



Reading About Us and Them: Moral but no Minimal Group Effects on Language-Induced Emotion

Björn 't Hart¹, Marijn Struiksma¹, Anton van Boxtel² and Jos J. A. van Berkum^{1*}

¹Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, Netherlands, ²Cognitive Neuropsychology, Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands

OPEN ACCESS

Edited by:

Pia Knoefeler,
Humboldt University of Berlin,
Germany

Reviewed by:

Johanna Maria Kissler,
Bielefeld University, Germany
David A. Havas,
University of Wisconsin–Whitewater,
United States

*Correspondence:

Jos J. A. van Berkum
j.vanberkum@uu.nl

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 31 July 2020

Accepted: 27 April 2021

Published: 20 May 2021

Citation:

't Hart B, Struiksma M, van Boxtel A
and van Berkum JJA (2021) Reading
About Us and Them: Moral but no
Minimal Group Effects on Language-
Induced Emotion.
Front. Commun. 6:590077.
doi: 10.3389/fcomm.2021.590077

Many of our everyday emotional responses are triggered by language, and a full understanding of how people use language therefore also requires an analysis of how words elicit emotion as they are heard or read. We report a facial electromyography experiment in which we recorded corrugator supercilii, or “frowning muscle”, activity to assess how readers processed emotion-describing language in moral and minimal in/outgroup contexts. Participants read sentence-initial phrases like “Mark is angry” or “Mark is happy” after descriptions that defined the character at hand as a good person, a bad person, a member of a minimal ingroup, or a member of a minimal outgroup (realizing the latter two by classifying participants as personality “type P” and having them read about characters of “type P” or “type O”). As in our earlier work, moral group status of the character clearly modulated how readers responded to descriptions of character emotions, with more frowning to “Mark is angry” than to “Mark is happy” when the character had previously been described as morally good, but *not* when the character had been described as morally bad. Minimal group status, however, did not matter to how the critical phrases were processed, with more frowning to “Mark is angry” than to “Mark is happy” across the board. Our morality-based findings are compatible with a model in which readers use their emotion systems to simultaneously simulate a character’s emotion and evaluate that emotion against their own social standards. The minimal-group result does not contradict this model, but also does not provide new evidence for it.

Keywords: psycholinguistics, communication, emotion, embodiment, morality, minimal groups, EMG, facial electromyography

INTRODUCTION

Part of the attraction of reading a story is that we can vicariously experience what it is like to be somebody else. For example, we can experience happiness when characters in a story find love, frustration when they quarrel, and sadness when they break up—all the while reclining in our armchairs or waiting for the train. Such vicarious experiences can take our mind off things or help us pass the time, provide entertainment, and help us learn about others, life, and possibly even ourselves.

Interestingly, “vicarious experience” may well have a literal meaning here. Partly motivated by the realization that the meaning of at least *some* concepts must be grounded in actual bodily experience (Barsalou, 2008), research on embodied language processing has indicated that reading or hearing a word can lead to a *simulation* of concrete experiences involving the concept, via the neural re-instantiation of perceptual, motor and other experience-induced states associated with what the

concept or phrasal combination of concepts is *about* (Barsalou, 2009; Vigliocco et al., 2009; Glenberg, 2011; Glenberg and Gallese, 2012; Havas and Matheson, 2013; Zwaan, 2014; Winkielman et al., 2015; Zwaan, 2016; Fingerhut and Prinz, 2018; Winkielman et al., 2018). For example, reading action words like “kick” or “pick” leads to activation of the motor cortex involved in *actually realizing* the described movements (Pülvermüller et al., 2005; Willems and Casasanto, 2011), and reading phrases such as “he saw an eagle in the sky” leads to a perceptual simulation of the described situation (Zwaan et al., 2002; Zwaan and Pecher, 2012). Such research suggests that when people process emotion words like “happy” or “angry”, they may actually reuse emotion-related neural systems (Anderson, 2010) to mentally simulate the emotional state described by the language at hand.

Compatible with this simulation idea, studies that use electromyography (EMG) to track subtle facial muscle activity have suggested that simply reading or hearing “angry” or “she is angry” leads to rapid contraction of the corrugator supercilli or ‘frowning muscle’, and, conversely, that reading or hearing “happy”, or “she is happy” leads to rapid contraction of the zygomaticus major, the cheek muscle involved in smiling (e.g., Foroni and Semin, 2009; Glenberg et al., 2009; Foroni and Semin, 2013; Künecke et al., 2015; Fino et al., 2016; see van Berkum et al., in press for review). The central idea here is not that people need to actually *move their face* to make sense of emotion words and phrases, but that the comprehension of an emotion word involves the spontaneous partial reinstatement of the described emotional state (including traces of the associated facial expression), as if one is having the emotion oneself. This reinstatement would occur as part of the retrieval of word meaning from memory (e.g., Foroni and Semin, 2009; Künecke, et al., 2015), and/or as part of constructing a situation model in which some concrete character is having an emotion (e.g., Glenberg et al., 2009; Fino et al., 2016).

Evidence that readers use their emotion systems to *simulate* linguistic meaning poses an interesting puzzle, because during everyday language comprehension, people obviously also need their emotion systems for their *primary* function, which is to—consciously or unconsciously—*evaluate* how events in the world relate to their own concerns (e.g., Lazarus, 1991; Frijda, 2007; Tooby and Cosmides, 2008; Scarantino, 2014; van Berkum, 2018; Scherer and Moors, 2019; see van Berkum, in press, for review). Emotional evaluation is what makes us feel good over a verbal compliment, scared when receiving an unfavourable medical diagnosis, worried over what we read in the newspaper, or surprised by a fictional character’s actions in a novel. These everyday examples suggest that we continuously use our emotion systems to evaluate what we read or hear. So how does such language-driven emotional evaluation mesh with language-driven emotion simulation? When processing language, do we *simultaneously* use our emotion systems to simulate somebody else’s described emotion *and* to evaluate, i.e., have our own emotions about, what is described? If so, how? And if not, which of the two potential uses of our emotion systems receives priority?

We explored this issue in two prior EMG-studies (’t Hart et al., 2018; ’t Hart et al., 2019), where we embedded phrases like “Mark

was angry” or “Mark was happy” in a narrative context that was designed to promote simulation as well as evaluation. Specifically, we compared the processing at negative or positive emotional state adjectives, e.g., “angry” vs. “happy”, in stories where the character experiencing those states had previously displayed morally good or morally bad behavior. We reasoned that any lexical and/or situation model simulation should in principle always generate more negative emotion at “Mark was angry” than at “Mark was happy”, independent of whether the character was morally good or bad. The reader’s moral evaluation of events, however, should depend on *who* the event is happening to, at least to some extent. When something bad happens to a morally *good* character, this should typically be seen as “unfair” or otherwise undesirable, and something good happening to him or her should be seen as desirable (as in a “feel-good” movie). Something bad happening to a morally bad character, however, should typically elicit a sense of fairness or “justice being served”, perhaps even a bit of Schadenfreude (e.g., Feather and Nairn, 2005; Singer et al., 2006; Leach and Spears, 2009; Cikara and Fiske, 2012), and something good happening to him or her should typically be seen as “unfair”.

Because our logic was cast in terms of the valence (positivity or negativity) of language-induced emotion, we looked for traces of reader emotion by recording EMG over the corrugator or “frowning” muscle, a sensitive and reliable indicator of valence (e.g., Larsen et al., 2003; Höfling et al., 2020; see van Boxtel, in press; van Berkum et al., in press, for reviews). The EMG-results were very clear. In both studies, phrases like “Mark was angry” led to stronger corrugator activity than phrases like “Mark was happy” when the character had previously acted in a morally good way, but *not* when the character had previously acted in a morally bad way—in the latter case, corrugator activity to negative and positive emotion adjectives did not differ. Because simple models involving *only* simulation or evaluation cannot easily explain these results, we converged on a *multiple-drivers* model of corrugator activity during language comprehension (see **Figure 1**; adapted from van Berkum et al., in press), which, in our materials, would involve both simulation (at the lexical and/or situation model) *and* evaluation of what is being asserted.

Our multiple-drivers account proposed that in the case of a good character, negative emotion induced by simulation at “Mark is angry” adds up with the negative emotional evaluation associated with an undesirable outcome, and positive emotion induced by simulation at “Mark is happy” adds up with the positive emotional evaluation associated with a desirable outcome, leading to a much stronger corrugator activity at negative emotion words, as compared to positive ones. In the case of a bad character, however, negative emotion induced by simulation at “Mark is angry” is *counteracted* by the positive emotional evaluation of a “fair” outcome, and positive emotion induced by simulation at, e.g., “Mark is happy” is counteracted by the negative emotional evaluation of an “unfair” outcome, to such an extent that, with our materials, no net valence effect at negative vs. positive emotion words remains.

While adequate, this account of the ’t Hart et al. (2018), ’t Hart et al. (2019) results is only modestly parsimonious, as it explains a null result as the net effect of two counteracting forces. The

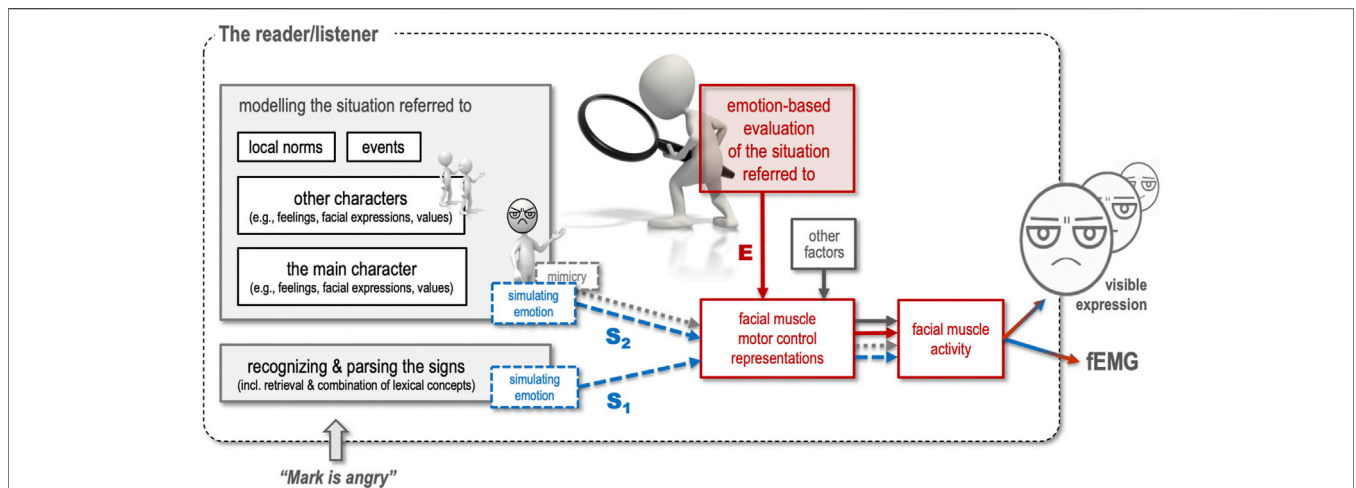


FIGURE 1 | A multiple-drivers model of emotional facial expressions and the associated EMG effects induced by language in simple (e.g., laboratory) communicative contexts. Apart from language-induced emotion simulation and emotional evaluation, the model also acknowledges mimicry and other factors as potential drivers (see the Discussion). Adapted from the *fALC* model, a broader model of what drives emotional facial expression during language processing (see van Berkum et al., in press).

