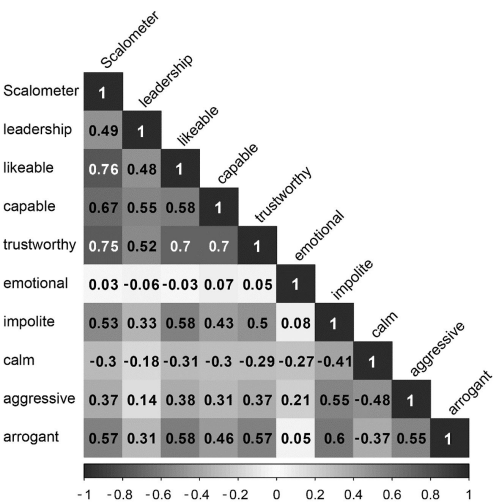
	Scalometer	leadership	likeable	capable	trustworthy	emotional	impolite	calm	aggressive	arrogant
Scalometer	1									
leadership	0.49	1								
likeable	0.76	0.48	1							
capable	0.67	0.55	0.58	1						
trustworthy	0.75	0.52	0.7	0.7	1					
emotional	0.03	-0.06	-0.03	0.07	0.05	1				
impolite	0.53	0.33	0.58	0.43	0.5	0.08	1			
calm	-0.3	-0.18	-0.31	-0.3	-0.29	-0.27	-0.41	1		
aggressive	0.37	0.14	0.38	0.31	0.37	0.21	0.55	-0.48	1	
arrogant	0.57	0.31	0.58	0.46	0.57	0.05	0.6	-0.37	0.55	1

-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1

Note: The figure displays the lower half of a full correlation matrix. Cells show correlation coefficients. The diagonal line displays eigenvalues.

Figure 8: Correlation Matrix of Leadership Trait Evaluations for Sigmar Gabriel

Note: The figure displays the lower half of a full correlation matrix. Cells show correlation coefficients. The diagonal line displays eigenvalues.

Figure 9: Correlation Matrix of Leadership Trait Evaluations for Politicians in General
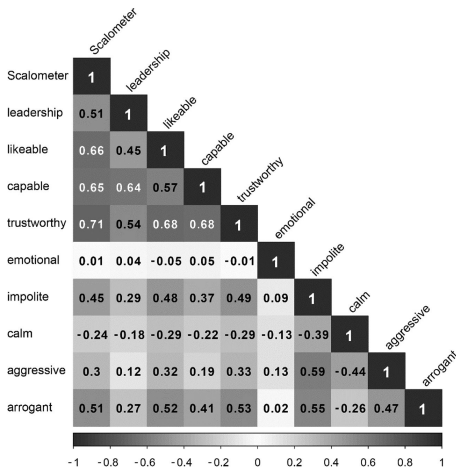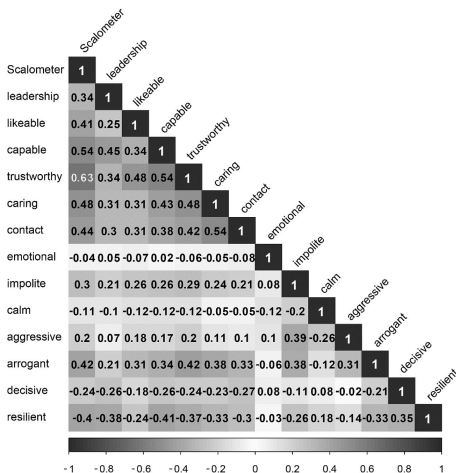

Note: The figure displays the lower half of a full correlation matrix. Cells show correlation coefficients. The diagonal line displays eigenvalues.

145

As a first step in undertaking an exploratory factor analysis, it has to be considered whether a factor analysis is indeed a suitable approach. To do so, the Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test of sphericity are typically required measurements to judge the adequacy of a factor analysis (Patil et al. 2008: 166). Regarding the Kaiser-Meyer-Olkin measure, the application of a factor analysis seems suitable in respect to all three politicians and politicians in general.[47] The overall measure of sampling adequacy (MSA) ranges from 0.89 for Gabriel, Merkel, and politicians in general to 0.90 for Gysi and therefore suggests an adequate sampling base for factor analysis, as values closer to 1 indicate a better adequacy (Field 2009: 647).

As a second measure, Bartlett's test of sphericity, is commonly analyzed.[48] Bartlett's test indicates that an exploratory factor analysis is an appropriate analysis for each of the three politicians and also for politicians in general in accordance with the previously conducted correlation analyses. All four Bartlett tests are highly significant, however, given the large sample sizes in contrast to the relatively small number of variables, the significant tests might not be surprising (Field 2009: 648). Nevertheless, the first two requirements to model an exploratory factor analysis have been fulfilled. In addition, another requirement for factor analysis is the ratio between the number of observations and number of variables.[49] The sample sizes are sufficiently large in all instances of this study.

Another requirement for factor analysis is an appropriate distribution of variables. The variables have to be roughly normally distributed (Field 2009: 650). This requirement is particularly challenging to fulfill with Likert scale variables. Investigating the distribution by inspecting Quantile-Quantile plots for each variable (graphics not shown here), some deviations from normal distribution can be observed; however, the overall im-

---

47 The Kaiser-Meyer-Olkin measure is based on the ratio of correlations among variables in the dataset to the partial correlation of variables within the dataset (Field 2009: 647).

48 This procedure tests whether the underlying matrix is an identity matrix, in which all variables do not correlate with each other (Field 2009: 648). Hence, a highly significant result is a desirable outcome as it indicates correlations among variables.

49 In general, it is advisable that at least 20 observations exist per variable. In this analysis, the number of observations range from each dataset between 1223 observations (Sigmar Gabriel) to 1629 observations (Gregor Gysi), 2446 observations (Angela Merkel) to 5695 observations (politicians in general).

pression is an acceptable approximation of normal distribution for the variables.

The number of factors that are retained can be determined by various measurements. Kaiser's eigenvalue-greater-than-one rule is one such widespread measure; however, it has also been widely critiqued over the last four decades and even more so in recent years (Patil et. al 2008; Zwick & Velicer 1982; Field 2009: 641). The general idea behind this measure is that one variable has an eigenvalue of one, and therefore a factor should capture at least as much variance as a single variable (Field 2009: 641).

The main critique regarding this measure is its tendency to over-extract factors and thereby fails to present parsimonious solutions (Patil et al. 2008: 163) and even worse, generates misleading factor solutions (Patil et al. 2008: 163).

One alternative or additional criterion to Kaiser's rule is the use of scree plots to determine the number of factors that are retained. The scree plot plots the eigenvalues of factors or components (y-axis) in relation to the number of factors (x-axis). A stark decline in the plot indicates the appropriate number of factors. However, while this measure is a preferable addition (Patil et al. 2008), the graphical interpretation can appear as a subjective decision in certain cases rather than a scientific endeavor (Patil et al. 2008: 164; Turner 1998).

A superior and certainly more scientific approach to the interpretation of a scree plot is facilitated by parallel analysis as introduced by Horn (1965) (Field 2009: 641; Patil et al. 2008; Sakaluk & Short 2017).[50] The scree plots with parallel analysis are shown in the online appendix for all politicians (see Figure A.1, Figure A.2, Figure A.3 and Figure A.4 in the online appendix). The analysis shows that a three-factor solution is sufficient for all specific politicians based on the factor analysis approach. When extracting principal components instead of factors, the principal component approach indicates that two components capture the variance within the variables for the three politicians – Angela Merkel, Gregor Gysi, and Sigmar Gabriel.

---

50 In a parallel analysis a scree plot is fitted to the data, while another scree plot is fitted to a dataset with random variables that are equivalent to the researcher's dataset in terms of the number of variables and number of observations. Typically, the process is iterated at least 100 times and the eigenvalues of the 95th percentile of the simulated data are displayed to avoid random biases in the simulated data.

