



Establishing the validity and robustness of facial electromyography measures for political science

Gijs Schumacher¹ , Maaïke D. Homan² , Isabella Rebasso³, Neil Fasching⁴, Bert N. Bakker⁵  and Matthijs Rooduijn¹

¹Department of Political Science, University of Amsterdam, Amsterdam, Netherlands, ²Organisational Behaviour Research Group, Utrecht University, ³Department of Government, University of Vienna, Vienna, Austria, ⁴Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, USA and ⁵Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands

Corresponding author: Gijs Schumacher; Email: g.schumacher@uva.nl

Abstract

Opinion formation and information processing are affected by unconscious affective responses to stimuli—particularly in politics. Yet we still know relatively little about such affective responses and how to measure them. In this study, we focus on emotional valence and examine facial electromyography (fEMG) measures. We demonstrate the validity of these measures, discuss ways to make measurement and analysis more robust, and consider validity trade-offs in experimental design. In doing so, we hope to support scholars in designing studies that will advance scholarship on political attitudes and behavior by incorporating unconscious affective responses to political stimuli—responses that have too often been neglected by political scientists.

Keywords: facial electromyography; robustness; validity; measurement

Introduction

Facial electromyography (fEMG) captures subtle activation of facial muscle areas and can be used to measure affective responses to stimuli (Bakker et al., 2021; Russell, 1980). In particular, activation of the so-called corrugator muscle above the eyebrow is indicative of a negative affective response (Hietanen et al., 1998; Lang et al., 1993; vanOyen Witvliet & Vrana, 1995), and activation of the so-called zygomaticus muscle on the cheek is indicative of a positive affective response (Cacioppo et al., 1986). This technique has several advantages compared to other approaches that measure affective responses: (1) compared to studies using skin conductance, fEMG moves beyond indications of emotional arousal and captures the direction or valence of the affective response; (2) compared to electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), fEMG is easier to capture, process, and analyze; and (3) compared to self-reports of emotion, fEMG captures unconscious affective processes while participants are exposed to a stimulus. While self-reports of emotion capture a cognitive evaluation—a feeling—after receiving a stimulus, fEMG registers a different, more unconscious aspect of affect that is not necessarily aligned with feelings (Evers et al., 2014; LeDoux & Pine, 2016). Despite these advantages, fEMG is rather unpopular in political science. Using Google Scholar, in the spring of 2023, we found only 18 studies published in peer-reviewed journals that used fEMG in the domain of politics. Only six of these were published in a political science journal (see Table A5 in the Supplementary Material).

   This article was awarded Open Data, Open Materials, and Preregistration badges for transparent practices. See the data availability statement for details.

© The Author(s), 2024. Published by Cambridge University Press on behalf of the Association for Politics and the Life Sciences. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

We identify four reasons why fEMG is unpopular. First, political scientists lack (or perceive that they lack) the skills and equipment necessary to conduct a study including fEMG measures. Second, psychophysiological measures are criticized as invalid measures of emotions (Marcus *et al.*, 2017; Osmundsen *et al.*, 2021). Third, there are too many researcher degrees of freedom in fEMG measurement and analysis, leading to a lack of robust findings. Fourth—and not necessarily unique to fEMG—there are concerns about the quality of experimental designs and the diversity of samples used in lab experiments.

We address these issues by (1) establishing fEMG as a supplementary measure, not an alternative, to capture a part of the dynamic unconscious and conscious processes that constitute an emotion episode. (2) We review different preprocessing and analysis techniques and provide recommendations to produce more robust results. (3) We evaluate a series of data collections to make recommendations regarding experimental design and sampling. To this end, we analyze data from five experimental designs, seven data collection sites, 98 treatments, and 585 individuals, for a total of 540,883 seconds of fEMG activity. We build on existing fEMG work in the field of psychophysiology (Blascovich *et al.*, 2011; Bucy & Bradley, 2011; Potter & Bolls, 2012; Tassinary *et al.*, 2007; van Boxtel, 2010), but we also rely on a series of recent developments regarding the modeling of such data (Hess *et al.*, 2017; 't Hart *et al.*, 2019).

Our article is aimed at (non-fEMG-specialist) researchers interested in measuring affective responses and theorizing about the position of emotions in politics more generally. We hope to make fEMG more popular by taking away concerns about validity and robustness and, at the same time, propose better experimental designs. Before doing this, we first explain fEMG in more detail. Yet be aware that for a specialized explanation regarding facial musculature, we refer to the relevant psychophysiological sources (Cacioppo *et al.*, 2007; Potter & Bolls, 2012; Tassinary *et al.*, 2007; van Boxtel, 2010).

What is fEMG?

Facial EMG can register rapid, automatic affective responses in the face while participants view stimulus material. Specifically, fEMG measures tiny bioelectric signals at the surface of the skin that are generated as a result of muscle contractions in a specific region (Blascovich *et al.*, 2011; Bucy & Bradley, 2011; Potter & Bolls, 2012; Tassinary *et al.*, 2007; van Berkum *et al.*, 2023; van Boxtel, 2010). These muscle contractions are not visible to the eye. This technique has a high temporal resolution, as it detects rapid changes in the contractions of facial muscles in a specific region (van Boxtel, 2010). It is primarily used as a measure of emotional valence (Larsen *et al.*, 2008; Tassinary *et al.*, 2007). The *corrugator supercilii* muscle region, the muscle above the eyebrow that draws the brow down and pulls the brows together (see Figure 1), registers negative affect. Indeed, increased corrugator activity has been recorded in response to negative images, negative words, and negative affective cues in language (Hietanen *et al.*, 1998; Lang *et al.*, 1993; van Berkum *et al.*, 2023). The *zygomaticus major* muscle region, which pulls the corners of the mouth up and back into a smile (Larsen *et al.*, 2003), registers positive affect. Indeed, zygomaticus activity increases in response to positively valenced images and videos (Cacioppo *et al.*, 1986; van Oyen Witvliet & Vrana 1995). For a full overview of experimental results with the zygomaticus and corrugator muscles, we recommend reading van Berkum *et al.* (2023), as they comprehensively review 55 studies.

Other muscles have been associated with specific discrete emotions, such as the *levator labii*, which is associated with disgust (Chapman *et al.*, 2009), or the *frontalis*, which is associated with fear. While there are concerns about whether the levator labii and frontalis validly and reliably capture these distinct emotions (Tassinary *et al.*, 2007), recent advances, particularly in data science, hold the promise of detecting discrete emotions using fEMG (Sharma *et al.*, 2019).

Facial muscles are part of the peripheral nervous system that guides muscular activity. This is unlike more commonly used psychophysiological measures in political science such as skin conductance, which is part of the autonomic nervous system. The autonomic nervous system controls our glands, and skin conductance specifically monitors glands on the hand or fingertips (for an introduction, see Soroka, 2019). We have little control over our glands. In contrast, most of our facial muscles can be used voluntarily. That said, facial muscle activation can also be automatic and expressed in response to a stimulus before conscious awareness (Dimberg *et al.*, 2000; Mehu *et al.*, 2012). In particular, automatic affective responses

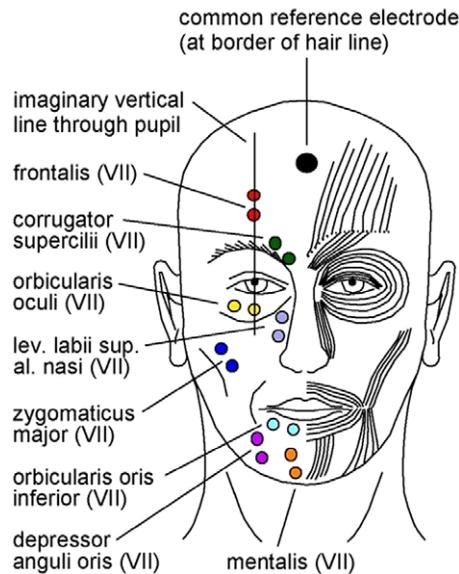


Figure 1. EMG electrode locations for several facial muscles. Figure courtesy of Anton van Boxtel (2010).

have been detected in evaluating the stimulus (Neumann et al., 2005) or by simulating the emotion presented in the stimulus (Niedenthal et al., 2010; 't Hart et al., 2019). Facial EMG responses also relate to specific activity in the motor cortex and other brain structures associated with emotions (Achaibou et al., 2008; Morecraft et al., 2004; Rymarczyk et al., 2018). Finally, the automaticity of fEMG responses is further underlined by the fact that they cannot be suppressed. If you ask people to not move their face, facial activity congruent with the valence of the presented stimuli is still present (Cacioppo et al., 1992; Dimberg et al., 2002). Even if participants are prompted to smile, they still frown upon exposure to a negatively valenced stimulus. In all, fEMG is a valid tool to register automatic, affective responses to stimuli.

Recent work uses emotion recognition algorithms to capture emotional expressions (Boussalis et al., 2021; Masch et al., 2021). These algorithms are only capable of detecting *visible* muscle contractions, while fEMG can capture tiny temporal changes in muscle contraction invisible to both the eye and computer vision (Perusquía-Hernández et al., 2019). In addition, it is problematic that there is little documentation of how these algorithms work and what training data they are based on. Therefore, while useful as a tool to measure expressions of political elites, emotion recognition algorithms cannot replace fEMG in the lab.