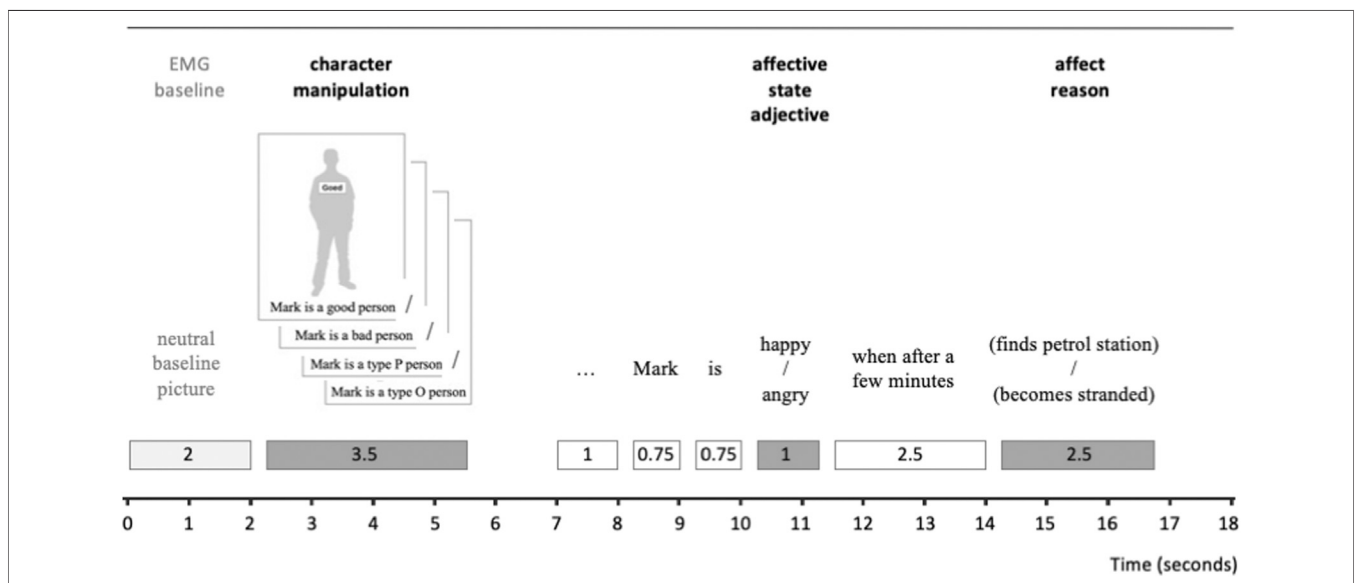


FIGURE 2 | Schematic trial structure, with (partly abbreviated) example item, component presentation duration in seconds, and associated EMG segment labels.

principal aim of the current experiment was to try to expose the counteracting drivers by “downtuning” the force of emotional evaluation. Morality is deeply intertwined with ingroup cohesion and intergroup competition (Haidt, 2012; Greene, 2014). As people tend to consider themselves as morally virtuous (Tappin and McKay, 2016), morally good people can be said to belong to a highly relevant ingroup, associated with strong positive feelings, and morally bad people can be said to belong to a highly relevant outgroup, associated with strong negative feelings. Taking this morality-based grouping as our starting point, we turned to a *minimal* group manipulation (e.g., Tajfel et al., 1971;

Diehl, 1990) to define in- and outgroups that are associated with *attenuated* emotional evaluations. In a minimal group paradigm, participants are divided into two or more groups on the basis of arbitrary characteristics, such as a coin flip, shirt color, or fake personality test score. Such classifications, although arbitrary, lead to subtle in- and outgroup biases with a preference for “us” and a dispreference for “them”, in face-to-face contact, but also when processing language (e.g., Morrison et al., 2012). We reasoned that with phrases like “Mark is angry”, the force of group-based emotional evaluation (e.g., a bit of Schadenfreude when something bad happens to an outgroup member) would be

weaker when the characters at hand belonged to *minimal* outgroup than when they belonged to a moral outgroup. With a smaller contribution of group-dependent evaluation, and on the assumption that at phrases like “Mark is angry”, language-driven lexical or situation-model simulation would remain the same, a simulation-based valence effect should begin to show up in corrugator EMG.

Before the critical task, participants filled out a fake personality questionnaire, which invariably scored them as a “type P” rather than a “type O” personality. Each participant subsequently viewed a series of composite stimuli as their EMG was being recorded. At each trial (see **Figure 2**), participants first saw a silhouette of a character together with a moral (“good” or “bad”) or a minimal (“type P” or “type O”) group classification, and then read a sentence in which the character was described having a positive or negative emotion because of some particular reason.

Each sentence contained three EMG-relevant segments. At the *character manipulation segment*, we predicted that designating a character as bad should elicit more corrugator activity than designating a character as good. Based on evidence for a mild negative bias toward minimal outgroups (e.g., Diehl, 1990), designating a character as type O to participants who themselves have been designated as type P might also elicit more corrugator activity, albeit to a lesser extent than with a *moral* outgroup designator.

At the *affective state adjective segment* (e.g., “angry” vs. “happy”), the critical segment for our study, predictions also depended on whether characters had been designated in terms of a moral or a minimal group dimension. For characters designated as morally good or bad, we expected to replicate the crucial corrugator EMG pattern observed in our two earlier EMG-studies: substantially more frowning at “(Mark was) angry” than at “(Mark was) happy”—i.e., a large adjective valence effect—for morally good characters because of simulation- and evaluation-driven activity adding up, but a zero, or close to zero, adjective valence effect for morally bad characters because of simulation- and evaluation-driven activity counteracting each other.

For characters designated as belonging to a *minimal ingroup* (type P, the same type as the participant), we again expected more frowning at “(Mark was) angry” than at “(Mark was) happy” because of simulation- and evaluation-driven activity adding up. However, because the fate of a member of an ingroup that the reader only weakly associates with should matter less than the fate of a member of an ingroup member that the reader strongly associates with, the size of the adjective valence effect with minimal ingroup characters should be smaller than with moral ingroup characters. Furthermore, for characters designated as belonging to a *minimal outgroup* (type O), we predicted that the negative emotion associated with simulating the meaning of “(Mark was) angry”, compared to “(Mark was) happy”, would not be fully counteracted by a weak outgroup-contingent positive evaluation of this particular outcome, leading to a small net adjective valence effect. Assuming *some* minimal ingroup favoritism, the net adjective valence effect should still be a bit larger with minimal ingroup characters (where any evaluation still aligns with simulation) than with minimal outgroup

characters (where it opposes simulation). But in *both* minimal group cases, adjective valence effects should lie between the adjective valence effects in the two moral group cases. With a smaller evaluation bias, EMG-responses should in the minimal-group part of the design be dominated more by language-driven simulation.

At the *affect reason segment*, the reason for the character’s emotion is revealed. Because the input provided here is distributed over a multi-word clause, with the reasons for positive and negative affect usually differing on more than one word, descriptions of affect reasons were much less well-controlled in terms of lexical variables and time-locking precision. We therefore made no detailed predictions for this segment. However, in line with the results of the one prior study where we had also temporally separated the affect reason from the affective state adjective (’t Hart et al., 2019), we expected a renewed phasic corrugator response to reasons for negative emotion, particularly for moral ingroup characters, but possibly also for other characters.

MATERIALS AND METHODS

Participants

We recruited 64 native speakers of Dutch (58 female and 6 male) aged between 18 and 27 ($M = 21.5$, $SD = 2.2$) from the Utrecht University Humanities faculty participant database, for an experiment on reading that focused on language, emotions and personality. None of the participants had been diagnosed with dyslexia, had taken Botox® injections in the face, or had participated in the earlier ’t Hart et al. (2018), ’t Hart et al. (2019) studies. Research procedures complied with Netherlands Code of Conduct for Academic Practice and with the Declaration of Helsinki. Participants gave written informed consent after reviewing a form that detailed the nature of the materials and the procedure, and emphasized their right to withdraw consent at any time without having to provide a reason and without losing financial compensation (€ 12,-). The study was approved by the Linguistics Chamber of the Faculty of Humanities Ethics Assessment Committee at Utrecht University.

Stimulus Materials and Design

We extracted the 64 critical story-final sentences from the larger Dutch stories of the ’t Hart et al. (2019) study, to now present them in a different context. All sentences described an affective episode according to the structure: <Character name> is/feels/becomes/notes <positive or negative affective state adjective> <neutral connector phrase> <reason for the affective state>. Positive and negative critical adjectives had comparable average length (positive: 8.0 letters, range 4–13; negative: 7.3 letters, range 4–14), and so did the reason fragments (positive: 40.3 characters, range 22–52; negative: 42.5 characters, range 21–58, including spaces). Each of the critical sentences was preceded by a neutral silhouette image of a male or female character together with a moral or a minimal group classification, signaled by a badge on their chest that said

either “good” (moral ingroup), “bad” (moral outgroup), “type P” (minimal ingroup) or “type O” (minimal outgroup), as well as by an accompanying qualification underneath the silhouette: “<Character name> is a really good person”, “. . . a really bad person”, “. . . a type P personality” or “. . . a type O personality” (see **Supplementary Section S1**).

Fully crossing character manipulation with critical sentence type yielded eight stimulus variants that realized our $2 \times 2 \times 2$ design: grouping dimension (moral vs. minimal) \times group (ingroup vs. outgroup) \times critical adjective valence (positive vs. negative). We constructed 8 pseudo-randomized 128-trials lists, such that (a) no specific stimulus variant was repeated in a list, (b) each list contained two pseudo-randomized blocks, with 64 moral-group items followed by 64 minimal-group items in four lists, and the reverse block order in the remaining lists, and (c) in each block, 32 items had a male character and 32 a female one. Each participant received one list only.

Procedure and Data Acquisition

After signing an informed consent form, participants first completed a (fake) digital personality test. The 22 items in this test, pseudo-randomly drawn from existing personality tests, queried aspects of personality unrelated to morality (e.g., “Sometimes I really lose myself in music”, “I have fairly fixed habits”, “I never worry”, and “I am very eager to learn”). Unbeknownst to the participants, the test automatically always classified them as “type P”. To make sure participants attended to their classification, they were asked to digitally enter their type themselves, and to wear a badge with a capital P for the remainder of the session.