The number of variables that were introduced into the factor analysis for politicians in general is slightly higher, because four additional variables were introduced – two items that relate to the external efficacy as well as two additional items on a semantic differential scale. For this model the number of suggested factors according to the parallel analysis is five factors or three components. In addition, it must be noted that parallel analysis tends to "overfactor" in samples with large sample sizes (Warne & Larsen 2014: 118-119). Therefore, another highly recommended approach is considered when determining the number of factors to retract is Velicer's Minimum Average Partial Factor Retention Method (MAP) (Patil et. al 2008). This method determines the number of factors to retract "by separating the common and unique variance and retaining only those factors that consist primarily of common variance" (Garrido et al. 2011: 553). This procedure can be slightly biased when ordinal scale variables are analyzed (Garrido et al. 2011).

According to the Minimum Average Partial Factor Retention Method, two factors should be retained for each politician – the same number of factors as the parallel analysis based on principal components. One main difference to the parallel analysis occurs for the politicians in general. According to the Minimum Partial Factor Retention Method, one factor is sufficient for the politicians in general, contrary to the three components that were suggested by parallel analysis. Trying to avoid over extracting of factors, the subsequent analysis focuses on two factors for each politician and a one-factor solution for politicians in general. Because the aim in this analysis is to find an underlying latent structure, a factor analysis instead of a principal component analysis is estimated as the next step. The exploratory factor analyses are estimated based on a maximum likelihood approach with an oblimin rotation that allows the factors to be correlated.

The following Table 13 presents the results of the factor analysis for each politician and politicians in general. The factor loadings indicate for each politician that the overall rating of the scalometer as well as the four items for candidate orientation form the first factor. The overall scalometer rating shows the highest factors loadings. The semantic differentials built a second factor, but the factor loadings are lower for this second factor. For politicians in general, a one-factor solution shows mostly acceptable factor loadings. The items "emotional–rational" and "calm–agitated" show the smallest factor loadings (0.04 and 0.19). Particularly, the item "emotional–rational" does not seem to belong to this one underlying construct.

Table 13: Parsimonious Factor Solutions for the Evaluation of Politicians

	Politicians in General	Angela Merkel		Gregor Gysi		Sigmar Gabriel	
	1	1	2	1	2	1	2
Scalometer	0.758	0.918		0.867		0.781	
Strong Leader	0.500	0.654		0.696	-0.164	0.740	-0.153
Likeable	0.538	0.757		0.779		0.687	0.162
Capable	0.692	0.820		0.782		0.861	
Trustworthy	0.762	0.840		0.849		0.797	0.112
Emotional-Rational		0.172	-0.216	0.167	-0.323		-0.162
Impolite-Polite	0.427		0.697	0.310	0.532	0.206	0.660
Calm-Agitated	0.185		0.539		0.563		0.474
Aggressive-Peaceful	0.283		0.735		0.852		0.842
Arrogant-Modest	0.565	0.242	0.564	0.371	0.491	0.348	0.461
Decisive-Unassertive	0.367						
Over-strained-Resilient	0.550						
Caring	0.644						
Contact Seeking	0.595						
Sum of Squared Loadings	3.997	3.332	1.686	3.442	1.710	3.188	1.687
Proportion of Variance	0.285	0.333	0.169	0.344	0.171	0.319	0.169
Cumulative Variance		0.333	0.502	0.344	0.515	0.319	0.487
Observations	5,695	2,446		1,629		1,223	

Note: Factor loadings are based on a maximum likelihood estimation with oblimin rotation.

Although the exploratory factor analysis has mostly yielded in two-factor solutions, these two dimensions did not clearly distinguish the two under-lying dimensions of warmth and competence. Rather, it reflects the two sets of measurements of candidate orientation as specific attributes and se-mantic differentials of more direct personal and observable attributions. Since some indicators, such as the parallel analysis, pointed towards three

underlying factors, exploratory three factor models are displayed in the following two tables (Table 14 and Table 15).

Table 14 shows that within a three-factor solution for politicians in general, the first two factors indicate two underlying factors of warmth and one of competence. The first factor has high ratings for the likeability and trustworthiness of politicians in general as well as their overall evaluation according to the scalometer rating. The two items measuring external efficacy also loaded onto the first factor as it could have been expected if the external efficacy is interpreted as political empathy (Kinder 1986: 241), which typically belongs to the dimension of warmth. Politicians' competence is reflected by the second factor which shows high values for leadership skills and problem-solving skills. In addition, two items that reflect performance-related evaluation such as decisiveness and resilience loaded onto this second factor. The third factor consists of several semantic differentials, which could reflect the politicians' level of arousal and their communication style. Compared to the two-factor solution, this three-factor solution has hardly improved the amount of cumulative variance that is accounted for by the three factors compared to a two-factor solution. While this three-factor solution indicates a fit that is more in line with theoretical expectations for politicians in general, the three-factor model for Angela Merkel shows a different factor pattern. Similar to the two-factor solution, the first factor shows the highest factor loadings regardless of whether items could theoretically be interpreted as belonging to the dimension of warmth or competence. Hence, the distinction between warmth and competence seems less pronounced for the evaluation of Angela Merkel.

Table 14: A Three-Factor Solution for Politicians in General and Angela Merkel

	Politicians in General			Angela Merkel		
	1	2	3	1	2	3
Scalometer	0.724			0.889		
Strong Leader		0.582		0.668		-0.311
Likeable	0.546			0.746		0.255
Capable	0.434	0.332		0.829		-0.207
Trustworthy	0.791			0.815		
Emotional–Rational	0.307	-0.215	-0.205	0.171	-0.221	0.286
Impolite–Polite	0.117	0.128	0.500	0.106	0.682	
Calm–Agitated	-0.162	0.231	0.365		0.546	-0.219
Aggressive–Peaceful			0.694		0.727	
Arrogant–Modest	0.438		0.284	0.244	0.569	
Decisive–Unassertive		0.431				
Resilient–Overstrained		0.603				
Caring	0.634					
Contact Seeking	0.577					
Sum of Squared Loadings	2.710	1.136	1.029	3.255	1.677	0.356
Proportion of Variance	0.194	0.081	0.074	0.325	0.168	0.036
Cumulative Variance	0.194	0.275	0.348	0.325	0.493	0.529
Observations	5,695			2,446		

Note: Factor loadings are based on a maximum likelihood estimation with oblimin rotation.

Table 15 examines the factor loadings for Gregor Gysi and Sigmar Gabriel. For Gregor Gysi, the factor loadings of a three-factor solution are similar to the three-factor solution that could be obtained for Angela Merkel. The first factor shows the highest loadings for the variables of candidate orientation and the highest values for competence. While some cross-loadings exists, the second factor could indicate an aroused communication style, while the third factor further exemplifies the underlying dimension of warmth with the high loadings for likeability. When estimating a three-factor solution, the evidence for two underlying factors is stronger for Sigmar Gabriel. The first factor shows high loadings for warmth, whereby the third factor shows high loadings for competence. The second factor captures the semantic differentials. Hence, these three-factor solutions partially support the notion of two underlying dimensions of warmth and competence, while the evidence is more ambiguous for Angela Merkel and Gregor Gysi. Furthermore, an additional third factor hardly increases the amount of explained variance across the models.