Facial EMG is a valid measure of affect and theoretically useful

The preceding section demonstrated the construct validity of fEMG, as evidenced by the expected fEMG responses to stimuli with positive or negative valence (Tassinari & Cacioppo, 1992; van Berkum et al., 2023). In our own analyses, we find high correlations between the valence of a stimulus (negative-to-positive) and both corrugator activity ($r = -.774$) and zygomaticus activity ($r = .501$) (see the section on Robustness of fEMG measures and analysis). Yet in political science, the convergent validity of psychophysiological measures has faced scrutiny because of their low correlation with self-reports of emotion (Marcus et al., 2017; Osmundsen et al., 2021). This misses the point. There are two reasons why we would not expect a substantive correlation between fEMG and self-reports. First, the lack of connection with self-reports emerges because conscious and unconscious processes are not necessarily connected (Evers et al., 2014; LeDoux & Pine, 2016). Different brain systems may even be responsible for the production of physiological fear responses versus cognitive fear responses (LeDoux & Pine, 2016). Keltner and Gross (1999) define emotions as “episodic, relatively short-term, biologically-based patterns of perception, experience, physiology, action, and communication that occur in response to specific physical and social

challenges and opportunities” (p. 468). Emotions bring about rapid physiological changes in, for example, heart rate or blood pressure. But emotions also convey our feelings and evaluations of the situation to others by means of posture, voice, and, indeed, facial expression (van Berkum *et al.*, 2023). These different elements of an emotional episode do not take place at the same, but precede or follow each other dynamically (Marcus, 2012). Longer, conscious responses may, in turn, affect new unconscious responses further developing the emotion episode (Scherer & Moors, 2019). In sum, fEMG may not correlate with self-reports because the fEMG readings may not enter or may have already exited conscious awareness.

Second, the willingness of research participants to openly share their genuine feelings plays a role in the correlation between fEMG and self-reports. Some participants may feel compelled to express or suppress certain emotions based on societal expectations or appropriateness (Friesen *et al.*, 2017). This point is underscored by some of the 18 political science studies using fEMG mentioned in the introduction (see Table A5 in the Supplementary Material). For instance, Ensari *et al.* (2004) demonstrated that participants rated politically sensitive out-groups favorably, while their fEMG activity indicated negative affect. However, when experimentally induced to justify hostility toward these out-groups, participants’ self-reported attitudes aligned with the negative fEMG activity. Similarly, Stewart *et al.* (2003) found no correlation between fEMG activity in response to interactions involving gay men and participants’ self-reported attitudes toward gay men. Interestingly, corrugator activity (measuring disgust) predicted antidiscrimination behaviors, whereas self-reported attitudes did not.

These two studies demonstrate that fEMG is relevant in contexts in which participants may be reluctant to report their views. In other cases, participants are very motivated to express their feelings, but with no discernible fEMG activity. For example, in the experiment conducted by Homan *et al.* (2022), participants reported feeling more anger when their in-party politician displayed anger instead of a happy or neutral display. But this difference between experimental conditions was not found in the corrugator and zygomaticus activity measures. Also, Bakker, Schumacher, and Homan (2020) showed that stronger partisans reported feeling more disgust toward an out-party politician, while the strength of partisanship did not significantly impact the fEMG response in the labii (disgust) muscle.

To gain a comprehensive understanding of emotional or affective responses to political stimuli, we advocate collecting both fEMG measures and self-reports of feelings. Each measure captures different aspects of this complex process. Rather than rejecting one of the two measures, researchers should theorize when they should align and when not (Arceneaux *et al.* 2024). This way, we can acquire a richer and more nuanced picture of the interplay of cognitive and affective processes in political decision-making.

Robustness of fEMG measures and analysis

Various approaches exist for preprocessing raw fEMG data, creating variables from fEMG measurements, and analyzing the effects of treatments on fEMG measures (van Berkum *et al.*, 2023). Experts in this field recognize the importance of adopting more standardized and transparent workflows, yet they also acknowledge how difficult this is given the sensitivity of fEMG readings (van Berkum *et al.*, 2023). Here we discuss the latest developments regarding preprocessing procedures, variable creation, and analysis. We use data that we previously collected to evaluate these developments and make specific recommendations. Before doing so, the next section first describes the data we collected.

Data collection

We combined data from five experimental designs gathered at seven data collection sites (note that several individual experimental designs have already been analyzed and published: Bakker, Schumacher, Gothreau *et al.*, 2020; Bakker, Schumacher, & Homan, 2020; Bakker *et al.*, 2021; Homan *et al.*, 2022; Schumacher *et al.* 2022). Yet the analyses conducted here are new. In total, this data set contains 98 treatments. Some of these treatments are images from the International Affective Picture System (IAPS) with positive, negative (or disgusting), or neutral valence. Other treatments are images of party

Table 1. Overview of experimental designs, number of participants and number of treatments

Design	Location	<i>N</i> (<i>p</i>)	<i>N</i> (<i>t</i>)	fEMG measures	Label treatments
Design 1	Lowlands	142	8	Corrugator, zygomaticus	Issues 1
Design 2	Lab	119	18	Corrugator, zygomaticus	Issues 1, Dominance and IAPS-threat pictures
Design 3	Media museum	47	18	Corrugator	Issues 2,
	EO Youth Festival	33	18	Corrugator	Party leaders and
	TT Assen	14	18	Corrugator	IAPS-neutral pictures
	Fair Tilburg	33	18	Corrugator	
Design 4	Film festival	41	12	Corrugator, labii	Leaders' moral violations, IAPS-disgust pictures
	Lab	54	40	Corrugator, labii	Leaders' moral violations, IAPS-disgust pictures IAPS-various
Design 5	Lab	102	56	Corrugator, zygomaticus	Leaders and emotions, ADFES expressions, IAPS (positive, negative), and spoken words

Notes: In the header, *p* stands for participants, *t* for treatments. IAPS refers to the International Affective Picture System. ADFES is the Amsterdam Dynamical Facial Expression System. The treatments are described in detail in Table A1 in the Supplementary Material. Regarding the locations: Lowlands is a large cultural festival, the EU Youth Festival is an evangelical youth festival, TT Assen is a bikers event, the Fair Tilburg is a fair throughout the city of Tilburg in the Netherlands. The Media Museum is a museum about media in Hilversum in the Netherlands. The Lab refers to the university lab. The locations are diverse in terms of geographical spread.

leaders or videos in which different stances on political issues are presented. Table 1 provides an overview of these data collections. Table A1 in the Supplementary Material describes each experimental design.

In total, the data set contains 585 unique participants from different locations. Participants in the university lab signed up through an online portal and received a financial compensation (equivalent to 10 euros per hour or course credit). Participants in the labs-in-the-field were recruited on-site and received no financial compensation. In all cases, participants first signed an informed consent form and completed a survey on a desktop computer (laboratory) or iPad (lab-in-the-field). Next, participants were connected to physiological measurement equipment by trained research assistants. Participants (lab-in-the-field) were also given noise-canceling headphones (Bose). We recorded events during the experiment in a logbook. For example, when electrodes came off or when a participant's mobile phone rang, we wrote this down to evaluate how it impacted our measurement. We return to this issue in the next section.

We recorded physiological responses per millisecond using the Versatile Stimulus Response Registration Program 1998 (Vsrrp98) software on laptops (lab-in-the-field data collection) or stationary computers running Windows 7 (laboratory data collection) at 1,000 Hz. The lab equipment we used was able to reliably and validly capture fEMG activity in earlier work in other domains (for references, see Bakker et al., 2021).

To measure negative affect, we measured the activity of the corrugator supercilii (see also Figure 1). We did this using two 70-millimeter Ag/AgCl electrodes that were filled with electrolyte gel (Signa, Parker Laboratories). Using double-sided adhesive tape, we placed the two electrodes just above the eyebrow—directly where the muscle is located (Fridlund & Cacioppo, 1986). A third electrode was placed on the middle of the forehead (just below the hairline) and served as the ground measure. The corrugator has no overlapping muscle groups, has a very limited representation in the motor cortex,

and “tends to be bilateral innervated” (Larsen et al., 2003, pp. 776–777). The measurement of the corrugator is therefore less subject to disruptions from the (voluntary) movement of other muscles.

We also measured the activity of the zygomaticus major by placing electrodes on the cheek where the zygomaticus muscle region is located (Larsen et al., 2003).¹ It is a difficult muscle to measure because it has greater contralateral innervation (Larsen et al., 2003), and the cheek is a particularly crowded area of the face with lots of muscles (Tassinari et al., 2007). This makes measures of the zygomaticus susceptible to “cross talk” (Larsen et al., 2003, p. 777). Irrespective of the difficulties of measuring the zygomaticus, it is a unique measure to capture positive affect.²

(Pre)processing raw fEMG data

Facial EMG data are highly sensitive. For one, because of technical errors, physiological measures or time markers of treatment onset may be (partially) missing and therefore cannot be analyzed. Second, fEMG measurement is susceptible to external noise. To reduce noise in the signal, raw fEMG data are typically band-pass filtered between 20 and 400 Hz, with an additional 50 Hz notch filter (van Boxtel, 2010). Recording software for fEMG typically integrates the raw signal. The output of this process is fEMG values expressed as microvolts per time unit of choice.