In the subsequent EMG-task, participants read a series of descriptions of events involving different characters, each preceded by a character description. Apart from trying not to move and blink too much, no other task was imposed. Stimuli were presented with the structure and timing shown in **Figure 2** on a 15.6-inch laptop monitor (Lenovo E531 ThinkPad) positioned at about 60 cm distance, in white on a gray background, with a character silhouette image of approximately 10° vertical angle, a 26 points Times New Roman font for the sentence, and with the same neutral baseline picture of a forest scene presented at the beginning of each trial (providing a mental reset and a trial-specific EMG-baseline). Participants pressed the space bar to advance to the next trial, with their left hand so as to prevent cable movement artifacts. Each block was preceded by two practice trials, and the blocks were separated by a pause that contained a short and easy distractor task. Sentence presentation parameters were identical to that of t Hart et al. (2019).

Facial EMG was recorded at 2048 Hz with a Nexus MKII biosignal system (Mind Media, Roermond-Herten), using reusable Ag/AgCl electrodes with a 2 mm contact surface, placed at standard recording sites over the right corrugator superciliaris and zygomaticus major (Fridlund and Cacioppo, 1986; van Boxtel, in press). As in the t Hart et al. (2018), t Hart et al. (2019) studies, we recorded from the right side of the face only (on average, spontaneous facial expressions do not differ between the left and right side of the face; Ekman et al.,

1981). Also as in our earlier studies, we defined predictions for *corrugator* EMG only. Although the corrugator and the zygomaticus are often used together to assess emotional valence, only the former muscle tracks valence in a relatively monotonic way (*zygomaticus* activity can increase with both positive and very negative stimuli, relative to neutral stimuli; Kunkel, 2018, Chapter 3; Larsen et al., 2003; Lee and Potter, 2018; see van Berkum et al., in press, for review). To allow for comparison to other work, we document average zygomaticus results in the **Supplementary Section S5**, with the raw data available in our online repository (<https://doi.org/10.24416/UU01-YM9VPP>).

After the EMG-task, participants filled out the Adolescent Measure of Empathy and Sympathy (AMES, Vossen et al., 2015), the Moral Foundations Questionnaire (MFQ, Graham et al., 2011), and a structured exit survey. The AMES and MFQ data were of exploratory interest and are reported in the **Supplementary Section S3**. Finally, participants were debriefed and paid. The average total session lasted about 75 min, with about 45 min on the EMG-task.

Data Preparation and Analysis

The raw EMG-data were filtered with a band-pass of 20–500 Hz (48 dB/octave roll-off) and a notch filter at 50 Hz to remove common artifacts (see van Boxtel 2010), followed by signal rectification and segmentation per trial, all using BrainVision Analyzer 2 (BrainProducts, Gilching). A trigger placement error resulted in loss of data for two of originally 64 tested participants. For the remaining 62 participants, we used visual inspection to select maximally long epochs of “quiet signal” (free of extreme bursts) within the 2,000 ms baseline segment, with a minimum length of 500 ms for each muscle. If a continuous artefact-free baseline epoch of at least 500 ms was not found, the trial was excluded from the analysis (resulting in 3.45% lost trials).

After baseline epoch selection, the data were exported to MatLab for further segmentation time-locked to the onset of the character manipulation picture (segment length 3,500 ms), the affective state adjective (segment length 1,000 ms), and the affect reason (segment length 2,500 ms). Each of the resulting EMG segments was then partitioned into consecutive 100-ms bins, known to strike a good balance between sufficient temporal resolution and sufficient random error reduction (van Boxtel, 2010). To reduce random variance both within and between individuals (van Boxtel, 2010), average EMG activity was expressed as a percentage of the pre-stimulus baseline epoch activity level.

The three segments were analyzed separately using the linear mixed models procedure in SPSS (IBM, v25). In linear mixed models the item- and participant variance are estimated simultaneously, resulting in a cross-classified model (Quené and Van den Bergh 2004; Quené and Van den Bergh 2008). For each segment, we constructed models for the corrugator data by iteratively adding potentially relevant components and testing for significant model improvement at each addition (using the likelihood ratio (-2LL difference chi-square) test, $p < 0.05$). We only kept components whose addition

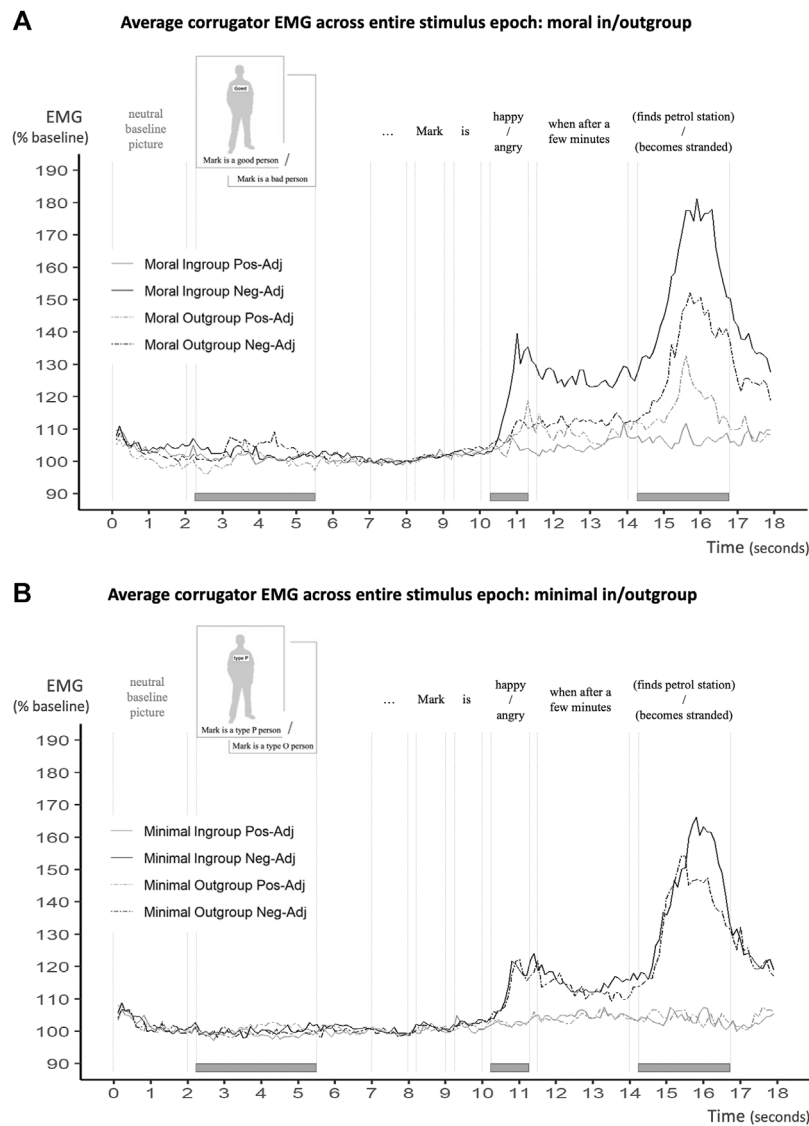


FIGURE 3 | Corrugator EMG across the entire stimulus epoch, for **(A)** moral and **(B)** minimal in- and outgroups, together with a (partly abbreviated) example item, the associated stimulus component presentation onsets and offsets (thin vertical lines) and the three critical EMG segments (gray bars). Note that on average, EMG activity in the first two seconds is above 100% because the values on which the baseline value is based may be a subset of all data points in this 2-s interval (see Methods for EMG baselining procedure).

explained a significant amount of variance or were necessary to test hypothesized interactions. Components that did not significantly improve the model were dropped in the next iteration (Winter, 2020).

Because we were not only interested in average corrugator activity in a segment but also in its development *over time*, we used a growth curve model approach (Peck and Devore, 2008; Mirman, 2015) with specific analysis designs that were optimized for assessing and comparing time trends across conditions. We first modeled participants and items as random factors. To assess the effect of our manipulations on the average activation across an entire segment, we subsequently added grouping dimension (moral vs.

minimal), group (ingroup vs. outgroup), and its interaction as a fixed factor in the model for the character manipulation segment, and grouping dimension (moral vs. minimal), group (ingroup vs. outgroup), affective state adjective/reason valence (positive vs. negative), and their 2-way and 3-way interactions as fixed factors in the model for the affective state adjective and affect reason segments. Afterward the most complex interaction was added to the random part of the model as a random slope. Next, linear, quadratic, and cubic trends were added as covariates in the fixed part of the model. Time trend (e.g., linear) components were added *per condition* to maintain flexibility in building the model, and to avoid forcing the model to fit, for example, a linear trend for all conditions

when only *one* condition contained a significant linear component. All trend components were centered to avoid correlation between trends (fixed effects final model intercepts therefore reflect the average corrugator activity across the entire segment, *not* the level at which corrugator activity intercepts the y -axis). By using trends up to the cubic component, we achieved some flexibility to fit responses without over-fitting or losing explanatory power (Mirman, 2015). Because we were particularly interested in temporal developments, the random part of the models always included random slopes for subjects for each time trend that initially improved the model (as well as standard random intercepts for subject and item).

To facilitate interpretation, in the final model the fixed factors grouping dimension, group and affective state adjective/reason valence were included as a single condition factor, which allowed for a no-intercept model where the estimates of the conditions reflect the segment average corrugator activation. This re-parametrization does not change the -2LL value, and as such still represents the optimal model. While trend components were fitted with a resolution of 100 ms, the associated parameter estimates (e.g., b for a linear slope) are reported on a 1-s basis. For the final model, custom two-tailed t -tests were used to assess theoretically relevant pairwise comparisons between condition averages. Theoretically relevant comparisons between two (e.g., linear) condition-specific trend components were done by explicitly comparing the difference between associated regression weights ($b_1 - b_2$) in a dedicated two-tailed t -test in case both components had been kept in the model, and by resorting to the simple fixed effects t -test for just one of them (e.g., b_1) when the other component had not been included in the final model (which effectively defined b_2 as 0). For each critical segment, **Supplementary Section S2** reports on model construction steps, followed by parameter tests and specific comparisons based on the final model (referring to our online repository for all original statistical analyses documents: <https://doi.org/10.24416/UU01-YM9VPP>).

RESULTS

Figure 3 shows average corrugator EMG responses across the entire stimulus epoch, together with an example item and the associated temporal structure. As can be seen, there is hardly any differential activity in the character manipulation segment, but substantial differential activity in the affective state adjective segment and the affect reason segment. In the following, we discuss these results per segment (see **Supplementary Section S2** for statistical details).

Character Manipulation

The character manipulation designated a character as morally good (moral ingroup), morally bad (moral outgroup), a type P personality (minimal ingroup) or a type O personality (minimal outgroup). **Figure 4** shows the corrugator EMG results for each

condition, time-locked to the onset of the character manipulation picture. For the average corrugator activity across the entire 3.5-s segment, the overall interaction test revealed a significant interaction between grouping dimension \times group (moral ingroup 103.9%, moral outgroup 112.0%, minimal ingroup 104.3%, minimal outgroup 104.4%, $F(4, 134.64) = 1542.44$, $p < 0.001$). We discuss all further effects for moral and minimal groups separately.