Table 15: A Three-Factor Solution for Gregor Gysi and Sigmar Gabriel

	Gregor Gysi			Sigmar Gabriel		
	1	2	3	1	2	3
Scalometer	0.521	0.109	0.392	0.603		0.265
Strong Leader	0.618	-0.112	0.111			0.746
Likeable	0.290	0.167	0.586	0.718		
Capable	0.931		-0.115	0.184		0.722
Trustworthy	0.643	0.103	0.229	0.590		0.293
Emotional–Rational		-0.325	0.336	0.385	-0.291	-0.261
Impolite–Polite		0.578	0.272	0.236	0.605	
Calm–Agitated	0.113	0.585	-0.137		0.528	0.174
Aggressive–Peaceful		0.871			0.836	
Arrogant–Modest	0.103	0.541	0.278	0.542	0.338	
Sum of Squared Loadings	2.056	1.897	0.862	1.770	1.550	1.352
Proportion of Variance	0.206	0.190	0.086	0.177	0.155	0.135
Cumulative Variance	0.206	0.395	0.481	0.177	0.332	0.467
Observations	1,629			1,223		

Note: Factor loadings are based on a maximum likelihood estimation with oblimin rotation.

While the assumptions for a factor analysis were largely fulfilled, the inclusion of semantic differential scales with dual endpoints could have impacted the factor solutions. One way to circumvent the use of these semantic differential scales is by conducting binary network models.

Recently, the use of network models has been suggested to model candidate evaluations (Dalege et al. 2016). In recent years, psychological networks have become a popular tool in psychology research to model attitude structures (Epskamp et al. 2018: 195). When fitting a network model, the variables are presented as nodes within a network graph, while the edges indicate the relationship between the variables (van Borkulo et al. 2014). Previous research has applied network models to the structure of attitudes and candidate evaluations regarding American politicians such as Ronald Reagan and his opponent in the election in 1984, Walter Mondale, using ANES survey data (Dalege et al. 2016). By doing so, Likert scale variables were recoded into dichotomous variables; therefore, violations against the requirement of normally distributed variables are not a concern for network models that are estimated by Ising Models, which are popular network model estimated by pairwise Markov Random Fields (Epskamp et al. 2018: 198). The Ising Model is appropriate and suitable for binary data (Dalege et al. 2016; Epskamp et al. 2018: 198; van Borkulo et al. 2014).

Since this procedure has been previously successfully applied to the field of candidate evaluations and concerns regarding the normal distributions

of each variable are diminished, a network model appears to be a worthwhile fit to the attitudes regarding German politicians. Furthermore, such network models do not only allow investigating of the structure of all variables, they also enable the researchers to detect changes in the relationships among variables that occur over time (Epskamp et al. 2018). In the case of this experiment, network models can also be estimated after the experimental treatment had occurred, to examine if and how the variables are related differently to each other once participants were exposed to the politicians in the video clips.

One requirement for the fit of a network model is a sufficient sample size. With each additional variable that is entered into the network model, the number of parameters that need to be estimated increases.[51] Contrary to most research in psychology, the sample size of this study is large enough so that the number of parameters that need to be estimated is not concerning. The large sample sizes have another advantage of improved estimations: "With increasing sample size, the parameters will be more accurately estimated from data" (Epskamp et al. 2018: 196).

The following network models mostly adheres to the model specifications proposed by Dalege et al. (2016) for the evaluation of presidential candidates. It is adjusted slightly to indicate the communities to which each variable belongs, similar to a latent class, underlying dimension or cluster (e.g., Bell & O'Driscoll 2018). As suggested by Dalege et al. (2016), the Likert scale variables are transformed into binary variables. The specific items of candidate orientation are coded in a way that they indicate the positive attribution of a characteristic: If participants indicated that they agreed or strongly agreed with a statement, the variable was coded as "1" and otherwise as "0". The semantic differentials were recoded so that the two highest or lowest end points, respectively, were coded as "1" and otherwise as "0" to indicate the presence of a certain attribution. For the overall rating (scalometer), the four highest values on an 11-point Likert scale

---

51 To give an example, in the case of fifteen variables that are estimated within a 15-node-network model for each of the three politicians, 120 parameters have to be estimated: 15 threshold parameters + 15 x 14/2 pairwise association parameters (see Epskamp et al. 2018: 198). In the case of politicians in general, the number of estimated parameters is noticeably higher. Within a 21-node-network model, 231 parameters need to be estimated: 21 threshold parameters + 21 x 20/2 pairwise association parameters.

are chosen as indication of a positive overall evaluation of each politicians and politicians in general.[52]

The following section displays network graphs for the three politicians and politicians in general. The models are fitted using the recommended Ising Model (van Borkulo et al. 2014). The network models for Angela Merkel, Sigmar Gabriel, and Gregor Gysi show similar network structures. The communities based on a spinglass community algorithm and the median number of clusters after 1000 repetitions indicate three communities within the networks for political leaders (see also Bell & O'Driscoll 2018; Fried 2016). The first community consists of the four items that measure the candidate orientations traditionally as well as the overall rating (scalometer). The next two communities, or latent classes, exist within the subgroup of semantic differentials, whereby positive evaluations such as calmness, modesty, and politeness form one community and their negations form the second cluster as they co-occur. By inspecting the edges between the nodes, the strength and nature of the relationship between the variables can be investigated. The stronger the edge, the stronger the association between both variables. A green edge indicates a positive association, red edges indicate negative relationships. The divide is partially a result of measuring specific character traits as semantic differentials instead of using unidimensional Likert scales. The associations between semantic differentials and items that measure candidate orientation are weak; however, for Angela Merkel; it is noticeably that arrogant evaluations are linked to overall evaluations and evaluations of trustworthiness, problem-solving and leadership skills (see Figure 10). The absence of arrogance and evaluation of modesty is positively linked to trustworthiness, overall ratings and likeability.

For Gregor Gysi, the same amount of communities occurs (see Figure 11). A strong link between his overall ratings and likeability occurs, while his leadership skills and likeability indicate only a weak positive relationship. Focusing on the link between semantic differentials and items of candidate orientations, evaluations of modesty and politeness relate most strongly to the candidate orientation. Positive evaluations of politeness and

---

52 The network models are based on the pre-test (Wave 1). In a later step, network models were also fitted for the anger condition of the post-test survey – the coding scheme had to be altered slightly in singular instances because not enough observations had occurred in the single experimental group. In these cases, a third answer category from the end point of the semantic differential scale closer was also coded as indication of a certain characteristic.

154

likeability co-occur, while a positive evaluation of his modesty is associated with a favorable overall rating and a favorable evaluation of his trustworthiness.

For Sigmar Gabriel, the same amount of communities occurs (see Figure 12). The semantic differentials that are particularly linked to his more general evaluation are his arrogance, politeness and calmness. Evaluations of arrogance have a negative relationship with overall ratings, trustworthiness, and problem-solving skills. Those participants who perceive him as being polite are also more likely to perceive him as being trustworthy, likeable, and favor him more strongly overall. Furthermore, evaluations such as calm have a positive relationship with his trustworthiness.

For politicians in general, the structure is slightly different (see Figure 13). The two items that measure external efficacy did not connect to the other variables, and correlated only weakly with them based on this binary specification. Hence, communities could only be calculated for the connected items, which resulted in a solution with two communities. The rest of the network model shows weak correlations that are weaker compared to the previous three models. In addition, semantic differentials and candidate orientations do not cluster together into two separate clusters; instead, two communities occur, whereby positive and negative evaluations can be distinguished from each other.