In addition to these standard procedures, we describe three more preprocessing steps that we have encountered in the literature. First, during the experiment, electrodes may come off or temporarily have problems transmitting the signal. These issues can be identified by visualizing fEMG activity and searching for flat lines or big temporary drops in the signal. We used two human coders—who were unaware of the experimental conditions—to visually inspect physiological activity and search for flat lines and temporary drops. On the basis of this, we identified 10% to 15% of the data as potentially problematic. Second, fEMG readings may suffer from the activity of other facial muscles. Zygomaticus readings may be suspect because of people clenching their jaw. Corrugator activity can be influenced by eye blinks (Bhowmik et al., 2017). Because such cross talk follows a particular pattern, specific algorithms can be applied to reduce this influence. In particular, we use a hamper filtering algorithm to identify and replace outliers (Bhowmik et al., 2017).³ Third, it is possible to use different statistical transformations such as winsorizing⁴ or removing statistical outliers.⁵ To evaluate the consequences of these preprocessing steps, we compare estimates of treatment effects on dependent variables constructed using different preprocessing steps. Figure 3 summarizes these results, but before we discuss them, we first need to take a few more steps to completely explain our measurement and analysis strategy.

fEMG measure construction

It is a general recommendation in the psychophysiology literature (Blascovich et al., 2011; Potter & Bolls, 2012) that the fEMG activity of a participant during a stimulus be contrasted with fEMG activity during a neutral baseline. Typically prior to a stimulus, participants see a blank screen—with a cross in the middle for focus—and are asked to relax. Baseline activity thus measures the physiological activity of a specific muscle when in a resting state.⁶ What is less clear in the psychophysiology literature is how long this baseline measurement should last. Participants are exposed to these blank screens between 8 and

¹On the line between the corner of the mouth and the middle of the ear.

²In addition, we measured skin conductance (all studies), heart rate (Studies 1 and 2), and levator labii activity (Study 4). We do not discuss these measures here. We encourage interested parties to contact us for these data.

³Outliers in corrugator readings caused by eye blinks take the form of single, repetitive high peaks. The hamper filtering algorithm identifies these peaks by *z*-standardization and replaces the peaks with the median value of surrounding observations.

⁴All values that deviate two standard deviations from the mean are set to the value of two standard deviations from the mean in the original direction.

⁵We removed data with *z*-values higher than 4.

⁶Physiological activity may also be affected by the duration of the experiment; therefore, it is best to have baseline activity measurement immediately before each of the treatments.

30 seconds, and part of the activity during this period is used to calculate baseline activity. Some researchers use the last 2, 5, or 7 seconds of the baseline, while others recommend finding those seconds before the treatment that are artifact-free. It is quite often not specified why a specific selection of the baseline is chosen. To test whether the choice of a specific baseline influences our results, we analyzed baseline activity in our data by taking five different centrality measures (e.g., the mean or the median) and selecting nine different periods from the baseline (e.g., all data, last 2 seconds). We compared these $9 \times 5 = 45$ baseline calculations to each other and to human coding to identify the general tendency in the baselines. Section 5 in the Supplementary Material provides the details of this analysis. On this basis, we give a weak recommendation to taking the median of the baseline (all observations). This measure produced among the highest correlations with human-coded data, and medians are generally more effective at reducing noise in signals than means. Also, taking all observations in the baseline avoids the need to select between rather arbitrary cutoff lines of 2, 3, or 5 seconds. That said, other choices here are unlikely to produce very different results.

By subtracting the baseline value from fEMG activity during the stimulus, one can construct a measure of fEMG activity that allows for comparisons between participants. This measure expresses the difference in microVolt between the stimulus and the individual baseline. In the literature, there are two ways to construct a measure of fEMG activity. First, divide the raw fEMG activity in the treatment by the baseline activity and multiply by 100. This way, all values can be expressed as a percentage increase or decrease. A second option is to *z*-standardize the fEMG data. The microvolt and percentage measures correlate very highly in our data set ($r = .9$). Both measures correlate at approximately $r = .7$ with the *z*-standardized measure. In this article, we use the second measure, which indicates percentage increase and decrease in microvolt fEMG activity. We chose this measure because we find it easier to interpret. Given the high correlations with the alternatives, we expect this choice to matter little.

So far, we have discussed analyzing the activity of single muscle regions. But researchers have also calculated a facial activity index by subtracting zygomaticus activity from corrugator activity (Olszanowski et al., 2020). This is intuitively appealing, because it creates a single valence measure. Yet this measure becomes problematic if participants have counterempathetic responses (Bucy & Bradley, 2004)—for example, when participants are amused by an angry-looking out-party politician (Homan et al., 2022). In this particular case, both corrugator and zygomaticus activity point in the same direction, and so subtraction produces zero facial activity even though there are clear physiological signals. In this case, capturing overall facial reactivity by adding zygomaticus activity to corrugator activity may be interesting.

Analysis strategy

Now that we have constructed a measure of fEMG activity, what analysis strategy should be adopted to analyze the effects of treatments? Various political science applications using psychophysiology have compared mean fEMG activity in different treatments (e.g., Settle et al., 2020). Alternatively, one can calculate the maximum fEMG activity or the area under the curve (AUC) (Osmundsen et al., 2021). More recently, however, fEMG activity during a treatment has not been collapsed into a single number but modeled by second as a multilevel model (Hess et al., 2017; Olszanowski et al., 2020; 't Hart et al., 2019). To see why this is useful, Figure 2 shows four corrugator responses to a negatively valenced image. All signals start at their own individual baseline set at 100, after which the treatment starts. The top (black) line shows strong increased corrugator activity. The bottom (blue) line shows a decrease in corrugator activity. The two middle lines show a response (a peak above 100) but also a decline. This decline is typical of fEMG responses and can be interpreted as activation-relaxation. Facial EMG activity typically reaches its full potential 750 milliseconds after stimulus onset, after which it returns in the direction of the baseline level (van Berkum et al., 2023). The bottom and top lines also show this pattern. The problem is that if we take the mean of these responses, the green line will be around 100 and interpreted as a nonresponse, even though there is a clear peak of 10% more activity than the baseline. The problem with

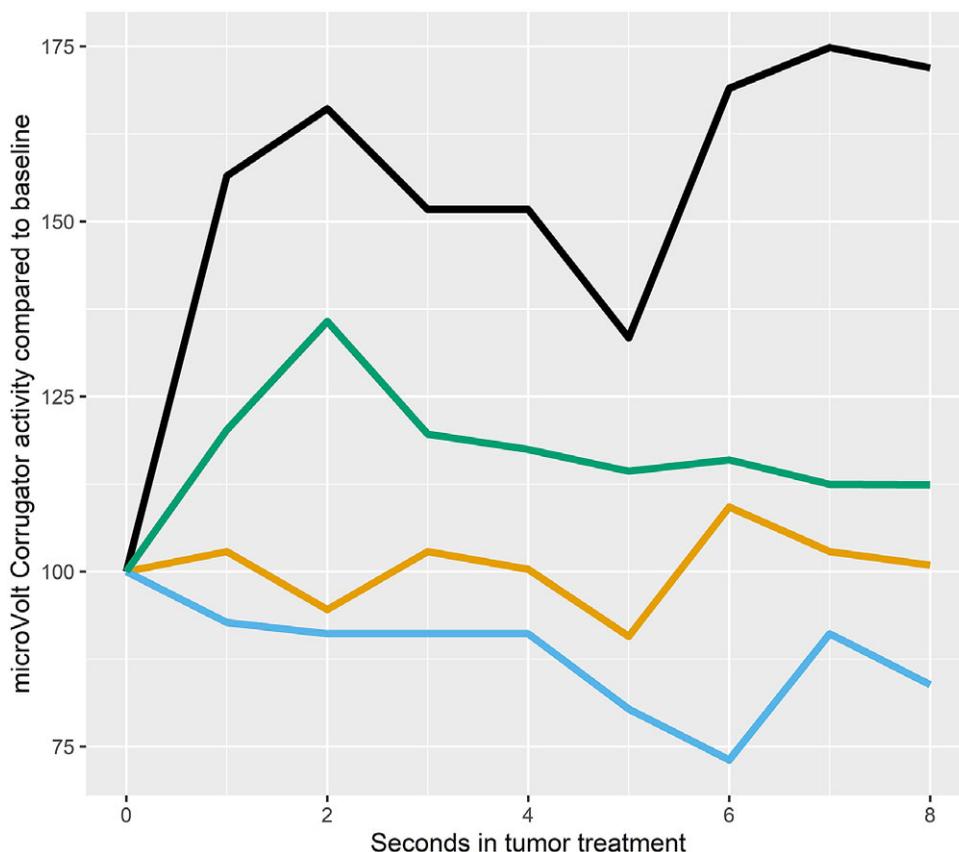


Figure 2. Corrugator activity during tumor treatment. The figure shows corrugator responses to the tumor treatment from four participants over the entire 8 seconds of exposure to the treatment. Corrugator activity is expressed as microVolt increase or decrease compared to an individual baseline (set at 100).

taking mean activity is that we may equate weak responses with no responses because the relaxation of the muscle is (close to) equal to the activation. The added advantage of a multilevel setup is that one can also model time dynamics.