Moral In- and Outgroup

As expected, **Figure 4** reveals increased frowning to characters designated as bad (dashed black line), and no such increase for characters designated as good (solid black line). In line with this, average corrugator activity across the entire segment differed significantly at morally good vs. morally bad character descriptions (difference_{ingr-outgr} = -8.02 , $t(184.17) = -2.48$, $p = 0.01$, 95% CI $[-14.41, -1.64]$). As for the trend components, the model fitted a flat line for good character descriptions, indicating no change in corrugator activity at all. For bad character descriptions, two marginal effects hint at the phasic nature of the response, with linear and quadratic trend components further improving the statistical model (see **Supplementary Section S2a**), and the associated b -estimates almost significantly different from zero (positive linear trend $b = 2.19$, $t(62.02) = 1.97$, $p = 0.053$, 95% CI $[-0.03, 4.41]$; negative quadratic trend $b = -2.89$, $t(61.97) = -1.93$, $p = 0.059$, 95% CI $[-5.89, 0.11]$). In all, seeing a silhouette with “bad” accompanied by “X is a really bad person” fairly rapidly elicits a bit of frowning, starting at around 1,000–1,100 msec in the actual (non-modeled) data.

Minimal In- and Outgroup

We had considered that designating a character as a *minimal* outgroup member might increase corrugator activity too, although not to the extent observed for *moral* outgroup designators. However, in **Figure 4**, the corrugator response to characters labeled as type O (dashed gray line) and type P

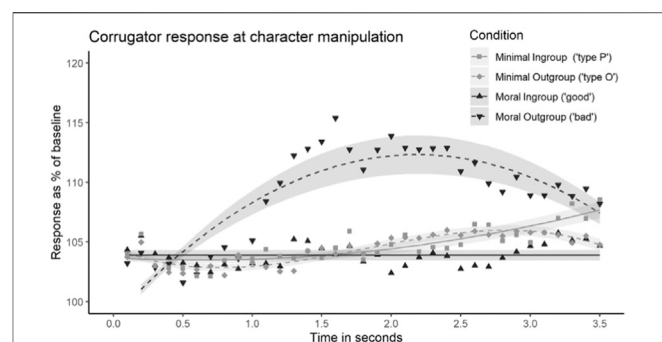
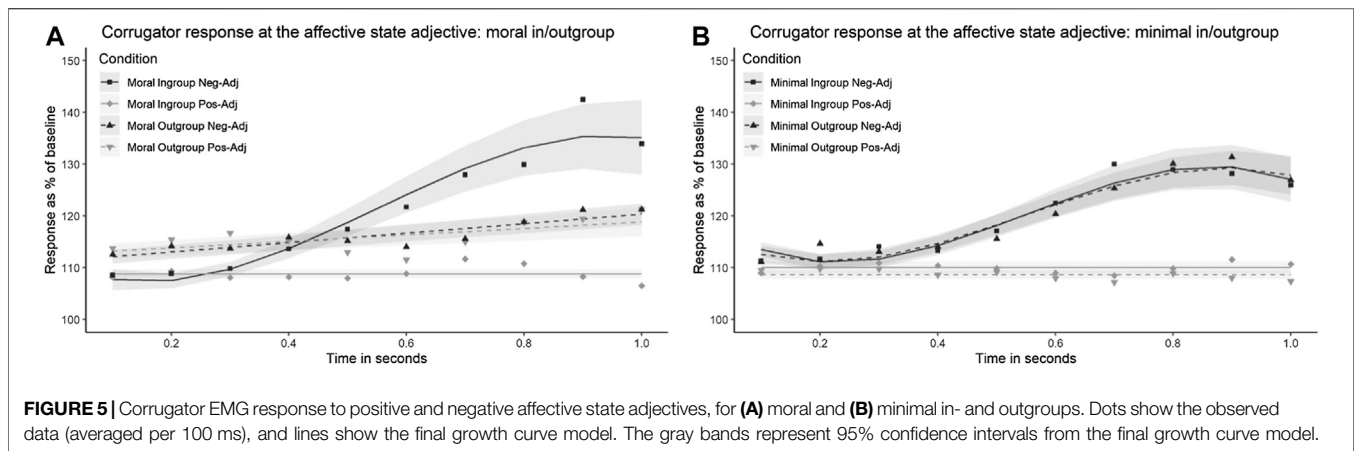


FIGURE 4 | Corrugator EMG response to the character manipulation. Dots show the observed data (averaged per 100 ms), and lines show the final growth curve model (incorporating all intercept and trend parameters that significantly improved the model). The gray bands represent 95% confidence intervals from the final growth curve model.



(solid gray line) are right on top of each other, and a pairwise comparison did not reveal a significant average corrugator activity difference between the two conditions (difference_{ingr-outgr} = -0.18, $t(184.15) = -0.05$, $p = 0.96$, 95% CI [-6.56, 6.21]). Also, trend analysis did not reveal any big differences there either. For minimal outgroup as well as minimal ingroup character descriptions, a positive linear trend component improved the model (see **Supplementary Section S2a**). For minimal outgroup characters, the positive linear trend significantly differed from zero ($b = 2.27$, $t(302.92) = 3.63$, $p < 0.001$, 95% CI [1.04, 3.50]), while for minimal ingroup characters the difference was only marginal ($b = 1.18$, $t(61.90) = 1.89$, $p = 0.06$, 95% CI [-0.07, 2.44]), but the pairwise comparison revealed no significant difference between these two trends (difference_{ingr-outgr} $b = 1.09$, $t(204.66) = 1.23$, $p = 0.22$, 95% CI [-0.66, 2.83]). Higher-order trend analysis revealed a significant but very small cubic trend in the minimal outgroup response (indicating a fall-rise-fall pattern, see **Supplementary Section S2b**), but as can be seen in **Figure 4**, the fitted curves for these two conditions are virtually on top of each other. In all, the corrugator EMG did not show a clear differential response to descriptions of minimal ingroup (type P) or outgroup (type O) characters. Participants did report feeling less similar (range: -3 *not similar at all*, 3 *very similar*) to minimal outgroup characters than to minimal ingroup characters ($M = -1.61$ vs. $M = 0.70$; $M_{diff} = -2.31$, $SD = 1.19$; two-tailed paired-samples t -test $t(61) = -9.06$, $p < 0.001$), but this did not translate to clearly differential EMG activity.

Affective State Adjective

At the affective state adjective (e.g., “happy/angry”), the most critical segment in our study, participants read about positive or negative emotion of the same character. This additional adjective valence factor expands the EMG-analysis to a 2 (grouping dimension: morality vs. minimal group) \times 2 (group: ingroup vs. outgroup) \times 2 (affective adjective valence: positive vs. negative) design. **Figure 5** displays the associated corrugator EMG-responses. Consistent with the first impression, analysis of the average EMG activity across the entire 1-s segment revealed a significant three-way-interaction of these factors ($F(8, 259.41) = 222.51$, $p < 0.001$).

As with the character manipulation, we discuss all further effects for moral and minimal groups separately.

Moral In- and Outgroup

For characters designated as morally good or bad, we expected to replicate the core result of our two earlier EMG-studies: substantially more frowning at “(Mark was) angry” than at “(Mark was) happy” for morally good characters because of simulation- and evaluation-driven corrugator activity adding up, but a zero, or close to zero, adjective valence effect for morally bad characters because of simulation- and evaluation-driven corrugator activity canceling each other out. As can be seen in **Figure 5A**, this is exactly what we observed.

For morally good characters, the EMG response showed a clear and rapid increase in frowning activity at negative state adjectives (solid black line), but no such increase at positive state adjectives (solid gray line), with the signals diverging from about 300–400 ms onwards. Statistical analysis of average EMG during the entire 1-s segment confirmed that participants frowned significantly more when a good character had a negative emotion than when he or she had a positive emotion (difference_{neg-pos} = 13.70, $t(428.27) = 3.13$, $p = 0.002$, 95% CI [5.10, 22.30]). Trends in the response also differed. The model fitted a flat line at positive affective state adjectives but included a significant linear increase in activation at negative adjectives ($b = 56.40$, $t(76.12) = 2.65$, $p = 0.01$, 95% CI [14.03, 98.78]), as well as a cubic trend (see **Supplementary Section S2d**).

For morally bad characters, the EMG response showed no such differential increase in frowning activity at negative state adjectives (dashed black line), relative to positive state adjectives (dashed gray line). The average segment EMG analysis confirmed that average frowning during this 1-s interval did not statistically differ at negative vs. positive state adjectives (difference_{neg-pos} = 0.72, $t(428.64) = 0.17$, $p = 0.87$, 95% CI [-7.88, 9.32]). As can be seen in **Figure 5A**, modest upward linear trend components improved the overall model, with a marginally significant b in the case of negative adjectives ($b = 9.67$, $t(56.35) = 1.92$, $p = 0.06$, 95% CI [-0.42, 19.75]), but not for positive adjectives ($b = 6.57$, $t(62.27) = 1.06$, $p = 0.30$, 95% CI [-5.86, 19.00]), and no significant difference between the two linear trends (difference_{neg-pos} $b = 3.10$, $t(115.74) = 0.39$, $p = 0.70$, 95% CI [-12.74, 18.95]).

Minimal In- and Outgroup

As can be seen in **Figure 5B**, we did not obtain the expected pattern of results in this part of the design. For minimal ingroup characters (i.e., designated before as a type P person), the EMG response showed a clear and rapid increase in frowning activity at negative state adjectives (solid black line) starting around 300–400 msec, and no such increase at positive state adjectives (solid gray line). However, for minimal *outgroup* characters (i.e., designated before as a type O person), the exact same result was observed, with a clear and rapid increase in frowning activity at negative state adjectives (dashed black line) starting at around 300–400 msec, and no such increase at positive state adjectives (dashed gray line). This suggests that minimal group membership did not modulate the net differences between responses to adjectives like “angry” and “happy”.

The statistical analysis of average EMG during the entire 1-s segment confirmed that participants frowned significantly more when a character had a negative emotion than when he or she had a positive emotion, for minimal ingroup characters (difference_{neg-pos} $b = 10.29$, $t(427.92) = 2.35$, $p = 0.02$, 95% CI [1.69, 18.89]), as well as for minimal outgroup characters (difference_{neg-pos} $b = 11.76$, $t(428.76) = 2.69$, $p = 0.01$, 95% CI [3.16, 20.36]). Furthermore, there was no difference in average frowning activity at *negative* state descriptions of minimal ingroup vs. outgroup members (**Figure 5B**, black lines; difference_{ingrp-outgrp} $b = 0.09$, $t(428.82) = 0.02$, $p = 0.98$, 95% CI [-8.51, 8.69]), nor at *positive* state descriptions of minimal ingroup vs. outgroup members (**Figure 5B**, gray lines; difference_{ingrp-outgrp} $b = 1.56$, $t(428.02) = 0.36$, $p = 0.72$, 95% CI [-7.03, 10.16]).