Figure 10: Network Graph of Angela Merkel's Character Traits Assessments
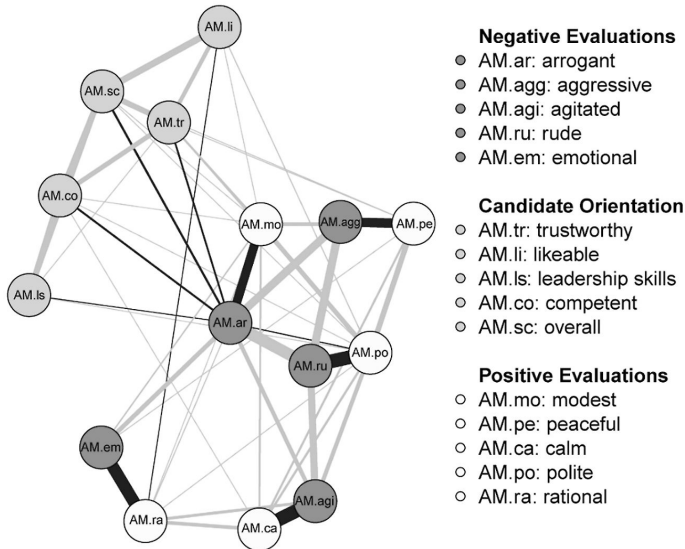
Negative Evaluations
- AM.ar: arrogant
- AM.agg: aggressive
- AM.agi: agitated
- AM.ru: rude
- AM.em: emotional

Candidate Orientation
- AM.tr: trustworthy
- AM.li: likeable
- AM.ls: leadership skills
- AM.co: competent
- AM.sc: overall

Positive Evaluations
- AM.mo: modest
- AM.pe: peaceful
- AM.ca: calm
- AM.po: polite
- AM.ra: rational

Note: Positive associations between items are displayed in light gray, while negative associations between items are displayed in dark gray.

Figure 11: Network Graph of Gregor Gysi's Character Traits Assessments
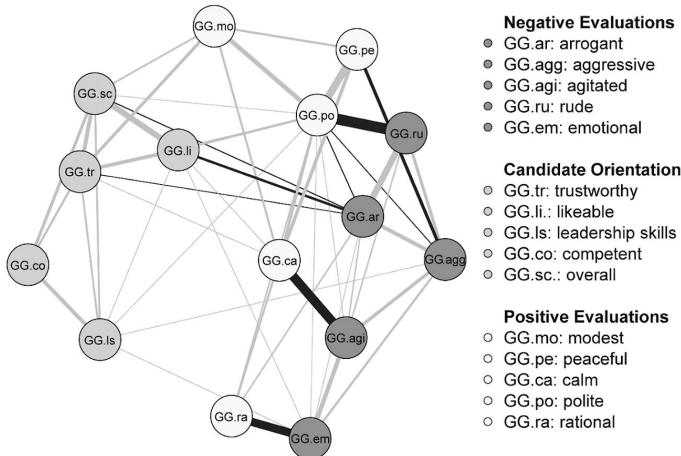
Negative Evaluations
- GG.ar: arrogant
- GG.agg: aggressive
- GG.agi: agitated
- GG.ru: rude
- GG.em: emotional

Candidate Orientation
- GG.tr: trustworthy
- GG.li.: likeable
- GG.ls: leadership skills
- GG.co: competent
- GG.sc.: overall

Positive Evaluations
- GG.mo: modest
- GG.pe: peaceful
- GG.ca: calm
- GG.po: polite
- GG.ra: rational

Note: Positive associations between items are displayed in light gray, while negative associations between items are displayed in dark gray.

Figure 12: Network Graph of Sigmar Gabriel's Character Traits Assessments

Negative Evaluations
- SG.ar: arrogant
- SG.agg: aggressive
- SG.agi: agitated
- SG.ru: rude
- SG.em: emotional

Candidate Orientation
- SG.tr: trustworthy
- SG.li.: likeable
- SG.ls: leadership skills
- SG.co: competent
- SG.sc: overall

Positive Evaluations
- SG.mo: modest
- SG.pe: peaceful
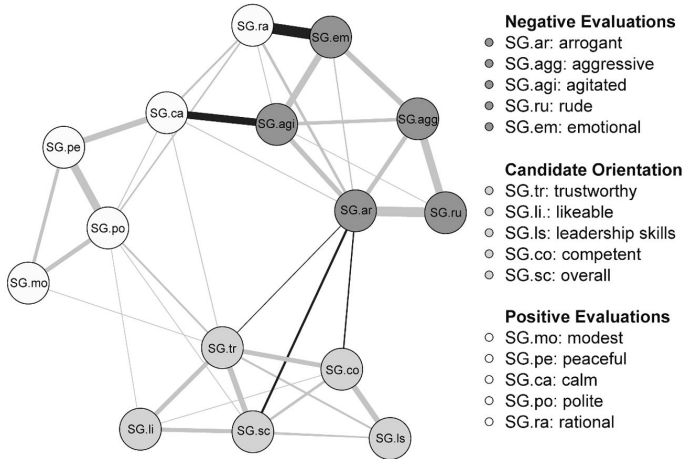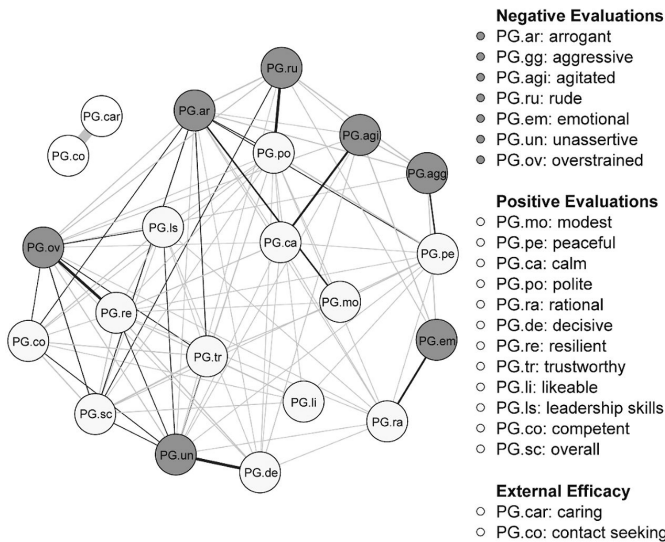- SG.ca: calm
- SG.po: polite
- SG.ra: rational

Note: Positive associations between items are displayed in light gray, while negative associations between items are displayed in dark gray.

Figure 13: Network Graph of Politicians' Character Traits Assessments

Negative Evaluations
- PG.ar: arrogant
- PG.gg: aggressive
- PG.agi: agitated
- PG.ru: rude
- PG.em: emotional
- PG.un: unassertive
- PG.ov: overstrained

Positive Evaluations
- PG.mo: modest
- PG.pe: peaceful
- PG.ca: calm
- PG.po: polite
- PG.ra: rational
- PG.de: decisive
- PG.re: resilient
- PG.tr: trustworthy
- PG.li: likeable
- PG.ls: leadership skills
- PG.co: competent
- PG.sc: overall

External Efficacy
- PG.car: caring
- PG.co: contact seeking

Note: Positive associations between items are displayed in light gray, while negative associations between items are displayed in dark gray.

4.4.2 Measurements of Covariates and Moderating Variables

This section presents the measurement of moderating variables which are tested in the following analysis to test heterogeneous treatment effects.[53] Based on the previously discussed theoretical assumptions, the impact of the experimental treatment could potentially be moderated by a range of variables. A moderation effect implies that the experimental treatment effect has a heterogeneous impact on the viewers, whereby the treatment effect varies by subgroups, which means that the treatment effect might be stronger or even only occurring for participants with a set of certain characteristics, such as sociodemographic properties or individual attitudes. These moderating effects can be analyzed by using interaction terms between the independent variable (the experimental treatment) and the moderating variable.[54] Because most of the tested moderator variables rely on self-reports, additional control variables are also taken into consideration to test for potential confounding factors.

Previous research has found that principles of homophily guide individual perceptions of political candidates. In general, voters tend to favor political candidates who are more similar to themselves in regard to age, gender, facial features, and ethnicity (Bailenson et al. 2008; Masters 1994; McDermott 1998). An individual's ethnicity cannot be considered for this study, since the survey questionnaire did not include a relatable question.[55] This analysis, however, controls for the gender and age of participants. Age

---

53 A mediation analysis, which could test whether the impact of the experimental effects is mediated by individual characteristics has not been conducted. Mediation analysis can introduce biases into the analysis if the mediating variable is not manipulated (Gerber & Green 2012: 321). The problem of attributing causality to factors that have not been randomly manipulated has also been expressed by Holland (1986) with his famous motto "no causation without manipulation" (Holland 1986: 959). Due to this skepticism towards the use mediating variables within the social sciences, a mediation analysis is not conducted.