To further substantiate the choice for a multilevel model, [Figure A3](#) in the Supplementary Material presents simulations in which we compare different analysis strategies. Specifically, we contrast using the mean, the maximum, and the AUC activity measures with a multilevel model using second-by-second activity measures. In the latter case, we retrieve the significant treatment effect in 95%, whereas in the other cases this is between 76% and 77%.

We recommend using a multilevel setup to analyze fEMG activity. This still gives some degrees of freedom regarding model specification, such as how to model time dynamics and what covariates to add. For this article, we ran a series of models including time dynamics and covariates—but no treatments—and selected the model with the best fit. Section 6 in the Supplementary Material provides more details regarding this. We preregistered the analysis strategy and preprocessing strategy discussed here (<https://osf.io/2wd8g/>). Our design is sufficiently powered to detect small effects (see Section 9 of the Supplementary Material). All subsequent analyses adopt these strategies.

Analysis of preprocessing steps

Figure 3 compares the preprocessing steps that we discussed earlier. We subset our data to two conditions: an IAPS picture with negative valence (a tumor) and an IAPS picture with positive valence

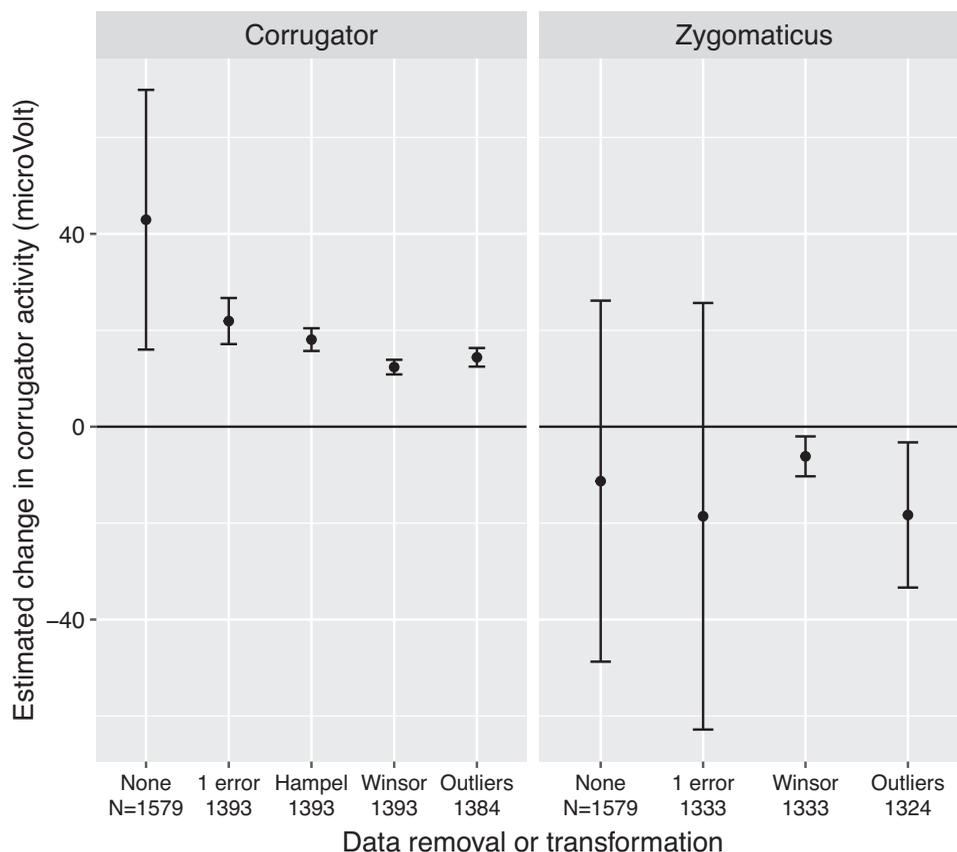


Figure 3. Treatment effects with different transformations of fEMG activity. The left panel shows estimates of the treatment (tumor versus baby condition) on corrugator activity (as a difference in microVolt compared to an individual baseline) using a multilevel analysis model. Each individual dot is a regression estimate from a single analysis, and the bar denotes the 95% confidence interval. The first estimate is based on all our observations. Then we remove those observations identified as erroneous by two coders. We also winsorize the data and remove statistical outliers (last estimate). The right panel has the same interpretation, but with zygomaticus activity as the dependent variable.

(a baby)—taken from Design 5 ($n = 102$). Using the percentage of fEMG activity compared to the baseline—as discussed in the preceding section—we estimate the difference between the two conditions. We expected that the negative image of a tumor would produce more corrugator activity and less zygomaticus activity than the positive image of a baby. To this end, we ran two multilevel models estimating the difference in corrugator activity (left-hand panel of Figure 3) and zygomaticus activity (right-hand panel).

The first estimate (a dot above the “none” label; the bars denote 95% confidence intervals) includes all recorded data. The unstandardized coefficient is 42.933. This means that—as expected—in the tumor treatment, participants have 42.9% more corrugator activity than in the baby treatment. Then we removed signals that at least one human coder identified as problematic (labeled *1 error*). This reduces the effect size by a factor of 2 and the standard error by a factor of 6.5. The next set of preprocessing steps further reduces the standard error, respectively by a factor of 2 (the hampel filter compared to the 1 error condition), 1.6 (winsorizing compared to hampel filter), and 1.2 (outlier removal compared to hampel filter). Effect sizes also gradually diminish, but by the same degree as the standard error.

We took the same steps to analyze zygomaticus activity (right-hand panel of Figure 3). As expected, the direction of the coefficient is negative, meaning that the tumor treatment produces less zygomaticus

activity than the baby treatment does. However, in steps 1 and 2, the confidence intervals are very large, and therefore the estimated difference is not statistically significant. Winsorization and the removal of remaining statistical outliers strongly reduce the standard errors in our estimates (to 9.3% and 34.1%, respectively), making the estimate statistically significant. Note that the hamper filter (for eye blinks) is only applicable for corrugator data.

For both corrugator activity and zygomaticus activity, we identified that most preprocessing steps reduced the error in our estimates—as one would expect. There is no indication of flipping signs of treatment effects. Our recommendation is to ensure robustness of results by presenting treatment effects using different preprocessing steps.

Improving designs including fEMG measures

In this section, we deal with questions of research design: What treatment characteristics work best to elicit fEMG responses? And what about the (lack of) participant heterogeneity in typical lab samples?

Data collections at the University of Amsterdam's lab were complemented with data collections in the field to reach a politically more diverse population. We also investigate differences between the lab and the lab-in-the-field and explore the role of participant characteristics and treatment characteristics to make recommendations regarding future experimental designs with fEMG measures.

What is the effect of treatment characteristics on fEMG?

The 98 treatments in our dataset differ in medium (video, image, or word), sound (no sound or sound), valence (positive, negative, or neutral), and political versus nonpolitical content. This variation allows us to evaluate the impact of treatment characteristics on fEMG activity. To analyze this, we coded each treatment as follows: (1) does the treatment contain positive or negative emotional content or is it neutral; (2) is the treatment a video (= 2), a word (= 1), or an image (= 0); (3) is the treatment with (= 1) or without audio (= 0); (4) does the treatment have political (= 1) or nonpolitical content (= 0); and (5) of those treatments with political content, does the treatment include a face of a politician (= 1) or does it describe an issue position (= 0). [Table A1](#) in the Supplementary Material describes how each specific treatment was coded.

[Figure 4](#) visualizes the results of our analyses.⁷ Before discussing the results, it is important to note that the two dependent variables produce by and large the same conclusions. The effects on the winsorized dependent variable look a little bit smaller, but this is because the variance in these variables is also smaller.⁸ For practical reasons, our presentation of the results concentrates on the first point estimate in each panel, which are from the analysis with erroneous signals and outliers removed and hamper filter applied.

First, we consider the effect of the valence of the nonpolitical treatments. We compare negative to neutral (first column) and positive to neutral (second column). Regarding corrugator activity, there is a small positive effect of negative versus neutral content ($b = 5.454$, standardized (b) = 0.176, $SE = 0.266$). Put differently, negative content produces 17.6% more corrugator activity than neutral content does. Positive content, as expected, produces less corrugator activity than neutral content ($b = -6.186$, standardized (b) = -0.199 , $SE = 0.352$) and, by implication, negative content. This means that the corrugator muscle relaxes during exposure to positive materials. Positive material also produces significantly more zygomaticus activity (second bottom panel) than neutral materials ($b = 17.442$, standardized (b) = 0.135, $SE = 1.724$). Surprisingly, negative material also produces more zygomaticus activity. In terms of effect size, this effect is a factor 4 smaller than the effect of positive material and only

⁷Some of the preregistered multilevel models experienced problems in converging. Therefore, we ran a simpler model with only one level: the participant level. These models converged, and importantly, produce results that are highly similar to the preregistered models (see [Table A3](#) in the Supplementary Material for a comparison of baseline models).