As clearly evident in **Figure 5B**, the *temporal development* of the corrugator EMG signal at positive and negative state adjectives also did not vary as a result of minimal group membership. Negative affective state adjectives led to a significant linear increase in corrugator activity both for minimal ingroup members ($b = 42.85$, $t(55.50) = 2.29$, $p = 0.03$, 95% CI [5.34, 80.36]) and for minimal outgroup members ($b = 39.76$, $t(144.53) = 3.47$, $p = 0.001$, 95% CI [17.12, 62.40]), with no significant difference between the two (difference_{ingrp-outgrp} $b = 3.09$, $t(99.49) = 0.14$, $p = 0.89$, 95% CI [-40.45, 46.64]). The model also included a negative cubic trend at negative affective state adjectives for minimal ingroup and outgroup members, but the patterns did not differ (see **Supplementary Section S2d**).

Our multiple-drivers model, and our additional assumption of weaker (but non-zero) group-dependent evaluation in the minimal group case than in the moral group case, had led us to expect that the differential adjective valence effect (e.g. “angry” vs. “happy”) would be smaller with minimal (type P) ingroup characters than with moral (good) ingroup characters, with the corrugator signal to a minimal ingroup character experiencing negative emotion to end up *below* that to a moral ingroup character experiencing the same emotion (i.e., black solid line in **Figure 5B** lower than black solid line in **Figure 5A**). However, although descriptively the EMG-response pattern is in the right direction, pairwise comparisons showed no significant

difference between these two signals, neither in terms of the 1-s segment average, nor in terms of the linear or cubic trend component (all p 's > 0.63). Also, we had expected the corrugator signal to a minimal ingroup character experiencing positive emotion to end up above that to a moral ingroup character experiencing the same emotion (i.e., gray solid line in **Figure 5B** higher than gray solid line in **Figure 5A**). However, both moral- and minimal ingroup-positive were fitted with a flat line that did not significantly differ in elevation ($p = 0.79$). For a full report of all estimates and comparison see **Supplementary Section S2d**.

All in all, in the morality part of the design, we replicate the core results of our earlier work: corrugator responses to negative and positive emotion adjectives strongly depend on who is experiencing the emotion described. In the minimal-group part, however, the identity of the character does not matter at all, with equally large adjective valence effects for minimal ingroup and minimal outgroup characters.

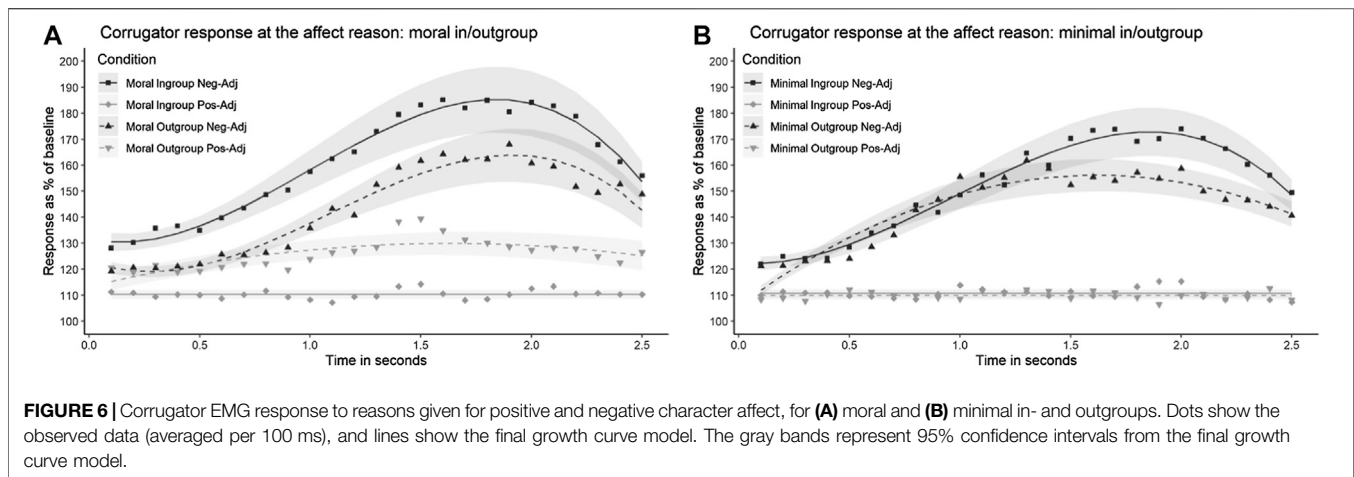
Affect Reason

At the affective reason segment, participants read about events that provided a reason for the character's emotion. The analysis at this segment involves a 2 (grouping dimension: morality vs. minimal group) \times 2 (group: ingroup vs. outgroup) \times 2 (affect reason valence: positive vs. negative) design. **Figure 6** displays the associated corrugator EMG-responses. One striking aspect of the EMG-patterns in **Figures 6A and B** is the renewed phasic corrugator response in all four conditions motivating a character's negative emotion, which suggests that these sentence fragments contained enough information to elicit *additional* differential corrugator activity. Also, as evident from the entire-epoch **Figure 3**, these new phasic corrugator responses ride on top of relatively stable corrugator differences that emerged at the prior affective state adjective, and that lasted for several more seconds, throughout the intermediate neutral connector phrase (e.g., “when after a few minutes”). Because corrugator activity is expressed as a percentage of the same pre-stimulus baseline at all three critical segments, these longer-lasting state adjective effects are responsible for the pre-existing differences at 0 s in **Figure 6**.

Analysis of the average EMG activity across the entire 2.5-s affect reason segment revealed a significant three-way-interaction of grouping dimension, group, and affect reason valence ($F(8, 255.28) = 41.79$, $p < 0.001$), an interaction that to some extent reflects these earlier adjective-triggered EMG effects. As before, we discuss all further effects for moral and minimal groups separately.

Moral In- and Outgroup

In line with t Hart et al. (2019), we had expected a renewed phasic corrugator response to events that were the reason for negative, as opposed to positive, emotion, particularly for good characters, but possibly also for bad characters. As can be seen in **Figure 6A**, there is indeed a clear and substantial increase for negative events befalling good characters (solid black line) and a flat-line response for positive events befalling these same characters (solid gray line). A smaller but descriptively comparable response difference emerged for bad characters, with negative events (dashed black



line) eliciting somewhat higher corrugator EMG activity than positive events (dashed gray line).

For good characters (solid lines), average corrugator activation across the segment was indeed significantly higher for negative events than for positive events (difference_{neg-pos} $b = 62.20$, $t(430.05) = 5.20$, $p < 0.001$, 95% CI [38.69, 85.71]). Furthermore, while negative events elicited a significant linear increase in corrugator activity ($b = 41.70$, $t(61.46) = 2.42$, $p = 0.02$, 95% CI [7.20, 76.20]), modulated by significant quadratic and cubic trends (both $p < 0.04$, see **Supplementary Section S2f**), positive events elicited a flat-line EMG response. For bad characters (dashed lines), however, average corrugator activation at negative and positive events was not significantly different (difference_{neg-pos} $b = 21.38$, $t(438.89) = 1.78$, $p = 0.08$, 95% CI [-2.25, 45.01]). Also, although negative events elicited an almost significant linear increase in corrugator activity ($b = 38.64$, $t(61.72) = 1.94$, $p = 0.06$, 95% CI [-1.16, 78.44]) while positive events did not ($b = 4.27$, $t(61.76) = 1.01$, $p = 0.32$, 95% CI [-4.20, 12.74]), the difference between the two linear trends was not significant (difference_{neg-pos} $b = 34.37$, $t(67.30) = 1.69$, $p = 0.10$, 95% CI [-6.25, 75.00]). Both events did elicit a significant quadratic trend ($p = 0.02$ and $p = 0.01$ for negative and positive events respectively, see **Supplementary Section S2f**).

With the two EMG-signals for good characters being much (and significantly) further apart than the two EMG-signals for bad characters, **Figure 6A** could be taken to suggest that readers are again more sensitive to the fate of good characters than to that of bad ones, just as at the adjective. However, the elevated average corrugator response to negative over positive events with moral ingroup characters is to a large extent already present at 0 s, and is as such presumably largely due to spill-over from the earlier adjective effect (see particularly **Figure 3**, and compare the EMG-pattern at segment onset in **Figure 6A** to the EMG-pattern at segment offset in **Figure 5A**). We therefore cannot confidently model this pattern of results as renewed differential sensitivity to the fate of good and bad characters. In all, the only informative result in this part of the design is a significant phasic rise-fall response when reading about bad events (happening to good or

bad people alike), and when reading about good events happening to bad people.

Minimal In- and Outgroup

As can be seen in **Figure 6B**, the dominant pattern of results is that of large phasic corrugator responses to negative events befalling both minimal ingroup (“type P”) and outgroup (“type O”) characters, and no responses to positive events. Statistical analysis confirms this. For minimal ingroup characters, average corrugator activation across the segment was higher for negative events than for positive events (difference_{neg-pos} $b = 52.06$, $t(429.90) = 4.35$, $p < 0.001$, 95% CI [28.55, 75.56]). Furthermore, while negative events happening to minimal ingroup characters elicited a significant linear increase in corrugator activity ($b = 40.57$, $t(59.99) = 2.26$, $p = 0.03$, 95% CI [4.67, 76.46]), which was modulated by a significant quadratic and marginally significant cubic trend ($p = 0.03$ and 0.07 , respectively, see **Supplementary Section S2f**), the corrugator response to positive events was modeled as a flat line. For minimal outgroup characters, average corrugator activation across the segment was also higher for negative events than for positive events (difference_{neg-pos} $b = 43.94$, $t(430.28) = 3.67$, $p < 0.001$, 95% CI [20.43, 67.45]). Furthermore, while negative events again elicited a significant linear increase in corrugator activity ($b = 12.38$, $t(61.92) = 2.71$, $p = 0.01$, 95% CI [3.25, 21.52]), which was modulated by a significant quadratic trend ($p = 0.02$, see **Supplementary Section S2f**), the corrugator response to positive events was again modeled as a flat line.

We had speculated that minimal group status might also have an impact on how negative vs. positive events affected the corrugator response. Although **Figure 6B** suggests a somewhat stronger EMG-response to negative events befalling minimal ingroup characters than befalling minimal outgroup characters, the statistics do not clearly support this: average corrugator activation over the entire 2.5-s segment did not differ (difference_{ingrp-outgrp} $b = 9.10$, $t(439.04) = 0.76$, $p = 0.45$, 95% CI [-14.53, 32.73]), nor did any of the trends (e.g., linear trend difference_{ingrp-outgrp} $b = 28.18$, $t(67.75) = 1.52$, $p =$

0.13, 95% CI [-8.77, 65.14]). EMG-responses to positive events befalling minimal ingroup vs. minimal outgroup characters were fitted with a flat line whose elevation did not differ either (difference_{ingroup-outgrp} $b = 0.99$, $t(421.35) = 0.08$, $p = 0.93$, 95% CI [-22.40, 24.38]).