54 Ideally, instead of using moderating variables, it is advisable to adjust the experimental design by using techniques such as blocking, if strong theoretical or empirical evidence exists (Gerber & Green 2012: 109–115).

55 Ethnicity is a sensitive topic in German surveys and not frequently measured in surveys unlike in English-speaking countries, such as the United Kingdom and the USA (e.g. British Social Attitudes Survey or the General Social Survey). Both of these countries are ethnically diverse. In a module for the International Social Survey Programme (ISSP), German participants were asked whether they identified with a "population group" instead of asking them about their race or ethnicity more directly (ISSP 2013/2014, ALLBUS 2014 (GESIS 2015)).

is measured as a mean-centered variable (M = 44.82, SD = 14.65). In addition to age and gender, an individual's party identification is controlled for, since it has been shown that party identification is a crucial category in the perception of political candidates (Rahn 1993).

In addition, some research has pointed out that particularly uninformed or less sophisticated voters rely on candidate appearances as heuristics (Lau & Redlawsk 2001). The political interest is used as an approximation of political sophistication. Hence, this study also controls for individual levels of political interest, whereby this 5-point Likert scale (ranging from "0" – not at all" to "4 – very strongly") is used as a mean-centered variable (M = 2.16; SD = 0.90 in Wave 1).

Education is not used to approximate sophistication for the models that are reported in the following. Education was measured only in relation to schooling without regard to college-education. Hence, education could only be used as a categorical variable of basic, medium and high levels of schooling in accordance with the three-tiered German school system.[56] However, education is not included in the presented models due to constraints based on the low number of participants, if party identification for the Left is also taken into consideration. Nonetheless, all models have also been conducted while controlling for educational level. These models did not indicate that the level of education acted as a confounding variable that changed the relationship between the experimental treatment, moderating factors, and changes in the assessment of political leaders measured as gain scores. An example can be found in the appendix (see Table A.31, Table A.32, Table A.33). Subsequently, the following variables are tested as moderating variables: party identification, personal predispositions and the order of video clips.

First, the order in which the politicians were seen is explored, since this factor was implemented by the experimental design as a random rotation of the order in which the two video clips (Merkel/Gysi and Merkel/Gabriel) were seen. This measurement is followed by individual predispositions such as party identification and further individual characteristics such as ideological beliefs, and personality factors. Party identification was measured by a standard question of German electoral research (for com-

---

56 A basic level of education stands for a certificate after nine years of schooling ("Hauptschulabschluss"), a medium level of education stands for a certificate after ten years of schooling ("Realschulabschluss") and a high level of education equals a German university entry-diploma ("Abitur") after twelve or thirteen years of schooling.

parison see Falter 1977; Rattinger et al. 2017), which can be translated as follows:

> "In Germany, many people lean towards a certain political party over time, although they occasionally vote for a different party. What about you: Generally speaking – do you lean towards one particular party? If so, which one?"

Participants could answer with one of the four parliamentary parties at the time – CDU/CSU, SPD, the Greens, and the Left – an open-ended response as well as checking "no party". Instead of further differentiating the strength of party identification, a dichotomous approach has been chosen for this study, whereby participants who identify with one party are coded as identifiers (1) and all other participants, whether they identified with any other party or no party at all, were coded as non-identifiers (0).

In this study, the ideological self-positioning was measured via two items which represent the cultural and socioeconomic dimensions of left and right. This study did not include an item which indicates self-positioning on an explicit left-right continuum. Both items were measured by an 11-point Likert scale, as is common (e.g., Rattinger et al. 2017). The cultural dimension was measured with an item reflecting support for immigration ranging from "immigration for foreigners should be made easier" (1) to "immigration for foreigners should be made more difficult" (11). The socioeconomic dimension of left and right was measured with an item focusing on participants' attitudes towards and support of taxation, ranging from "less taxes, even if it results in less social benefits" (1) to "more social benefits, even if it results in more taxes" (11). For the analysis, both items were rescaled so that they ranged from 0 to 10. Furthermore, the socioeconomic item was recoded so that it reflected the left-right scale with lower values for more social benefits and higher values for less taxation. After rescaling both items only correlated weakly ($r = 0.24$); similarly, Cronbach's alpha showed a low construct validity (alpha = 0.37). Hence, this following analysis only focused on a single item regarding the socioeconomic dimensions of the left-right scale, as this dimension has traditionally been the origin of left and right.

Two personality factors – neuroticism and extraversion – were measured with two items each according to a Big Five Short Scale (Rammstedt et al. 2013). The items were measured on a 5-point Likert scale and partially recoded before mean indices were computed, whereby high values represent

high values for neuroticism and extraversion, respectively. The items ranged from "completely agree" (4) to "completely disagree" (0).[57]

When moderation effects are considered as part of the analysis, the dependent variable of an overall assessment (scalometer) as well as two indices of warmth and competence were considered. Warmth and competence were measured as two mean indices, whereby likeability and trustworthiness measure warmth and strong leadership and political capabilities measure competence. The mean indices were then divided by the respective number of items each. The following Table 16 shows Cronbach's alpha values for each pairing based on measurements in the pre-test condition. Overall, Cronbach's alpha is mostly acceptable in all instances, given that it was calculated based on only two items. Only the reliability measures for politicians as a social group are less than desirable.

Table 16: Overview of Reliability Measures Based on Cronbach's Alpha

	Politicians in General	Angela Merkel	Gregor Gysi	Sigmar Gabriel
Warmth	0.64	0.82	0.82	0.81
Competence	0.62	0.77	0.71	0.78
Observations	5,704	2,450	1,631	1,226

Note: Cronbach's alpha is calculated based on the pre-test measurements in the first wave.

4.5 Manipulation Checks of the Experimental Treatment

This chapter presents manipulation checks of the experimental treatment and considers the external validity of this study. The first section looks at the self-reported elicited emotions and perceptions of the emotions expressed by the politician; thereafter, the external validity of the treatment conditions is analyzed by determining these conditions more thoroughly regarding the emotional expressions as well as a more in-depth analysis of their verbal contents.

---

57 Neuroticism was measured by the following two items:"I see myself as someone who is relaxed, handles stress well." (-) and "I see myself as someone who gets nervous easily." (+); the first item was recoded. Extraversion was measured by the following items: "I see myself as someone who is outgoing, sociable." (+) and "I see myself as someone who is reserved." (-); the second item was recoded.

4.5.1 Self-Reported Affective and Cognitive Responses

Cognitive and affective responses are two potential mechanisms which can have a simultaneous and interactive impact on the processing of political information such as emotional expressions by politicians (Redlawsk & Pierce 2017: 425). In the following sections, cognitive responses, such as conscious perceptions, and conscious affective responses are considered.

4.5.1.1 Perceptions of Emotional Expressions

Previous studies have found that self-reports of feelings as well as emotional expressions of others suffer severely from measurement errors in surveys (Scherer 2005; Siegert et al. 2011; Marcus et al. 2017). The problem of relying on self-reported emotional responses after experimental treatments of emotional expression by politicians has been noted before (Sullivan et al. 1991: 201). This concern in survey research is further amplified by the theoretical assumption that respondents do not necessarily have to be consciously aware of the emotion expression in order to form an impression, since these judgements and inferences can be made rapidly and automatically, and pre-consciously (Todorov et al. 2005, Brader & Marcus 2013: 171–174; Fiske & Taylor 2017: 70). Despite these concerns, this section presents the manipulation checks that were administered directly after each video clip regarding the perceptions of emotional expressions.