⁸For corrugator activity, the variance of the winsorized variable is a factor 0.7 smaller, and for zygomaticus activity it is a factor of 0.2.

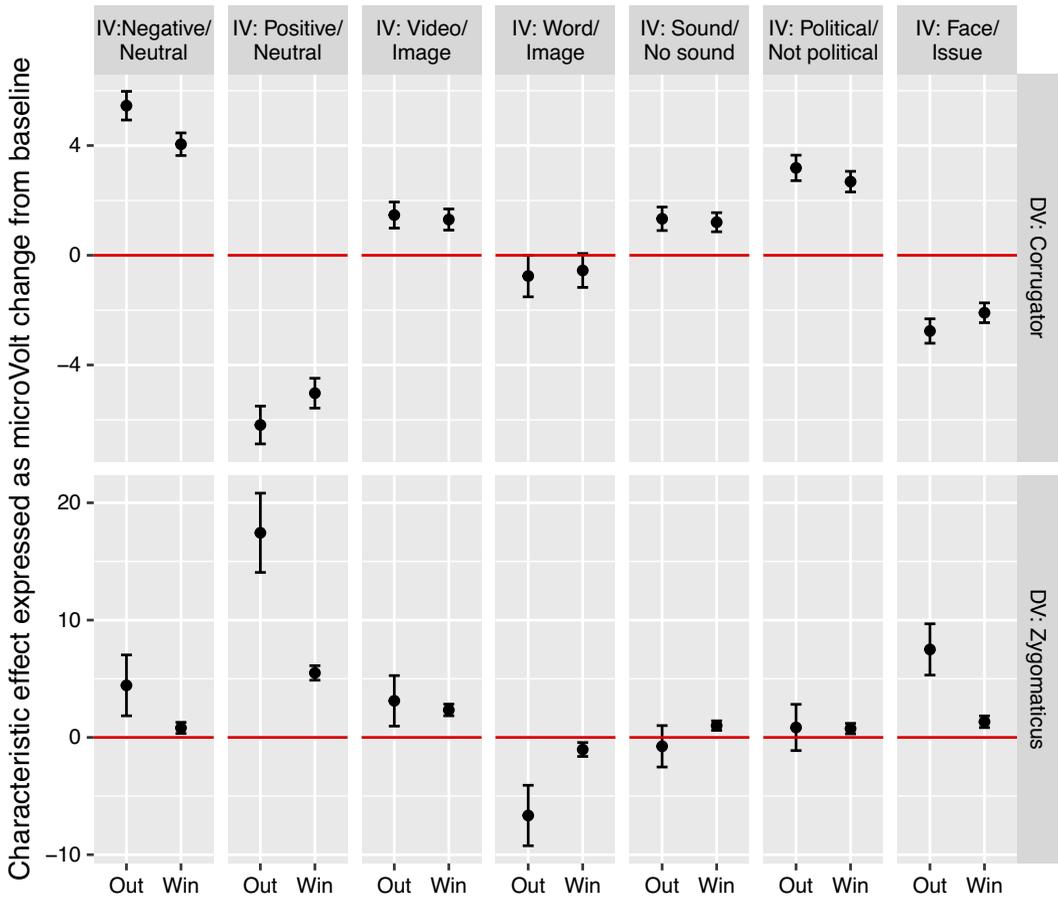


Figure 4. Multilevel regression effects on fEMG activity of treatment characteristics. The figure presents results from 24 multilevel models. The dots are the beta coefficients, and the bars denote 95% confidence intervals. In some cases, the bars equal the size of the dots and therefore are poorly visible. There are four dependent variables: (1) corrugator with statistical outliers removed, (2) corrugator winsorized, (3) zygomaticus with statistical outliers removed, and (4) zygomaticus winsorized. Five sets of independent variables are displayed: (1) valence, including negative, positive, and neutral; (2) video, word, and image; (3) sound/no sound; (4) political versus nonpolitical treatments; and (5) treatments that show faces or present issues. For each set, we ran a separate multilevel analysis model. In each multilevel analysis model, there is a random intercept for the individual participant and additional controls to model time in the treatment. The estimates are expressed as microvolt change from the baseline.

explains a difference of 1/30th standard deviation in zygomaticus activity ($b = 4.436$, standardized ($b = 0.034$, $SE = 1.323$)). As such, it is mostly neutral material that does not produce much zygomaticus activity. Another explanation is that very negative stimuli also produces zygomaticus activity (van Berkum et al., 2023).

Videos produce slightly more physiological activity than images do, both regarding corrugator activity ($b = 1.466$, standardized ($b = 0.046$, $SE = 0.242$)) and zygomaticus activity ($b = 3.118$, standardized ($b = 0.031$, $SE = 1.099$)). Words produce slightly less corrugator activity ($b = -0.756$, standardized ($b = -.023$, $SE = 0.387$)) and zygomaticus activity ($b = -6.658$, standardized ($b = 0.066$, $SE = 1.314$)) than images do.

Treatments with audio produce slightly more corrugator activity ($b = 1.331$, standardized ($b = 0.041$, $SE = 0.218$)) than treatments without audio do. Regarding zygomaticus activity, the picture is more diffuse with one measure suggesting a null effect and the other measure a tiny positive effect.

Treatments with political content produce more corrugator activity than treatments with nonpolitical content ($b = 3.185$, standardized ($b = 0.099$, $SE = 0.237$)), while there seems to be no difference regarding

zygomaticus activity. Political treatments including faces—as opposed to issues—produce less corrugator activity ($b = -2.761$, standardized ($b = -0.086$, $SE = 0.227$) and more zygomaticus activity ($b = 7.499$, standardized ($b = 0.074$, $SE = 1.115$). Overall, the treatment characteristics evoke corrugator and zygomaticus activity in predicted ways, further validating the measures.

What is the best design of a treatment to elicit fEMG activity? Prior to collecting the data for our first design, we reasoned that as most people do not care about politics, our experimental treatments needed to be very strong to elicit any bodily response. Therefore, we designed videos in which strongly worded political positions were defended by a voice actor. The fact that treatments with political content elicit stronger corrugator activity than treatments with nonpolitical content indicates that our initial reasoning was wrong. People do have strong physiological responses to politics. In one experiment, we measured fEMG activity in response to images of in-party leaders, out-party leaders, and ordinary citizens with different manipulated emotional expressions. These manipulations were identical. Even with such highly similar stimuli, participants have stronger responses to the political stimuli compared to the nonpolitical ones.

We do not need to make our stimuli dramatic to elicit a fEMG response. In fact, most of fEMG research uses very simple stimuli such as words or sentences. In the latter case, the words in these sentences are interspersed with sufficient time to reliably capture an fEMG response that can be associated with the stimulus (‘t Hart *et al.*, 2019; van Berkum *et al.*, 2023). Such a design has stronger internal validity than the videos we developed, because in the videos, a lot happens at the same time. We do not know whether people respond to the images or to what is said. The videos look like real campaign messages. Therefore they are more ecologically valid, but this may also produce more noise in fEMG activity. In designing fEMG studies, it is important to consider this trade-off between internal validity and ecological validity.

What is the effect of sample characteristics on fEMG?

Facial EMG data are typically collected in a university lab. A concern with lab studies is that by using small samples of typically university students, we lack the heterogeneity in, for example, political attitudes to answer our research questions. Because we shared this concern, we ran several of our studies outside the lab at a field location. This clearly increased variation in ideology, age, and education in our sample (Bakker, Schumacher, Gothreau *et al.*, 2020), but it also brings some challenges regarding the comparability of the results because of, for example, differences in temperature and surroundings. Also, some of our field locations were festivals, and some of our participants had been drinking.⁹

To assess differences between data collections in the lab and the field labs, we conduct the same analysis as in the preceding section but now add a dummy variable for lab versus field location. Regarding both measures of corrugator activity, we find no statistically significant difference between data collected in the lab or in the field lab. For zygomaticus activity, however, we find systematically higher zygomaticus activity in the lab than in the field lab ($b = 28.88$, standardized ($b = 0.301$, $SE = 4.927$). In this case, this only concerns the Lowlands field lab location, because we did not collect zygomaticus data in the other field labs. In contrast, corrugator activity at Lowlands was significantly lower than in all other locations. We ran a number of robustness checks that could not account for the difference.¹⁰

The previous analyses contrasts level differences in fEMG activity. It does not analyze whether treatment effect would differ depending on the location or participant characteristics that are more common in one location than another. We turn to two such analyses.

⁹Although this did not affect our results, we do believe it is ecologically valid that some people process politics while having a beer or two.