DISCUSSION

When processing language, do we *simultaneously* use our emotion systems to simulate somebody else’s described emotion *and* to evaluate, i.e., have our own emotions about, what is described? We explored the viability of a multiple-drivers model for language-driven emotion (’t Hart et al., 2018; ’t Hart et al., 2019; van Berkum et al., in press) by “downtuning” the force of character-dependent emotional evaluation via a minimal-groups paradigm, such that corrugator EMG responses would reveal character-independent emotion simulation to a larger extent. Also, we aimed to replicate the findings of ’t Hart et al. (2018), ’t Hart et al. (2019), generalizing those earlier morality-based observations to a situation where characters were simply *declared* as good or bad, rather than shown to be so earlier in a story. As for morality, we indeed replicated the core result of our earlier studies: substantially more frowning to negative emotion adjectives than to positive ones when the character having the emotion was seen as morally good, but *not* when he or she was seen as morally bad. However, and in contrast to our expectations, defining characters as belonging to a *minimal* (rather than a moral) in- or outgroup did not matter to how much more readers frowned to negative as opposed to positive emotion adjectives. We first discuss the EMG-results per segment, and then turn to a more general discussion.

Processing Character Descriptions

In our study, introducing some unknown fictional character as a member of a minimal in- or outgroup did not elicit any differential frowning. As for *morally defined* groups, however, things were different: declaring some unknown fictional character as “really bad” led to a small but significant phasic increase in frowning, whereas declaring a character as “really good” did not affect the corrugator. It is perhaps tempting to relate this to differences at the level of situation modeling (see **Figure 1**), i.e., of imagining a concrete bad character in some real or imaginary context (with a silhouette providing extra input). However, because isolated negative words are known to elicit more frowning than positive words (e.g., Larsen et al., 2003; Kunkel, 2018; see van Berkum et al., in press, for review), this effect may very well also—or exclusively—hinge on automatic responses associated with the retrieval of negative vs. positive words (“bad vs. “good”). Either way, it is interesting to compare the very modest current effect to the very large corrugator responses to descriptions morally bad and good character *behavior* in our earlier two studies. In ’t Hart et al. (2018), ’t Hart et al. (2019), phasic corrugator increases were some 50–90% higher at peak relative to baseline, when participants read about a main character committing a concrete moral transgression (e.g., deliberately speeding up to soak a pedestrian in the rain) than

when reading about that character displaying morally good behavior (e.g., deliberately slowing down to *not* soak the pedestrian). In the current study, however, seeing a silhouette simply described as really bad generated a phasic corrugator increase which was only some 10% higher at peak relative to baseline, as compared to a silhouette described as really good. Although adequately controlled within-experiment comparisons are required to explore the matter further, this comparative observation could be taken to suggest that describing a concrete bad action in some detail is frowned upon to a much larger extent than simply defining somebody as a bad person, an interpretation that is in line with the idea that our brains evolved to deal with concrete events and actions, and are as such much more sensitive to narrative than to non-narrative descriptions (e.g., Boyd, 2009; Boyd, 2018).

Processing Character Affect

Our predictions for the impact of character morality on reading a subsequent adjective that described *an emotion of that character* were confirmed. With good characters, readers frowned more at negative affective state adjectives like “angry” than at positive affective state adjectives like “happy”, with the difference emerging very rapidly, within only a few hundred milliseconds after adjective presentation. With bad characters, our earlier work had led us to predict that this differential valence effect would be reduced to (close to) zero, which was indeed what we observed in the current study too. Taken together, these EMG-results constitute a direct replication of the ’t Hart et al. (2018), ’t Hart et al. (2019) findings. Like the original findings, the new findings are compatible with a multiple-drivers account in which the valenced emotional responses associated with language-driven simulation and evaluation align for good characters, but counteract each other in the case of bad characters.

Our current morality-based EMG-findings also extend the morality-based ’t Hart et al. (2018), ’t Hart et al. (2019) results from a paradigm where characters were described as actually *doing* something good or bad to a paradigm where characters are simply described as *being* good or bad. Note that the size of the EMG-effect at the critical state adjective (a difference at peak of about 30% relative to baseline) is comparable to the corresponding effect at the critical affective state adjective observed by ’t Hart et al. (2019); a difference at peak of about 20% relative to baseline. Thus, although declaring rather than showing somebody as bad strongly attenuates the differential EMG-response of readers at the character segment, the *downstream impact of this* on how readers respond to various character emotions is not attenuated by that factor at all.

Furthermore, our findings are in line with other EMG-evidence that the social identity of characters can affect later language-driven processing. In an EMG-study on social unexpectedness, for example, descriptions of moral transgressions generated a larger corrugator response if they were committed by characters previously described in a positive, rather than a negative, way (Bartholow et al., 2001). Also, in an EMG-study involving Italian in- or outgroup politicians (e.g., Berlusconi; Fino et al., 2019) the corrugator responded strongly to negative vs. positive emotional

expression descriptions (e.g., “Berlusconi frowns” vs. “Berlusconi smiles”) if the politician belonged to the participant’s political ingroup, but not if he or she belonged to the participant’s political outgroup. The overall pattern of results in the latter study is actually strikingly similar to the pattern in our current and two earlier studies, with average corrugator EMG-responses to outgroup politicians that are not only indifferent to the characters’ emotional state, but that are also positioned between the very different corrugator signals to negative vs. positive emotions of ingroup politicians. This makes sense: political and moral orientations are strongly related (e.g., see van Berkum et al., 2009; Haidt, 2012), and both are associated with strong in- and outgroups. Still, the stability of this crucial finding across labs and materials is reassuring.

In the minimal-group part of the design, the EMG results here were predicted to be an attenuated version of those in the moral-group part of the design, with an intermediately sized adjective valence effect for both in- and outgroup characters, and *some* group-dependent modulation of this effect. However, although the adjective valence effects for minimal in- and outgroups were indeed of an intermediate magnitude, they also were exactly the same. Under a multiple-drivers account, this suggests that when reading, say, “Mark is angry”, readers not only *simulate* negative emotion at the lexical and/or situation-model level similarly for minimal in- and outgroup characters, but also *evaluate* the unhappy event in the same way. This evaluation may or may not be neutral. Importantly, however, it does not differ as a result of whether a type P or type O person is being angry.

A major goal of the current study was to look for new traces of the “power struggle” between language-driven simulation and evaluation, beyond what is visible when working with moral materials. We tried to do so by reducing the impact of *character-dependent* evaluation while keeping the force of lexical and situation-model simulation intact. But this part of the endeavor did not succeed. The reason may well be that the current minimal-groups manipulation is too subtle, and that when applied to fictional people, the resulting group bias is simply too weak to generate any detectable character-dependent evaluation at the critical emotional state adjective. We return to the implication of this after discussing our findings at the third segment.

Processing Reasons for Character Affect

Although the experimental logic hinged on the EMG results at critical adjectives describing the character’s positive or negative emotion, EMG responses to the later verbal *explanation* for that emotion also provided some information. First of all, the explanations for negative character emotion elicited renewed rise-fall phasic corrugator responses in all four character conditions (of at least an additional 30% relative to the signal at 0 msec), whereas the explanations for positive character emotion elicited zero responses in three out of four cases, and only a small (~10%) phasic increase when the positive emotion involves a bad character. Example explanations for negative character emotion involve such phrases as “(because) her

shares turned out to be worthless”, “(because) he stared at her and ignored her”, “(because) somebody pushed her aside to get in more quickly” and “(because the waitress) responds in a grumpy way and looks angrily at her”. The phasic corrugator effects that these reasons for negative character emotion elicit in the reader can thus be explained in many ways, including frowning on moral transgressions, imagining unpleasant states of affairs, or simulating the negative emotions of secondary characters. It is also conceivable that reading about a reason for negative emotion can briefly boost the situation-model simulation of the main character being in that negative state—after all, knowing that somebody’s anger has a reason that fully justifies it, and that you can identify with, may well deepen one’s mental representation of that anger. Because the affect reason segments were not controlled to allow us to discriminate between these various options, these are all issues for future research.

A second and theoretically more interesting finding is that, at the onset of this affective reason segment, the corrugator activation levels by and large echo those at the end of the affective state adjective segment (compare **Figures 5** and **6**). As can be seen in **Figure 3**, the reason is that the corrugator response to descriptions of character emotion are to a large extent maintained throughout the intervening 3 s, during which people read neutral connector phrases such as “. . .when after a few minutes . . .” or “. . .when he arrives at the station and . . .”. In the case of moral in- and outgroup characters, this sustained corrugator behavior replicates what we observed at neutral connector phrases in the t Hart et al. (2019) study. As discussed in our earlier paper, this could be taken to indicate that the emotion simulation induced by phrases like “Mark is angry” is more likely to occur at the level of the situation model (where the character is modeled as angry) than at the—presumably more short-lived—level of simulation as part of retrieving the meaning of the word “angry” from memory. Of course, under the current multiple-drivers account for our morality-based EMG-results at the state adjective, sustained simulation would need to be matched with equally sustained group-dependent evaluation. Also note that the degree of stability over these three intervening seconds is not perfect, which could be taken to indicate dynamic fluctuations in simulation, evaluation, or both. Still, we find it striking that the reader’s emotional state, as indexed by the corrugator, remains relatively stable for several seconds after the critical adjective, not just with moral in- and outgroup characters, but also with minimal ones.

Counteracting Simulation and Evaluation Drivers, or Something Else?

What are the implications for our theoretical model? The first question we should ask is whether the absence of a minimal group effect at the affective state adjective falsifies the multiple-drivers model. We don’t think it does. If emotion simulation and emotional evaluation both drive corrugator EMG, but evaluation is the same for both minimal groups (e.g., people care as much, or as little, about a type-P character’s feelings than about a type-O character’s feelings), no modulation of the adjective valence effect is

to be expected. So, rather than rejecting the multiple-drivers model on these grounds, a more sensible strategy at this point is to look for other techniques that may selectively down- or up-regulate the force of one of the presumed drivers (e.g., using story materials in which “bad” characters commit severe, moderate or mild moral transgressions). Also, our study does replicate the original findings that led us to adopt the multiple-drivers model in the first place, extending the relevant phenomenon to situations where characters are simply declared—rather than shown—to be good or bad. The lack of increased frowning to negative state adjectives like “angry” over positive state adjectives like “happy”, for morally bad characters, can therefore be explained by the same account that we provided for those earlier findings, a tie between lexical and/or situation model simulation pushing corrugator activity up and fairness-based evaluation pushing it down.

As we already pointed out in our earlier publications, a simple account that involves lexical or situation-model *simulation only* cannot explain why the corrugator faithfully tracks the valence of emotion adjectives when the sentence is about a good character, but *not* when it is about a bad character. Also, it is difficult to account for our morality-based results in terms of *evaluation only*. The results in **Figure 5A** might tempt one to infer that readers care about what happens to good, but not *bad*, characters, and that this differential evaluation alone can parsimoniously explain the EMG results. However, this interpretation seems unlikely. Part of the joy of written or streamed fiction comes from caring about what happens to good as well as bad characters. Also, if we would not care about what happens to bad people, gossip would become dysfunctional, and Schadenfreude would not exist. Of course, in a boring lab, things could be different. However, Schadenfreude has also been established in laboratory studies (e.g., Leach and Spears, 2009; Feather and Nairn, 2005; Singer, et al., 2006), and has even been shown to influence corrugator activity (Cikara and Fiske, 2012). More generally, why would the lab context lead people to become indifferent to the fate of bad people, but not good people?