The perception of emotional expressions was designed as a treatment check that was administered as a first set of questions immediatelyafter the experimental treatment – the video clips – were administered. First, participants were asked whether they had the impression that the displayed politician(s) showed feelings, followed by a multiple choice question in which participants could choose among up to six different discrete emotions as possible impressions they had gained. The item wording of these discrete emotions was designed as adjectives to facilitate the self-report of emotions, as previous research in psychology has used adjectives rather than nouns to measure such self-reports, for example the Positive and Negative Affect Schedule (PANAS) (Watson et al. 1988; Krohne et al. 1996). Adjectives might be closer to everyday language than nouns and therefore potentially aided the measurement of emotional expressions. Furthermore, scholars have suggested to at least use two to three synonyms when measuring emotions (Brader & Marcus 2013: 188), however, for reasons of parsimony in online surveys, the measurement was restricted to one adjective

162

per discrete emotion. Out of the six discrete emotions three emotions had a positive valence, and three emotions a negative valence in order to avoid any biases: anger, sadness, happiness, pride, indignation, and amusement.

The following Table 17 shows the percentages of participants who perceived anger and/ or indignation within the experimental treatment conditions that featured male and female politicians. Within the negative-active treatment conditions more than six out of ten participants recognized anger and/or indignation as such for male and female politicians. A few participants also recognized these emotional expressions within the neutral condition – 14.4 percent out of those participants who saw male politicians and 9.1 percent of participants who saw female politicians. However, the percentages are particularly low for those who saw positive emotional displays – 0.6 percent and 1.5 percent, which by and large indicates a positive outcome of this treatment check.

Table 17: *Percentages of Perceived Discrete Emotions for Politicians as Social Groups*

Politicians	Treatment	Discrete Emotions				
		Anger		Indignation		
		Percent	N	Percent	N	N(Group)
Male Politicians						
	Positive	1.47	5	2.35	8	340
	Negative-Active	62.86	220	64.00	224	350
	Negative-Passive	21.33	74	11.82	41	347
	Neutral	14.41	51	14.12	50	354
Female Politicians						
	Positive	0.56	2	0.56	2	358
	Negative-Active	62.87	210	65.27	218	334
	Negative-Passive	27.98	101	30.75	111	361
	Neutral	9.06	31	11.11	38	342

Note: Cells display frequencies and percentages of perceived expressions of anger and indignation for each experimental treatment group.

Table 18 displays the results for the three political leaders. An equal number of participants recognized Gregor Gysi's anger (61.5 percent), while almost three quarters of participants recognized his indignation. The experimental treatment for Angela Merkel was recognized by roughly more than a third of participants as anger (35.0 percent; 33.6 percent), with slightly higher percentages for her negative-active expressions as indignation (37.7 percent; 36.9 percent). Sigmar Gabriel's expressions of negative-active emotions were also only perceived as anger by slightly more than a third of par-

ticipants (36.3 percent) and hardly perceived as indignation by roughly more than a quarter of the participants (28.6 percent). Indignation as an emotional expression is more strongly related to moral anger compared to incivility, hence, this finding could be explained by the different sides of anger (Hess 2014).

Table 18: Percentages of Perceived Discrete Emotions for Political Leaders

Politicians	Treatment	Discrete Emotions				
		Anger		Indignation		
		Percent	N	Percent	N	N (Group)
Merkel						
With Gysi	Positive	0.61	2	0.30	1	328
With Gysi	Negative	35.00	119	37.65	128	340
With Gabriel	Negative	33.63	113	36.90	124	336
With Gysi	Neutral	3.00	10	6.61	22	333
With Gabriel	Neutral	4.07	14	4.07	14	344
Gysi						
	Positive	2.13	7	2.74	9	328
	Negative	61.47	209	74.41	253	340
	Neutral	24.32	81	26.13	87	333
Gabriel						
	Negative	36.31	122	28.57	96	336
	Neutral	4.07	14	5.23	18	344

Note: Cells display frequencies and percentages of perceived expressions of anger and indignation for each experimental treatment group.

In addition to the perception of emotional expressions, participants also indicated whether they felt affected in any way by the experimental treatments. The next section covers whether further differences between the anger conditions can be discovered.

4.5.1.2 Affective Responses Based on Emotional Expressions

Since emotional expressions can potentially evoke an affective response within others, this section describes the self-reported affective impact that the video clips had on viewers. The affective response towards the video clips was measured at the end of the questionnaire in the last question that asked whether participants felt emotionally affected by the video clip. Participants could answer this question with a simple yes/no followed by a 5-point Likert scale that assessed the strength of the affective response; however, the valence of the affective response was not measured.

164

Table 19 presents the percentages and frequencies of those participants within each experimental condition who felt an affective reaction towards the videos. The experimental groups that focused on politicians as social groups show overall low levels of affective responses. For the groups with male politicians almost one in five participants (18.9 percent) and almost one in four participants for the group with female politicians (24.3 percent) felt an emotional reaction due to the anger conditions. Similarly, only slightly more than one in five participants (22.9 percent) felt affected by the video clips of Angela Merkel's and Sigmar Gabriel's negative-active expressions. In contrast, almost half of the participants who saw Merkel's and Gysi's negative-active expressions reported an affective response (48.5 percent). The percentage is equally high for those who saw their positive emotions (45.7 percent) and a small percentage of participants even felt an emotional reaction due to the neutral displays of both politicians (17.7 percent).

Table 19: Percentages of Affective Responses to Anger Expressions

Politicians	Treatment			
		Percent	N	Group(N)
Male Politicians				
	Positive	17.35	59	340
	Negative-Active	18.86	66	350
	Negative-Passive	11.24	39	347
	Neutral	4.52	16	354
Female Politicians				
	Positive	29.05	104	358
	Negative-Active	24.25	81	334
	Negative-Passive	17.17	62	361
	Neutral	4.09	14	342
Treatment E2/ E3				
Merkel/Gysi	Positive	45.73	150	328
Merkel/Gysi	Negative-Active	48.53	165	340
Merkel/Gabriel	Negative-Active	22.92	77	336
Merkel/Gysi	Neutral	17.72	59	333
Merkel/Gabriel	Neutral	9.88	34	344

Note: Cells display frequencies and percentages of self-reported affect for each experimental treatment group.

Despite the difficulties in measuring self-reported emotions and perceptions of emotional expressions these treatment checks indicate that Gregor Gysi's anger had been noticeably perceived by a larger number of participants in the negative-active treatment conditions, as participants reported a higher affective response towards his (and/or Merkel's) expressions of

anger and indignation. The failure to replicate a similar number of partici-pants who recognized Merkel's emotional expressions when also seeing Sigmar Gabriel illustrates that these differences can be attributed to Gysi's displays of anger and indignation. These differences could reflect the vary-ing quality of the real-world video material, whereby politicians have indi-vidually varying levels of expressivity. In order to assess the differences in emotional expressivity for each anger condition, the next chapter presents findings from a classification attempt of facial micro expressions.

4.5.2 The Facial Expressions of Anger

The first section discusses the validity of the emotional expressions and thereby the internal validity of the experimental design. For this study, emotional expressions were defined as being expressed either verbally, fa-cially, gesturally, or vocally. While emotions can be expressed in various ways, strong emotional expressions should be expressed via more than one channel. Anger is expressed by a vocal intensity of a raised voice, gestures, possibly verbal expressions, and corresponding facial expression. Previous studies that have focused on the impact of emotional expressions, particu-larly in non-verbal communication, have first and foremost focused on fa-cial expressions. This comes as no surprise, considering the dominant role of facial expressions in interpersonal communication: "Faces are intrinsi-cally the focus of attention in any social interaction" (Fiske & Taylor 2017: 65). Therefore, a vast majority of studies have adhered to Ekman's Facial Action Coding System (FACS) to determine the emotional valence of fa-cial expressions (e.g., Stewart & Ford Dowe 2013; for an overview: Stewart et al. 2009b).