¹⁰Robustness analyses demonstrate that the following explanations do not explain the difference: (1) At Lowlands, the readings on one computer were affected by electric noise produced by a generator nearby, but we found no differences between the two computers that we used there, as our filters likely removed the noise. (2) The difference with Lowlands also remains if we subset the data to include only the “issues 1” experiment that we ran at Lowlands and a number of other locations. (3) Lowlands visitors were not significantly happier than visitors to other locations because we compare each participant to their baseline activity.

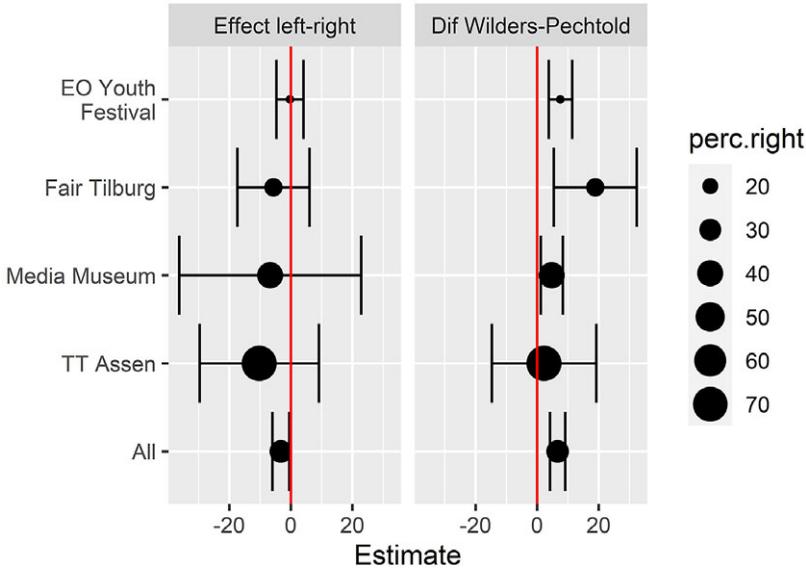


Figure 5. Differences in regression estimates of corrugator activity per location. Both panels present results from separate multilevel regression models per location, and one for all locations. The first panel shows regression estimates of left/right ideology on corrugator activity while exposed to the picture of the Dutch radical-right leader Geert Wilders using the multilevel setup used throughout the article. The second panel shows the effect on corrugator activity (microVolt difference from baseline) of the Wilders condition (leader populist radical-right party) compared to the Pechtold condition (his most vocal opponent). See Table 1 for an explanation of the locations.

First, would different locations lead us to different treatment effects? Figure 5 seeks to illustrate this problem. We list the four locations in which we conducted Design 3. There is quite a variance in the percentage of voters who reported voting for a right-wing party, as indicated by the different sizes of the points in Figure 5. At the EO Youth Festival—an evangelical youth festival—only 15.6% percentage of the participants voted for a right-wing party.¹¹ Yet at TT Assen—a biker’s event—71.4% voted for a right-wing party. To evaluate the impact that such variation can have, we first regressed the left/right ideology question on corrugator activity during exposure to an image of Geert Wilders—the leader of a radical-right party in the Netherlands (see left panel of Figure 5). The overall effect (see “All” on the y-axis) is statistically significant and negative. This means that left-wing individuals experience more corrugator activity to Wilders than right-wing people do. In none of the individual locations is the effect statistically significant, yet in three out of four cases, it is negative. Only at the EO Youth Festival, where we have relatively few right-wing people, is the effect precisely 0. The right panel shows a regression estimate from comparing two experimental conditions—Wilders to his main critic Alexander Pechtold, the leader of D66, a social-liberal party. Here, we retrieve the general positive effect of Wilders producing more corrugator activity than Pechtold in three out of four locations. Now in the case with the most right-wing voters, we also find a positive but not significant effect. These two cases illustrate that imbalances in a variable like ideology can affect the results of a study. It shows the importance of drawing participants from a pool that is not necessarily representative for the population, but minimally balanced.

Second, do participant characteristics moderate treatment effects? For reasons of space, we concentrate here on analyzing differences between participants in how they processed neutral, negative, and positive treatments. We consider the moderating effect of participant characteristics, including gender, age, left/right ideology, and education, as these are common sources of variation in political attitudes.¹²

¹¹The main political party for evangelicals in the Netherlands is the Christian Union, which is considered a centrist party.

¹²We regressed the four fEMG activity measures used so far on these participant characteristics using the same multilevel setup discussed earlier. Although we do find some effects for gender and education, they are not systematic across specifications

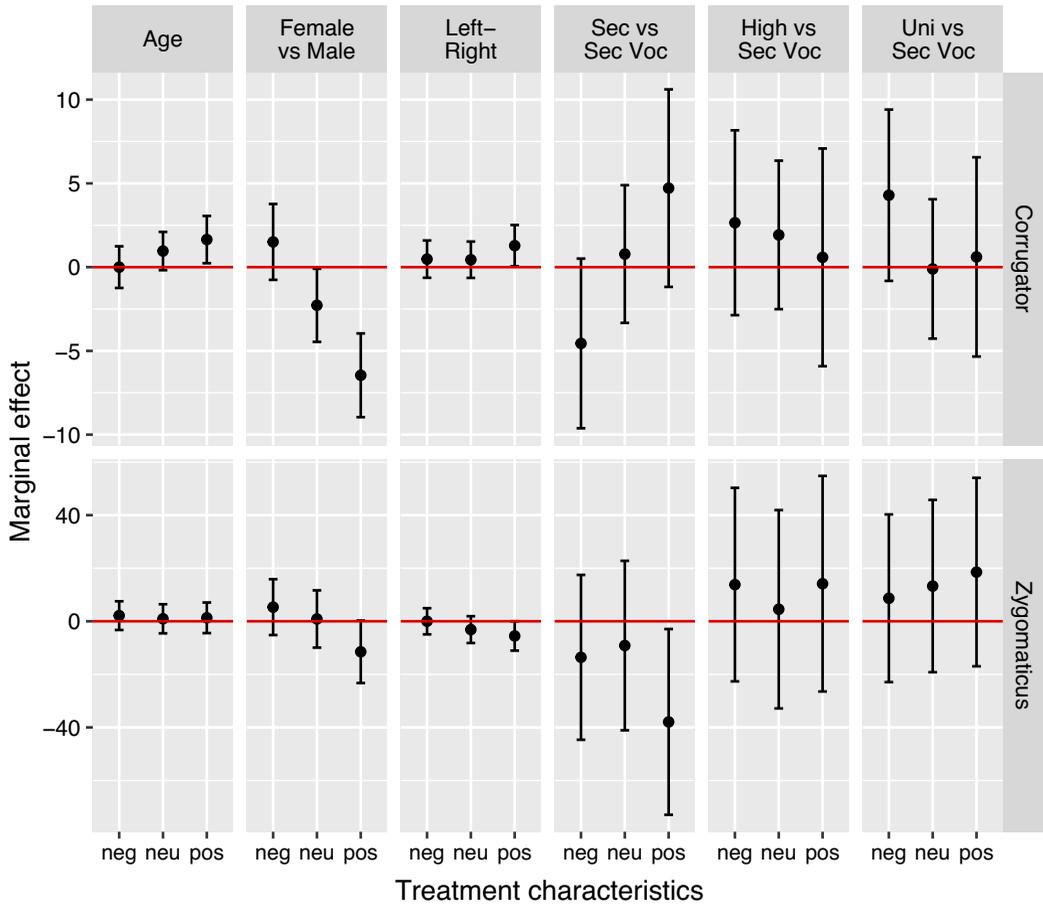


Figure 6. Interaction predictions of effect treatment characteristics per participant characteristics. The figure presents results from eight multilevel models (one for age, one for gender, one for left/right, and one for education). Each model interacts a participant characteristic with the negative, neutral, or positive content of a treatment. Age and left/right ideology are modeled as continuous variables; panels 1 and 3 show the marginal effect of age and left/right ideology, respectively, in the negative, neutral, and positive treatments. Gender and education are measured as categories; therefore, we report the difference between two categories in the neutral, negative, and positive treatments, respectively. For gender, the difference is between females and males. For education, the reference category is secondary vocational education completed (sec voc); this is compared to sec (finished secondary education—mostly students), high (finished higher vocational education) and uni (finished university education). The dots are the beta coefficients, and the bars denote 95% confidence intervals. Compared to Figure 4, we only report analyses with outliers removed; otherwise, the analyses are the same. Note that all dependent variables have been standardized.

We asked a gender question with three categories (0 = male, 1 = female, 2 = nonbinary). We asked age in years. Left/right ideology was measured with a single question that asked participants to place themselves on a 0 (= left) to 10 (= right) scale. Finally, we asked participants to indicate the highest level of education that they finished. We identified four levels: secondary vocational, higher vocational, secondary, and university. In the Supplementary Material, [Table A2](#) and [Figures A1a](#) and [A1b](#) provide the descriptive statistics of these measures.