With *simple* simulation-only and evaluation-only accounts dismissed, the multiple-drivers account displayed in **Figure 1** remains an attractive one for our morality-based EMG results, with positive or negative emotional responses associated with language-driven simulation and evaluation aligning for good characters but counteracting each other for bad characters. The explanatory power and flexibility of this multi-factor model is of course also a vulnerability. It is therefore crucial to obtain independent evidence for our assumption that, at least in our materials, simulation and evaluation fully cancel each other out when reading about the emotions of bad people.

Furthermore, although we did not consider them before running the current study, other theoretical explanations for our results may be on the table as well. One possibility is that with immoral characters, readers are somehow less inclined to engage in embodied simulation of what is being described, so less

likely to *simulate* an angry or happy character. This *selective-simulation* idea fits with recent ideas on embodied language processing, where it is becoming clear that language-driven simulation is not an all-or-none concept but depends on all kinds of contextual factors (Willems and Casasanto, 2011; Havas and Matheson, 2013; Zwaan, 2014; Pecher and Zwaan, 2017; Pecher 2018; Winkielman et al., 2018). Identification, or liking, could be one of those factors (Hoeken and Sinkeldam, 2014). As indicated in **Figure 1** and discussed more fully elsewhere (van Berkum et al., in press), we also cannot exclude that *emotional mimicry*, in response to vividly imagined character affect, partly drives emotional facial expressions during language processing. Such mimicry might occur more for good characters than for bad ones either because emotions of the former are simulated to a stronger extent, or because mimicry itself is selective, and more likely to occur with ingroup or otherwise likable characters than with other characters (see Hess and Fischer, 2014, for a review of relevant findings, and Fino et al., 2019, for EMG-results interpreted in terms of language-driven mimicry).

The possibility of selective simulation and/or selective mimicry illustrates the fact that we are dealing with a very complex situation here. Although we currently prefer our multiple-drivers account over post-hoc accounts in terms of selective simulation and/or selective mimicry—if only because it was conceived of *before the experiment*—we acknowledge that our studies are only scratching the surface. Language can lead to emotion in many different ways, and disentangling them will remain a challenge for some time.

LIMITATIONS

We end with some limitations of the current study. First, the multiple-drivers account illustrated in **Figure 1** inevitably introduces several free parameters in our modeling of language-driven facial EMG data. Of course, only *some* of the depicted drivers may actually be at work (i.e., explain EMG variance) at any given time. Furthermore, like so many other workings of the human brain, language-driven emotion may simply be this complex. Nevertheless, the explanatory power of the current model is also a vulnerability, which will need to be addressed in future work. Second, most of our participants were female, with only 6 males in a group of 64 participants. With small empathy-related sex differences in corrugator EMG reported elsewhere van der Graaff et al. (2016), this may matter. Third, we did not use a deliberate strategy to prevent people from guessing that their facial expressions were the object of study (e.g., attach dummy electrodes elsewhere, cf. Fridlund and Cacioppo, 1986). Although corrugator activity is in part automatic (e.g., Cacioppo et al., 1992; Dimberg et al., 2000; Neumann et al., 2005; Tamietto and de Gelder, 2010; van Boxtel, in press), and although it is plausible that participants soon forgot about the electrodes (see Nordin, 1990, for evidence of rapid facial sensory habituation), it may be wise to consider such a strategy in future work. Fourth, in our

growth curve analysis, only linear, quadratic and cubic trends are fitted, and they were constrained to fit the signals in a segment of a predefined duration. Although this worked out reasonably well in our data, the segment constraint obviously imposes limitations on how the data can be modeled—our procedure would not work well, for instance, when most of the segment contained a flat line, with a huge effect in the narrow last bit of the signal only. Fifth, we assessed emotion in terms of valence only—this simplified the research logic, but it also ignores some of the richness of language-induced emotion. Finally, we made relatively simple working assumptions about how characters are perceived (e.g., as good or bad), and about how people evaluate, say, something bad happening to a bad character. We think that given our materials, those assumptions are reasonable. However, people are layered, and so is their response to other people's fate. The study of language-driven human emotion will sooner or later need to take on this additional complexity.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Data publication platform of Utrecht University <https://doi.org/10.24416/UU01-YM9VPP>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Linguistics Chamber of the Faculty of

REFERENCES

- Anderson, M. L. (2010). Neural Reuse: A Fundamental Organizational Principle of the Brain. *Behav. Brain Sci.* 33, 245–266. doi:10.1017/s0140525x10000853
- Barsalou, L. W. (2008). Grounded Cognition. *Annu. Rev. Psychol.* 59, 617–645. doi:10.1146/annurev.psych.59.103006.093639
- Barsalou, L. W. (2009). Simulation, Situated Conceptualization, and Prediction. *Phil. Trans. R. Soc. B* 364 (1521), 1281–1289. doi:10.1098/rstb.2008.0319
- Bartholow, B. D., Fabiani, M., Gratton, G., and Bettencourt, B. A. (2001). A Psychophysiological Examination of Cognitive Processing of and Affective Responses to Social Expectancy Violations. *Psychol. Sci.* 12 (3), 197–204. doi:10.1111/1467-9280.00336
- Boyd, B. (2009). *On the Origin of Stories: Evolution, Cognition, and Fiction*. Harvard University Press. doi:10.2307/j.ctvj9xvk
- Boyd, B. (2018). The Evolution of Stories: from Mimesis to Language, from Fact to Fiction. *Wires Cogn. Sci.* 9 (1), e1444. doi:10.1002/wcs.1444
- Cacioppo, J. T., Bush, L. K., and Tassinari, L. G. (1992). Microexpressive Facial Actions as a Function of Affective Stimuli: Replication and Extension. *Pers Soc. Psychol. Bull.* 18 (5), 515–526. doi:10.1177/0146167292185001
- Cikara, M., and Fiske, S. T. (2012). Stereotypes and Schadenfreude. *Soc. Psychol. Personal. Sci.* 3 (1), 63–71. doi:10.1177/1948550611409245
- Diehl, M. (1990). The Minimal Group Paradigm: Theoretical Explanations and Empirical Findings. *Eur. Rev. Soc. Psychol.* 1 (1), 263–292. doi:10.1080/14792779108401864
- Humanities Ethics assessment Committee at Utrecht University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

B'tH, MS, and JvB designed the study, AvB provided specific EMG expertise for study design and data analysis, B'tH conducted the study, B'tH and MS analyzed the results, and JvB, MS, and B'tH wrote the paper.

FUNDING

Partly supported by NWO Vici grant #277–89–001 to JvB.

ACKNOWLEDGMENTS

We thank Ella Bosch and Eletta Damen for help with running the experiment, Huub van den Bergh for his input during the analysis, and members of the UiL OTS Language and Communication research group for feedback. Results were initially published in a doctoral dissertation (t Hart, 2017). Correspondence: JvB j.vanberkum@uu.nl.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2021.590077/full#supplementary-material>

- Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious Facial Reactions to Emotional Facial Expressions. *Psychol. Sci.* 11 (1), 86–89. doi:10.1111/1467-9280.00221
- Ekman, P., Hager, J. C., and Friesen, W. V. (1981). The Symmetry of Emotional and Deliberate Facial Actions. *Psychophysiol.* 18, 101–106. doi:10.1111/j.1469-8986.1981.tb02919.x
- Feather, N., and Nairn, K. (2005). Resentment, Envy, Schadenfreude, and Sympathy: Effects of Own and Other's Deserved or Undeserved Status. *Aust. J. Psychol.* 57 (2), 87–102. doi:10.1080/00049530500048672
- Fingerhut, J., and Prinz, J. J. (2018). Grounding Evaluative Concepts. *Phil. Trans. R. Soc. B* 373 (1752), 20170142. doi:10.1098/rstb.2017.0142
- Fino, E., Menegatti, M., Avenanti, A., and Rubini, M. (2016). Enjoying vs. Smiling: Facial Muscular Activation in Response to Emotional Language. *Biol. Psychol.* 118, 126–135. doi:10.1016/j.biopsycho.2016.04.069
- Fino, E., Menegatti, M., Avenanti, A., and Rubini, M. (2019). Unfolding Political Attitudes through the Face: Facial Expressions when Reading Emotion Language of Left-And Right-Wing Political Leaders. *Scientific Rep.* 9 (1), 1–10. doi:10.1038/s41598-019-51858-7
- Froni, F., and Semin, G. R. (2013). Comprehension of Action Negation Involves Inhibitory Simulation. *Front. Hum. Neurosci.* 7, 1–7. doi:10.3389/fnhum.2013.00209
- Froni, F., and Semin, G. R. (2009). Language that Puts You in Touch with Your Bodily Feelings. *Psychol. Sci.* 20 (8), 974–980. doi:10.1111/j.1467-9280.2009.02400.x
- Fridlund, A. J., and Cacioppo, J. T. (1986). Guidelines for Human Electromyographic Research. *Psychophysiol.* 23 (5), 567–589. doi:10.1111/j.1469-8986.1986.tb00676.x