  For this study the emotions within the experimental treatment were coded by the researchers of the research project, who decided unanimously that anger expressions were displayed within the selected video clips (see also Subchapter 4.3). Since the researchers were highly involved within the processes of data collection and emotion recognition, it seems worthwhile to assess the validity of this researchers' coding externally. Recent develop-ments in machine learning have led to algorithms that can detect discrete emotions within facial expressions due to training data and decision rules such as Ekman's FACS (e.g., Microsoft Cognitive Services).[58]

---

58 The algorithm built by Microsoft Cognitive Services was accessed via an API. For further information on the available Microsoft algorithms for face detections and

In order to classify the facial expressions within the experimental treatment, the following detection of emotional expressions is based on the algorithm built by Microsoft Cognitive Services. The video clips for each experimental treatment were classified by video frames, whereby each second of the videos resulted in 24 frames. When analyzing the video clips for Angela Merkel, Sigmar Gabriel, and Gregor Gysi, those frames that showed other politicians were not included in the analysis. Some of the frames could not be classified because the algorithm could not detect faces from certain angles. This number of misclassified frames was higher in the experimental groups with several politicians. 9.1 percent of frames with female politicians and 15.0 percent of frames with male politicians could not be classified, compared to the videos of political leaders – only 3.3 percent of frames could not be classified for Merkel, 2.0 percent for Gysi, and 0.1 percent for Gabriel. The algorithm detected the following emotions: anger, contempt, disgust, fear, happiness, sadness, and surprise as well as neutral expressions.[59] The number of frames per video clip ranged from 2303 to 2854 for the five different groups (for summary statistics of each experimental group see Table A.2, Table A.3, Table A.4, Table A.5, Table A.6 in the online appendix).

The classification of emotional expressions across all frames for each anger treatment is presented in the following Figure 14, which displays grouped boxplots for the classification of eight emotional expressions within each anger treatment. The grouped boxplots show that in each experimental condition, the vast majority of expressions were most likely neutral expressions. The x-axes show the classification of emotions ranging from 0 to 1, which can be interpreted as the estimated probability for each emotion (multiplied by 100).

From a theoretical standpoint, the classification of anger expressions is of the highest interest. Based on the interquartile ranges (or the boxes), three-quarters of the frames were classified with very low probabilities for

---

emotion classifications see the description by Microsoft Cognitive Services, online available at: https://azure.microsoft.com/en-us/services/cognitive-services/ (last accessed: 09 January 2020).

59 Microsoft Cognitive Services provide more information, available online at: https://docs.microsoft.com/en-us/dotnet/api/microsoft.azure.cognitiveservices.vision.face.models.emotion?view=azure-dotnet (last accessed: 04 May 2020). When accessing the algorithm via an API, the classifications of disgust and contempt were labelled as being under development. Since contempt and disgust were still in the experimental development stage, their respective classifications are not discussed in further detail.

facial expressions of anger. All boxplots show that high probabilities for anger expressions were outliers; therefore, only a few frames within the anger expressions were classified as likely showing anger based on the politicians' facial expressions. This is not necessarily unexpected in longer video clips, which also show frames leading up to an expression of anger. Furthermore, this finding might partially reflect the discourse in German politics at the time of the data collection, which has been at times described as tame and consensual (e.g., Tenscher 2013; Zimmermann 2014), and could be seen as less emotionally driven and more rational in a cross-cultural comparison.

Figure 14: External Classification of Facial Expressions Within the Anger Conditions
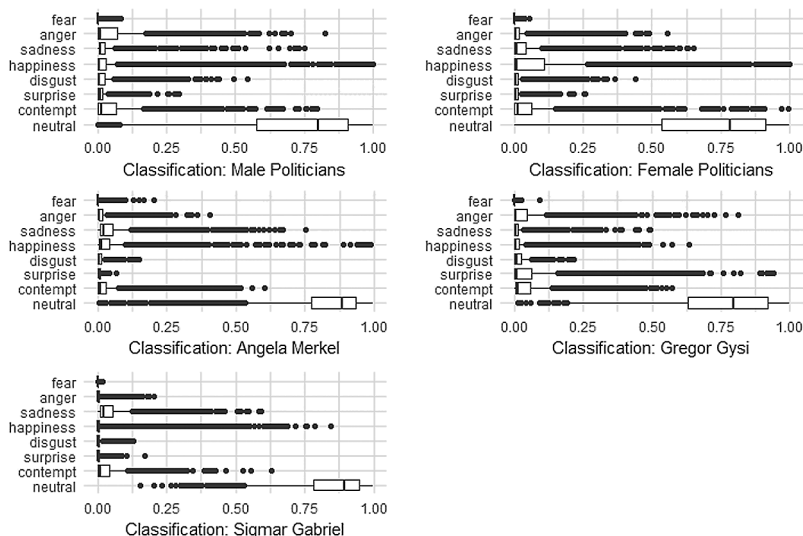
Note: Figure based on author's own illustration. Each boxplot displays the classifications of video frames as probabilities for emotion expressions ranging from 0 to 1. *Source*: Microsoft Cognitive Services API.

Nonetheless, the experimental anger conditions for male politicians and Gregor Gysi contained 22 and 26 frames, respectively, for which displays of anger were predicted with an above 50 percent probability. Only one frame within the experimental treatment condition for female politicians had a likelihood of displaying anger above 50 percent and no frames with

such a high probability were classified for Angela Merkel or Sigmar Gabriel. Then Finance Minister Wolfgang Schäuble (CDU), at 82.3 percent, displayed the highest classification for the experimental group showing male politicians. Within the group of female politicians, a frame showing Rebecca Harms (the Greens) had the highest prediction of showing anger at 55.6 percent. Judging from the maximum of predicted anger, the expressions of the three leading politicians differ: Gregor Gysi had the highest prediction at 81.1 percent, while Angela Merkel's maximum prediction for anger reached 40.6 percent, and anger was only predicted with a chance of 20.3 percent for Sigmar Gabriel's displays of anger. This could indicate differences in their emotional expressiveness, at least based on the selected video material (see Figure 15).

Figure 15: Strongest Anger Expression Within the Experimental Groups of Male Politicians, Female Politicians, Angela Merkel, Gregor Gysi, and Sigmar Gabriel

Note: The figure shows the frames with the highest anger expressions within the experimental treatment condition of male politicians (Wolfgang Schäuble), female politicians (Rebecca Harms), Angela Merkel, Gregor Gysi and Sigmar Gabriel (from top left to bottom right).

4.5.3 Political Issues and Anger Expressions

In order to investigate differences that might have occurred due to different verbal expressions of anger and related political issues or other politicians, the verbal content of the video clips has been transcribed. Since anger is an emotion that highly depends on its context (Hess 2014, Knutson 1996: 177–179), the topics mentioned within the video clips should be considered in order to assess varying effects. If anger is assessed as appropriate, for example in cases in which anger is directed towards injustices or less than ideal states of affairs, it has a positive connotation of change that is intended by the person who addresses the underlying issues. The negative side of anger emerges if individuals are attacked personally and uncivil behavior occurs. This kind of anger can be perceived as inappropriate. Based on the theoretical assumptions about the two sides of anger (Hess 2014), the content of the experimental treatment is investigated in more detail. This in-depth analysis of the experimental anger conditions can shed some light on the hypothesis H3 that the political messages of the treatment present a crucial contextual factor for the evaluation of politicians' anger expressions.

All video clips had a length of roughly 1.5 minutes. All video clips showing negative-active emotions shared a similar amount of words (see Table 20). After cleaning each corpus by deleting the most common words within a language, such as articles and pronouns, with a low informational value, the number of words ranged from 99 (for female politicians) to 116 (Gysi and Gabriel) per video clip.

Table 20: Number of Words Within the Negative-Active Treatment Conditions

Politicians	Number of Words Without Stop Words	Number of Words with Stop Words
Female Politicians	99	281
Male Politicians	109	291
Angela Merkel	100	270
Gregor Gysi	116	323
Sigmar Gabriel	116	340

Note: Frequencies are based on author's own calculations according to transcriptions of the video clips.