[Figure 6](#) illustrates the marginal effects of participant characteristics in negative, neutral, and positive treatments. Starting on the left, the higher the age of the participant, the more corrugator activity there is

(see [Table A4](#) in the Supplementary Material). Therefore, we conclude there is no robust evidence that these participant characteristics directly impact fEMG activity.

in the neutral and positive conditions, but not in the negative condition. This effect is very small: the effect of age in the positive condition is 1.646. This translates to an increase in standardized corrugator activity of 0.05 for every standard deviation increase in age.

Regarding gender differences—displayed in column 2 of Figure 6—female participants have weaker corrugator responses to positive and neutral treatments than male participants. Particularly, the corrugator responses of female participants to positive treatments are substantively lower compared to male participants (difference (female – male) = -6.45 , $SE = 1.27$). This is a substantial reduction of 0.21 standard deviation in corrugator activity. We are not familiar with an explanation of this differences from the psychophysiology literature (see also Soroka et al., 2016). Regarding negative treatments, the sign is positive, yet there is no statistically significant difference between male and female participants (difference (female – male) = 1.51 , $SE = 1.16$). We find no gender differences in zygomaticus activity.

Left/right ideology differences are displayed in column 3 of Figure 6. Left/right ideology does not have a statistically significant effect on corrugator activity in any of the treatments separately or when we compare the trend lines between the categories, although in some of the comparisons, the p -values are just above the mark for statistical significance. For zygomaticus activity, however, we find that for positive treatments right-wing participants have somewhat lower zygomaticus activity than left-wing participants. The effect is not substantive: a one standard deviation shift in left/right ideology is associated with a -0.04 standard deviation reduction in zygomaticus activity.

The effects of the education categories are displayed in columns 4–6 of Figure 6. Although there are some effects that do not cross the zero line, none of the effects is statistically significant when we adjust for multiple comparisons using the bonferroni method.

In sum, we report some differences regarding age, ideology and gender. The first two effects are very small. In all, these results do not provide strong evidence that lack of heterogeneity may affect treatment results. There are of course other reasons—such as statistical power—for seeking more heterogeneous samples.

Conclusion

The goal of this article is to encourage the adoption of fEMG in political science. We identified four issues that block this wider adoption. We addressed three of these four issues.

First, we used old and new material to show the construct validity of fEMG. Next, we identified that concerns about the concurrent validity of fEMG—that is, the correlation with self-reports—are unfounded. Self-reports and fEMG measure different aspects of complex, dynamic, partially conscious and partially unconscious emotion episodes. We should not expect them to correlate. Rather, the next step in the literature should be when to expect alignment or disalignment between self-reports and fEMG (or other psychophysiological or neurological measures) (Arceneaux et al., 2024).

Second, there is a lack of robustness in fEMG measurement and analysis due to too many researcher degrees of freedom. Following recent developments and own analyses, we recommend specific pre-processing steps, how to construct a baseline activity measure, and different ways of calculating and analyzing fEMG activity measure. Our analyses of different preprocessing steps show remarkably robustness across specifications. That said, we also recommend preregistering analysis and preprocessing plans with fEMG data (van Berkum et al., 2023).

Third, we addressed concerns about the quality of experimental designs and samples used in lab experiments. Using the data we collected and reflecting on our own experiences we identify a trade-off between internal validity and ecological validity. We analyzed treatment heterogeneity, for which we found limited evidence.

The problem we did not discuss is that most political scientists lack the skills and equipment necessary to conduct a study including fEMG measures. Most of the authors of this article have the luxury of working at a university with a faculty-level lab with the relevant equipment and expertise. We did not have to pay the start-up cost of approximately €30,000 to get the laptops, wires, signal amplifier, and

relevant software to conduct an fEMG experiment. To run an experiment is fairly cheap, because you have few disposables: wires, adhesive gel or stickers, and cleaning utensils. To get expertise is, of course, more complicated. We hope that this article has helped make fEMG more accessible to political scientists by providing clear recommendations on various processing steps. Yet we recommend that scholars interested in using fEMG seek collaborations with teams already using fEMG protocols.

As a final note, we want to say that fEMG is not a miracle measure. However, we believe it is part of a larger effort to transform the study of political behavior. So far, the field has relied almost exclusively on self-reported measures of feelings, beliefs, and opinions. As a consequence, the processes that lead to such reports are ignored, and we fail to appreciate the mechanisms underlying feelings, beliefs and attitudes. On the positive side, political scientists are increasingly using techniques that tap into unconscious or preconscious processes, such as the implicit association task, automated emotion recognition algorithms based on computer vision techniques, EEG, fMRI, and psychophysiological measures such as skin conductance, heart rate, and, the focus of this article, fEMG. We hope to place this study into this growing body of work investigating the affective-cognitive mechanisms behind political attitudes, beliefs, and feelings.

Acknowledgements. Support for this research was provided by the Dutch National Science Foundation (B.N.B., G.S. & M.R.), the Royal Dutch Academy of Sciences (B.N.B., G.S., & M.R.), the broadcaster NTR (B.N.B., G.S., & M.R.), BKB: The Campaign Agency (B.N.B., G.S., & M.R.), the University of Amsterdam (B.N.B., G.S., & M.R.), the Amsterdam School of Communication Research (B.N.B.), the Research Priority Area Communication (B.N.B.), the European Union's Horizon 2020 research and innovation programme under grant agreement No. 750443 (B.N.B.), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 759079 (G.S.), and an NWO-Veni grant No 451-16-020 (M.R.). We want to thank the organizers of "Lowlands Science" and in particular Noortje Jakobs, Rewan Janssen, and the late Erik van Bruggen for the facilitation of the data collection during Lowlands. Gé Teunissen, Sjoukje Kerman, and Robin Meijers facilitated the creation of the stimulus material. We want to thank the team of the Dutch TV-show "De Kennis van Nu" and in particular Dirk de Bekker, Diederik Jekel, Ines Kaal, Susanne Linssen, Elisabeth van Nimwegen, Roland Vissers, and Marcia van Woensel for the inspiring conversations about our project and the facilitation of the data collection at Design 2. Throughout this project, Ming Boyer, Boris van den Berg, Cas Woudstra, Emke de Vries, Sander Kunst, Judith Meijer, Denise van de Wetering, and Myrthe Willems provided excellent assistance during the data collection, and we want to thank Bert Molenkamp, Marco Teunisse, and Jasper Wijnen for their technical assistance.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/pls.2023.26>.

Data availability statement. This article earned the Open Data, Open Materials, and Preregistration badges for open science practices. The data, replication code, and preregistration for this study are available at <https://osf.io/2wd8gj/>.

Authors contributions. Conceptualization: G.S.; Data curation: G.S.; Formal analysis: G.S.; Funding acquisition: G.S., B.N.B. and M.R.; Investigation: G.S., M.D.H., I.R., N.F., B.N.B. and M.R.; Methodology: G.S., M.D.H., B.N.B. and M.R.; Project administration: G.S.; Resources: G.S.; Supervision: B.N.B. and M.R.; Validation: G.S.; Visualization: G.S.; Writing – original draft: G.S., M.D.H. and I.R.; Writing – review & editing: G.S., M.D.H., I.R., N.F., B.N.B. and M.R.

References

- Achaibou, A., Pourtois, G., Schwartz, S., & Vuilleumier, P. (2008). Simultaneous recording of EEG and facial muscle reactions during spontaneous emotional mimicry. *Neuropsychologia*, 46(4), 1104–1113.
- Arceneaux, K., Bakker, B. N., & Schumacher, G. (2024). *Being of one mind: Does alignment in physiological responses and subjective experiences shape political ideology?* Political Psychology, OSF. <https://doi.org/10.17605/OSF.IO/AQHSP>
- Bakker, B. N., Schumacher, G., Gothreau, C., & Arceneaux, K. (2020). Conservatives and liberals have similar physiological responses to threats. *Nature Human Behaviour*, 4(6), 613–621.
- Bakker, B. N., Schumacher, G., & Homan, M. D. (2020). Yikes! Are we disgusted by politicians? *Politics and the Life Sciences*, 39(2), 135–153.
- Bakker, B. N., Schumacher, G., & Rooduijn, M. (2021). Hot politics? Affective responses to political rhetoric. *American Political Science Review*, 115(1), 150–164.
- Bhowmik, S., Jelfs, B., Arjunan, S. P., & Kumar, D. K. (2017). Outlier removal in facial surface electromyography through Hampel filtering technique. In *2017 IEEE Life Sciences Conference (LSC)* (pp. 258–261). IEEE. <https://doi.org/10.1109/LSC.2017.8268192>