- Frijda, N. H. (2007). *The Laws of Emotion*. Mahwah, NJ: Erlbaum.
- Glenberg, A. M., and Gallese, V. (2012). Action-based Language: A Theory of Language Acquisition, Comprehension, and Production. *cortex* 48 (7), 905–922. doi:10.1016/j.cortex.2011.04.010
- Glenberg, A. M. (2011). How Reading Comprehension Is Embodied and Why that Matters. *Int. Electron. J. Elem. Educ.* 4 (1), 5–18.
- Glenberg, A. M., Webster, B. J., Mouilso, E., Havas, D., and Lindeman, L. M. (2009). Gender, Emotion, and the Embodiment of Language Comprehension. *Emot. Rev.* 1, 151–161. doi:10.1177/1754073908100440
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., and Ditto, P. H. (2011). Mapping the Moral Domain. *J. Personal. Soc. Psychol.* 101 (2), 366–385. doi:10.1037/a0021847
- Greene, J. D. (2014). “The Cognitive Neuroscience of Moral Judgment and Decision-Making,” in *The Cognitive Neuroscience V*. Editor M. S. Gazzaniga (Cambridge, Massachusetts, USA: MIT Press), 1013–1024.
- Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. London, United Kingdom: Penguin Books.
- Havas, D. A., and Matheson, J. (2013). The Functional Role of the Periphery in Emotional Language Comprehension. *Front. Psychol.* 4, 294. doi:10.3389/fpsyg.2013.00294
- Hess, U., and Fischer, A. (2014). Emotional Mimicry: Why and when We Mimic Emotions. *Social Personal. Psychol. Compass* 8, 45–57. doi:10.1111/spc3.12083
- Hoeken, H., and Sinkeldam, J. (2014). The Role of Identification and Perception of Just Outcome in Evoking Emotions in Narrative Persuasion. *J. Commun.* 64, 935–955. doi:10.1111/jcom.12114
- Höfling, T. T. A., Gerdes, A., Föhl, U., and Alpers, G. W. (2020). Read My Face: Automatic Facial Coding versus Psychophysiological Indicators of Emotional Valence and Arousal. *Front. Psychol.* 11, 1388. doi:10.3389/fpsyg.2020.01388
- Künecke, J., Sommer, W., Schacht, A., and Palazova, M. (2015). Embodied Simulation of Emotional Valence: Facial Muscle Responses to Abstract and Concrete Words. *Psychophysiol* 52 (12), 1590–1598. doi:10.1111/psyp.12555
- Kunke, A. E. (2018). *The Processing Of Moral Transgressions: Investigating The Role Of Affective Evaluations*. Unpublished doctoral dissertation. Tübingen, Germany: Eberhard Karls Universität.
- Larsen, J. T., Norris, C. J., and Cacioppo, J. T. (2003). Effects of Positive and Negative Affect on Electromyographic Activity over Zygomaticus Major and Corrugator Supercilii. *Psychophysiol* 40 (5), 776–785. doi:10.1111/1469-8986.00078
- Lazarus, R. S. (1991). *Emotion and Adaptation*. Oxford: Oxford University Press.
- Leach, C. W., and Spears, R. (2009). Dejection at In-Group Defeat and Schadenfreude toward Second- and Third-Party Out-Groups. *Emotion* 9 (5), 659–665. doi:10.1037/a0016815
- Lee, S., and Potter, R. F. (2018). The Impact of Emotional Words on Listeners’ Emotional and Cognitive Responses in the Context of Advertisements. *Commun. Res.* 47 (8), 1155–1180.
- Mirman, D. (2015). Growth Curve Analysis. Philadelphia, PA: Drexel University. Available at: www.danmirman.org (Accessed March 24, 2015).
- Morrison, S., Decety, J., and Molenberghs, P. (2012). The Neuroscience of Group Membership. *Neuropsychologia* 50, 2114–2120. doi:10.1016/j.neuropsychologia.2012.05.014
- Neumann, R., Hess, M., Schulz, S., and Alpers, G. (2005). Automatic Behavioural Responses to Valence: Evidence that Facial Action Is Facilitated by Evaluative Processing. *Cogn. Emot.* 19 (4), 499–513. doi:10.1080/02699930441000364.001
- Nordin, M. (1990). Low-threshold Mechanoreceptive and Nociceptive Units with Unmyelinated (C) Fibres in the Human Supraorbital Nerve. *J. Physiol.* 426, 229–240. doi:10.1113/jphysiol.1990.sp018135
- Pecher, D. (2018). Curb Your Embodiment. *Top. Cogn. Sci.* 10 (3), 501–517. doi:10.1111/tops.12311
- Pecher, D., and Zwaan, R. A. (2017). Flexible Concepts: a Commentary on Kemmerer (2016). *Lang. Cogn. Neurosci.* 32 (4), 444–446. doi:10.1080/23273798.2016.1274413
- Peck, R., and Devore, J. (2008). *Statistics: The Exploration and Analysis of Data*. 6th ed.. London, United Kingdom: Thomson Higher Learning.
- Pulvermüller, F., Shtyrov, Y., and Ilmoniemi, R. (2005). Brain Signatures of Meaning Access in Action Word Recognition. *J. Cogn. Neurosci.* 17 (6), 884–892. doi:10.1162/0899892054021111
- Quené, H., and van den Bergh, H. (2008). Examples of Mixed-Effects Modeling with Crossed Random Effects and with Binomial Data. *J. Mem. Lang.* 59, 413–425. doi:10.1016/j.jml.2008.02.002
- Quené, H., and van den Bergh, H. (2004). On Multi-Level Modeling of Data from Repeated Measures Designs: a Tutorial. *Speech Commun.* 43, 103–121. doi:10.1016/j.specom.2004.02.004
- Scarantino, A. (2014). The Motivational Theory of Emotions. *Moral Psychol. Hum. Agency*, 156–185. doi:10.1093/acprof:oso/9780198717812.003.0008
- Scherer, K. R., and Moors, A. (2019). The Emotion Process: Event Appraisal and Component Differentiation. *Annu. Rev. Psychol.* 70, 719–745. doi:10.1146/annurev-psych-122216-011854
- Singer, T., Seymour, B., O’Doherty, J. P., Stephan, K. E., Dolan, R. J., and Frith, C. D. (2006). Empathic Neural Responses Are Modulated by the Perceived Fairness of Others. *Nature* 439, 466–469. doi:10.1038/nature04271
- t Hart, B. (2017). *That’s Frowned upon: Using Facial EMG to Track Evaluation and Simulation during Affective Language Processing*. Utrecht: LOT Dissertation Series.
- t Hart, B., Struiksma, M. E., Van Boxtel, A., and van Berkum, J. J. A. (2018). Emotion in Stories: Facial EMG Evidence for Both Mental Simulation and Moral Evaluation. *Front. Psychol.* 9, 613. doi:10.3389/fpsyg.2018.00613
- t Hart, B., Struiksma, M. E., Van Boxtel, A., and van Berkum, J. J. A. (2019). Tracking Affective Language Comprehension: Simulating and Evaluating Character Affect in Morally Loaded Narratives. *Front. Psychol.* 10, 318. doi:10.3389/fpsyg.2019.00318
- Tajfel, H., Billig, M. G., Bundy, R. P., and Flament, C. (1971). Social Categorization and Intergroup Behaviour. *Eur. J. Soc. Psychol.* 1, 149–178. doi:10.1002/ejsp.2420010202
- Tamietto, M., and De Gelder, B. (2010). Neural Bases of the Non-conscious Perception of Emotional Signals. *Nat. Rev. Neurosci.* 11 (10), 697–709. doi:10.1038/nrn2889
- Tappin, B. M., and McKay, R. T. (2016). The Illusion of Moral Superiority. *Soc. Psychol. Personal. Sci.* 8, 1–9. doi:10.1177/1948550616673878
- Tooby, J., and Cosmides, L. (2008). “The Evolutionary Psychology of the Emotions and Their Relationship to Internal Regulatory Variables,” in *Handbook of Emotions*. Editors M. Lewis, J. M. Haviland-Jones, and L. F. Barrett (New York: Guilford), 114–137.
- van Berkum, J. J. A. (2018). “Language Comprehension, Emotion, and Sociality: Aren’t We Missing Something?,” in *Oxford Handbook of Psycholinguistics*. Editors S. A. Rueschemeyer and G. Gaskell (Oxford: Oxford University Press), 644–669.
- van Berkum, J. J. A. (in press). “A Survey of Emotion Theories and Their Relevance to Language Research,”. To Appear in *Language and Emotion: An International Handbook*. Editors G. L. Schiewer, J. Altarriba, and B. C. Ng (Berlin: Germany: De Gruyter Mouton).
- van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., and Murre, J. (2009). Right or Wrong?. *Psychol. Sci.* 20, 1092–1099. doi:10.1111/j.1467-9280.2009.02411.x
- van Berkum, J. J. A., Struiksma, M. E., and t Hart, B. (in press). “Using Facial EMG to Track Emotion during Language Comprehension: Past, Present, and Future,” in *Language Electrified: Techniques, Methods, Applications, and Future Perspectives in the Neurophysiological Investigation of Language*. Editors M. Grimaldi, Y. Shtyrov, and E. Brattico (Springer).
- van Boxtel, A. (in press). “Electromyographic (EMG) Responses of Facial Muscles during Language Processing,” in *Mirko Grimaldi, Yury Shtyrov & Elvira Brattico Language Electrified: Techniques, Methods, Applications, and Future Perspectives in the Neurophysiological Investigation of Language* (Springer).
- van Boxtel, A. (2010). “Facial EMG as a Tool for Inferring Affective States,” in *Proceedings of Measuring Behavior 2010, Eindhoven, Netherlands, August 24–27, 2010*. Editors A. J. Spink, F. Grieco, O. E. Krips, L. W. S. Loijens, L. P. J. J. Noldus, and P. H. Zimmerman.
- van der Graaff, J., Meeus, W., de Wied, M., van Boxtel, A., van Lier, P. A. C., Koot, H. M., et al. (2016). Motor, Affective and Cognitive Empathy in Adolescence: Interrelations between Facial Electromyography and Self-Reported Trait and State Measures. *Cogn. Emot.* 30 (4), 745–761. doi:10.1080/02699931.2015.1027665
- Vigliocco, G., Meteyard, L., Andrews, M., and Kousta, S. (2009). Toward a Theory of Semantic Representation. *Lang. Cogn.* 1, 219–247. doi:10.1515/langcog.2009.011
- Vossen, H. G. M., Piotrowski, J. T., and Valkenburg, P. M. (2015). Development of the Adolescent Measure of Empathy and Sympathy (AMES). *Personal. Individual Differences*, 74, 66–71. doi:10.1016/j.paid.2014.09.040

- Willems, R. M., and Casasanto, D. (2011). Flexibility in Embodied Language Understanding. *Front. Psychol.* 2. doi:10.3389/fpsyg.2011.00116
- Winkielman, P., Coulson, S., and Niedenthal, P. (2018). Dynamic Grounding of Emotion Concepts. *Phil. Trans. R. Soc. B* 373 (1752), 20170127. doi:10.1098/rstb.2017.0127
- Winkielman, P., Niedenthal, P., Wielgosz, J., Eelen, J., and Kavanagh, L. C. (2015). Embodiment of Cognition and Emotion. In *APA Handbook of Personality and Social Psychology, Vol. 1. Attitudes and social cognition*. Editors M. Mikulincer, P. R. Shaver, E. Borgida, and J. A. Bargh. Washington, DC: American Psychological Association. 151–175. doi:10.1037/14341-004
- Winter, B. (2020). *Statistics for Linguists: An Introduction Using R*. New York: Routledge.
- Zwaan, R. A. (2014). Embodiment and Language Comprehension: Reframing the Discussion. *Trends Cogn. Sci.* 18, 229–234. doi:10.1016/j.tics.2014.02.008
- Zwaan, R. A., and Pecher, D. (2012). Revisiting Mental Simulation in Language Comprehension: Six Replication Attempts. *PLoS One* 7, 51382. doi:10.1371/journal.pone.0051382
- Zwaan, R. A. (2016). Situation Models, Mental Simulations, and Abstract Concepts in Discourse Comprehension. *Psychon. Bull. Rev.* 23 (4), 1028–1034. doi:10.3758/s13423-015-0864-x
- Zwaan, R. A., Stanfield, R. A., and Yaxley, R. H. (2002). Language Comprehenders Mentally Represent the Shapes of Objects. *Psychol. Sci.* 13, 168–171. doi:10.1111/1467-9280.00430

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 't Hart, Struiksmā, van Boxtel and van Berkum. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.