To gain some further insight into potential differences between the video clips, the topics within the negative-active treatment were coded according to categories implemented by Rattinger et al. (2015d) in the GLES media campaign study, as shown in Table 21. Before going into detail about the

topics mentioned in each video clip, it can be seen that the experimental groups that saw video clips of male and female politicians saw a higher number of shorter video sequences compared to the groups that saw political leaders. This difference is particularly visible for the groups of female and male politicians, which had twice as many video sequences as the group that featured Sigmar Gabriel and almost twice as many as the group featuring Angela Merkel. The number of sequences within a video clip for Gregor Gysi falls between those extremes with 13 sequences. In general, shorter sequences are likely to contribute less context, while longer-lasting sequences provide more context, which could be used for an appraisal of the scene.

Table 21: Number of Topics Mentioned Within the Negative-Active Treatments

Politician	Treatment	Topics	Number of Topics Mentioned
Gysi	Negative -Active		
		Election Campaign	1/13
		Social Policies	6/13
		Labor Market Policy	3/13
		Fiscal Policy	3/13
Merkel	Negative -Active		
		EU Politics	1/9
		Rule of Law	1/9
		Labor Market Policy	2/9
		Economic Policy	2/9
		Fiscal Policy (Euro Crisis)	2/9
		Election Campaign	1/9
Gabriel	Negative -Active		
		Fiscal Policy	2/8
		Labor Market Policy	1/8
		Economic Policy	1/8
		Social Policy	1/8
		Election Campaign (Party Politics)	3/8

Politician	Treatment	Topics	Number of Topics Mentioned
Male Politicians	Negative -Active		
		Economic Policy	2/17
		Infrastructure	2/17
		Not Identifiable	3/17
		Fiscal Policy	2/17
		Election Campaign (Party Politics)	6/17
		Social Policy	1/17
		International Conflicts	1/17
Female Politicians	Negative -Active		
		Fiscal Policy	5/18
		Social Policy	2/18
		Defense Policy	2/18
		Election Campaign (Party Politics)	2/18
		Environmental Policy	1/18
		Social Policy	3/18
		EU Policy	1/18
		Infrastructure	2/18

Note: The last column indicates the number of video sequences that mentioned a topic within the video clips. One topic is coded per video sequence. The codings are based on transcriptions of the video clips.

In addition to the topics within the anger condition, the most frequent words with a frequency larger than two are displayed in the next section (see Figure 16). Overall, the politicians only mention a few words more than once;[60] nonetheless, differences occurred between the politicians. Most noticeably, Gregor Gysi's focus on social policies becomes apparent even from a few words, which focused on education and ensuring an equal education across the sixteen German federal states. Furthermore, his anger was directed towards unequal tax burdens across the social strata by pointing out a perceived societal mismatch between "rich managers who earned billions" and the working class,[61] while a minimum wage had not been

---

60 This refers to the root of a word, since the word clouds were produced after stemming and then completed for the word clouds to the most frequent realization of the stemmed word.

61 "Aber es is bei uns doch völlig maßlos geworden. Der braucht keine Milliarde, das ist doch albern. Aber andere wissen nicht, wovon sie sich die nächste Speise kaufen sollen. Das geht einfach nicht."

172

implemented in Germany at the time of the data collection.[62] He also added that he could not understand why the current tax system favored him as a member of parliament and other civil servants in regard to retirement payments for social security compared to ordinary employees. That was something he disagreed with and could not explain to "any cashier at Lidl".[63] Therefore, his anger was even directed at a policy that was not disadvantageous to him personally.

In a stark contrast, social policies were less frequently mentioned by Sigmar Gabriel. Although he mentioned issues such as the potential loss of workplaces and a proposition to increase the highest taxation rate, the most frequent words signal personal attacks – even when talking about these political issues.[64] In one video sequence, he said to Ursula von der Leyen, CDU, "to be less over-excited" and questioned whether she had behaved that way because she "smoked something" prior to the show.[65] Furthermore, he asked in several clips to be listened to without interruptions, while he described himself as "well-behaved" and someone who listened to others. The frequent use of words such as "Mrs" and "folks" further underlines that he addressed other politicians often directly.[66] Furthermore, he attacked other politicians with colloquialisms, in order to imply that they were too afraid to implement certain policies, or were not telling the truth.[67]

---

62 "Also stimmt an vielen Gesetzen bei uns was nicht. Wir haben keine Steuergerechtigkeit. Wir haben keinen flächendeckenden gesetzlichen Mindestlohn. Was soll denn 'ne Frisörin in Thüringen mit 3,50 € Brutto anfangen."

63 "Ich kann das keiner Lidl-Kassiererin erklären, warum Sie die Mütterrente bezahlt und ich nicht. Es tut mir leid, ich finde diesen Ansatz falsch."

64 "Sondern lass uns das verbinden mit der Erhöhung des Spitzensteuersatzes, und davor hatten Sie Schiss."

65 "Das heißt junge Leute nach guter Ausbildung und Studium finden zur Hälfte neue Arbeitsplätze vor, die befristet oder Leiharbeit sind; Ich sage Ihnen, Sie können das Kindergeld vervierfachen. Hören Sie doch einfach mal zu; Sie scheinen immer, sind aufgedreht, sind aufgedreht, als hätten Sie vorher etwas geraucht. Ich glaube, das ist nicht so; Nein, das ist doch gar nicht so schwer. Ich höre Ihnen doch auch die ganze Zeit zu. Lassen Sie mich doch wenigstens einmal den Satz ausreden; Das kann doch nicht so schwer sein. Selbst für jemanden wie Sie kann das doch mal drin sein."

66 "Ich halte es nicht... ; Ich halte es nicht... [unverständlich/ incomprehensible], Frau Klöckner. Ich bin einfach dafür, dass wir uns gegenseitig zuhören und uns dann sagen, warum der andere Unrecht hat. Ich finde es… ."

67 "Ich bin nicht sauer darüber, was bei den Grünen passiert, sondern dass Frau von der Leyen hier so, sagen wir mal, relativ unverfroren die Unwahrheit sagt."

Angela Merkel's anger expressions focused more strongly on the polity dimension of political life with a focus on the law, and shared European values.[68] Nonetheless, she also mentioned the importance of caring and benefiting companies and creating jobs while "others just talk".[69, 70] Because she made some of her appearances at party conventions, she addressed her fellow party members as "dear friends"[71] and mentioned the "Christian Democratic Union" by its full name to emphasize the values and ideological position of the party.[69]

Figure 16: Word Clouds Based on the Experimental Treatments

Note: Author's own illustration. The word clouds are based on word stems with a frequency of at least two occurrences within the treatment conditions.

---

68 „Aber ich sage auch, das Recht des Stärkeren kann nicht über das Recht stehen und deshalb müssen wir uns auch für das Recht einsetzen, sonst sind alle unsere Sonntagsreden leere Worthülsen."

69 "Wir Mitglieder der Europäischen Volkspartei und der Christlich Demokratischen Union werden [bei] einer Arbeitsteilung nicht mitmachen. Die einen kümmern sich darum, dass Arbeitsplätze entstehen in Unternehmen und die anderen kümmern sich ums Verteilen."

70 "Nein, wir wissen doch aus Deutschland, dass die soziale Marktwirtschaft uns stark gemacht hat. Erst muss etwas erarbeitet werden und deswegen müssen wir gute Bedingungen für die kleinen und großen Unternehmen auf unserem Kontinent haben."

71 "Und ich möchte zu den Fehlern der Vergangenheit nicht zurückkehren. Das hat mein Vorgänger gemacht und was daraus geworden ist, haben wir gesehen mit der Euro-Krise. Das darf sich nicht wiederholen, liebe Freunde."