- Blascovich, J., Vanman, E. J., Mendes, W. B., & Dickerson, S. (2011). *Social psychophysiology for social and personality psychology*. Sage Publications.
- Boussalis, C., Coan, T. G., Holman, M. R., & Muller, S. (2021). Gender, candidate emotional expression, and voter reactions during televised debates. *American Political Science Review*, *115*(4), 1242–1257.
- Bucy, E. P., & Bradley, S. D. (2004). Presidential expressions and viewer emotion: Counterempathic responses to televised leader displays. *Social Science Information*, *43*(1), 59–94.
- Bucy, E. P., & Bradley, S. D. (2011). What the body can tell us about politics: The use of psychophysiological measures in political communication research. In E. P. Bucy and R. L. Holbert (Eds.), *The sourcebook for political communication research: Methods, measures, and analytical techniques* (pp. 525–540). Routledge.
- Cacioppo, J. T., Bush, L. K., & Tassinari, L. G. (1992). Microexpressive facial actions as a function of affective stimuli: Replication and extension. *Personality and Social Psychology Bulletin*, *18*(5), 515–526.
- Cacioppo, J. T., Petty, R. E., Losch, M. E., & Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology*, *50*(2), 260–268.
- Cacioppo, J., Tassinari, L. G., & Berntson, G. G. (2007). *The handbook of psychophysiology*. Cambridge University Press.
- Chapman, H. A., Kim, D. A., Susskind, J. M., & Anderson, A. K. (2009). In bad taste: Evidence for the oral origins of moral disgust. *Science*, *323*(5918), 1222–1226.
- Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, *11*(1), 86–89.
- Dimberg, U., Thunberg, M., & Grunedal, S. (2002). Facial reactions to emotional stimuli: Automatically controlled emotional responses. *Cognition & Emotion*, *16*(4), 449–471.
- Ensari, N., Kenworthy, J. B., Urban, L., Canales, C. J., Vasquez, E., Kim, D., & Miller, N. (2004). Negative affect and political sensitivity in crossed categorization: Self-reports versus EMG. *Group Processes & Intergroup Relations*, *7*(1), 55–75.
- Evers, C., Hopp, H., Gross, J. J., Fischer, A. H., Manstead, A. S., & Mauss, I. B. (2014). Emotion response coherence: A dual-process perspective. *Biological Psychology*, *98*, 43–49.
- Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, *23*(5), 567–589.
- Friesen, A., Smith, K. B., & Hibbing, J. R. (2017). Physiological arousal and self-reported valence for erotica images correlate with sexual policy preferences. *International Journal of Public Opinion Research*, *29*(3), 449–470.
- Hess, U., Arslan, R., Mauersberger, H., Blaison, C., Dufner, M., Denissen, J. J., & Ziegler, M. (2017). Reliability of surface facial electromyography. *Psychophysiology*, *54*(1), 12–23.
- Hietanen, J. K., Surakka, V., & Linnankoski, I. (1998). Facial electromyographic responses to vocal affect expressions. *Psychophysiology*, *35*(5), 530–536.
- Homan, M. D., Schumacher, G., & Bakker, B. N. (2022). Facing emotional politicians: Do emotional displays of politicians evoke mimicry and emotional contagion? *Emotion*, *23*(6), 1702–1713.
- Keltner, D., & Gross, J. J. (1999). Functional accounts of emotions. *Cognition & Emotion*, *13*(5), 467–480.
- Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261–273.
- Larsen, J. T., Berntson, G. G., Poehlmann, K. M., Ito, T. A., & Cacioppo, J. T. (2008). The psychophysiology of emotions. In R. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *The handbook of emotions* (3rd ed., pp. 180–195). Guilford Press.
- Larsen, J. T., Norris, C. J., & Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, *40*(5), 776–785.
- LeDoux, J. E., & Pine, D. S. (2016). Using neuroscience to help understand fear and anxiety: A two-system framework. *American Journal of Psychiatry*, *173*(11), 1083–1093.
- Marcus, G. E. (2012). *Political psychology: Neuroscience, genetics and politics*. Oxford University Press.
- Marcus, G. E., Neuman, W. R., & MacKuen, M. B. (2017). Measuring emotional response: Comparing alternative approaches to measurement. *Political Science Research and Methods*, *5*(4), 733–754.
- Masch, L., Gassner, A., & Rosar, U. (2021). Can a beautiful smile win the vote? The role of candidates' physical attractiveness and facial expressions in elections. *Politics and the Life Sciences*, *40*(2), 213–223.
- Mehu, M., Mortillaro, M., Bänziger, T., & Scherer, K. R. (2012). Reliable facial muscle activation enhances recognizability and credibility of emotional expression. *Emotion*, *12*(4), 701–715.
- Morecraft, R. J., Stilwell-Morecraft, K. S., & Rossing, W. R. (2004). The motor cortex and facial expression: New insights from neuroscience. *The Neurologist*, *10*(5), 235–249.
- Neumann, R., Hess, M., Schulz, S. M., & Alpers, G. W. (2005). Automatic behavioural responses to valence: Evidence that facial action is facilitated by evaluative processing. *Cognition and Emotion*, *19*(4), 499–513.
- Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The simulation of smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *The Behavioral and Brain Sciences*, *33*(6), 417–433, discussion 433–480.
- Olshanowski, M., Wróbel, M., & Hess, U. (2020). Mimicking and sharing emotions: A re-examination of the link between facial mimicry and emotional contagion. *Cognition and Emotion*, *34*(2), 367–376.
- Osmundsen, M., Hendry, D., Laustsen, L., Smith, K., & Petersen, M. B. (2021). The psychophysiology of political ideology: Replications, reanalysis and recommendations. *The Journal of Politics*, *84*(1), 50–66.

- Perusquía-Hernández, M., Ayabe-Kanamura, S., Suzuki, K., & Kumano, S.** (2019). The invisible potential of facial electromyography: A comparison of EMG and computer vision when distinguishing posed from spontaneous smiles. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–9). Association for Computing Machinery.
- Potter, R. F., & Bollis, P. D.** (2012). *Psychophysiological measurement and meaning: Cognitive and emotional processing of media*. Routledge.
- Russell, J. A.** (1980). Acircumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.
- Rymarczyk, K., Żurawski, L., Jankowiak-Siuda, K., & Szatkowska, I.** (2018). Neural correlates of facial mimicry: Simultaneous measurements of EMG and BOLD responses during perception of dynamic compared to static facial expressions. *Frontiers in Psychology*, *9*, 52.
- Scherer, K. R., & Moors, A.** (2019). The emotion process: Event appraisal and component differentiation. *Annual Review of Psychology*, *70*, 719–745.
- Schumacher, G., Rooduijn, M., & Bakker, B. N.** (2022). Hot populism? Affective responses to antiestablishment rhetoric. *Political Psychology*, *43*(5), 851–871.
- Stettin, J. E., Hibbing, M. V., Anspach, N. M., Carlson, T. N., Coe, C. M., Hernandez, E., Peterson, J., Stuart, J., & Arceneaux, K.** (2020). Political psychophysiology: A primer for interested researchers and consumers. *Politics and the Life Sciences*, *39* (1), 101–117.
- Sharma, K., Castellini, C., van den Broek, E. L., Albu-Schaeffer, A., & Schwenker, F.** (2019). A dataset of continuous affect annotations and physiological signals for emotion analysis. *Scientific Data*, *6*(1), 196.
- Soroka, S., Gidengil, E., Fournier, P., & Nir, L.** (2016). Do women and men respond differently to negative news? *Politics & Gender*, *12*(2), 344–368.
- Soroka, S. N.** (2019). Skin conductance in the study of politics and communication. In G. Foster (Ed.), *Biophysical measurement in experimental social science research* (pp. 85–104). Elsevier.
- Stewart, T. L., Amoss, R. T., Weiner, B. A., Elliott, L. A., Parrott, D. J., Peacock, C. M., & Vanman, E. J.** (2003). The psychophysiology of social action: Facial electromyographic responses to stigmatized groups predict antidiscrimination action. *Basic and Applied Social Psychology*, *35*(5), 418–425.
- Tassinary, L. G., & Cacioppo, J. T.** (1992). Unobservable facial actions and emotion. *Psychological Science*, *3*(1), 28–33.
- Tassinary, L. G., Cacioppo, J., & Vanman, E.** (2007). The skeletomotor system: System electromyography. In J. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 267–299). Cambridge University Press.
- ’t Hart, B., Struiksma, M. E., van Boxtel, A., & van Berkum, J. J. A.** (2019). Tracking affective language comprehension: Simulating and evaluating character affect in morally loaded narratives. *Frontiers in Psychology*, *10*, 318.
- van Berkum, J. J. A., Struiksma, M. E., & ’t Hart, B.** (2023). Using facial EMG to track emotion during language comprehension. Past, present and future. In M. Grimaldi, E. Shtyrov, & Y. Brattico (Eds.), *Language electrified: Techniques, methods, applications, and future perspectives in the neurophysiological investigation of language* (pp. 687–729). Springer.
- van Boxtel, A.** (2010). Facial EMG as tool for inferring affective states. In A. Spink, F. Grieco, O. Krips, L. Loijens, L. Noldus, & P. Zimmerman (Eds.), *Proceedings of measuring behavior 2010* (pp. 104–108). Noldus Information Technology.
- van Oyen Witvliet, C., & Vrana, S. R.** (1995). Psychophysiological responses as indices of affective dimensions. *Psychophysiology*, *32*(5), 436–443.

Cite this article: Schumacher, G., Homan, M. D., Rebasso, I., Fasching, N., Bakker, B. N., & Rooduijn, M. (2024). Establishing the validity and robustness of facial electromyography measures for political science. *Politics and the Life Sciences*, 1–18. <https://doi.org/10.1017/pls.2023.26>