Affective Language Processing: Language-driven Evaluation of Character Affect in Morally Loaded Narratives

Li Kloostra

6986471

RMA Thesis Linguistics

Utrecht University 2022

Words: 10.071

(excl. abstract, acknowledgements, and reference list)

Supervisors:

Prof. Dr. J.J.A. (Jos) van Berkum

Dr. M. (Marijn) Struiksma

Utrecht Institute of Linguistics (UiL) OTS/Institute for Language Sciences, Utrecht University, Utrecht, Netherlands

29 November 2022

Abstract

Research has shown that the valence of emotional linguistic stimuli is reflected in corrugator supercilii ('frowning muscle') activation, which can be measured using facial electromyography. But in a larger context, the corrugator response may be additionally affected by emotional evaluation, i.e., how we feel about some emotion. Our earlier study from 2019 found that descriptions of moral characters experiencing negative emotions elicited substantially more corrugator activity than positive emotions, thus reflecting valence, but they found no effect of valence for adjectives describing immoral characters. The authors therefore suggested that both mental simulation of word valence and moral evaluation may drive corrugator activation, e.g., mentally simulating 'angry' activates the corrugator whereas 'happy' relaxes it, whereas evaluation may elicit relaxation for both moral-positive and immoral-negative stories based on the idea that good people deserve good things and bad people deserve bad things. More research was needed to corroborate the existence of these two effects opposing each other. This current study therefore used the same narratives but with an additional task that requested participants to explicitly judge the behavior with a rating task. This was hypothesized to increase their tendency to evaluate what they read and subsequently change the net result that the postulated two effects should have. Critically, we expected more frowning for bad people experiencing positive emotions than negative emotions since bad people feeling good should be seen as very unfair. However, this is not what we found, and our study in fact replicated the findings from 2019: a null result of valence for adjectives describing immoral characters and a large valence-based effect for adjectives describing moral characters. Surprisingly, we did find the expected evaluation-based pattern in responses on an exploratory task where we asked participants to rate the fairness of every narratives' ending. Finally, we found preliminary evidence that a person's attitude towards justice, i.e., their preference or desire for morally good people being rewarded and morally bad people being punished, as a personality trait may modulate the effect that morality and valence have on their corrugator responses.

Keywords: affect, embodiment, emotion, facial electromyography, morality, psycholinguistics

Introduction

Humans can completely immerse themselves in a good book. A heartwarming situation where two long lost lovers finally meet each other again can make our hearts flutter, whereas a frustrating situation such as someone frantically trying to find their child in a busy park can quickly make us feel stressed. Theories of grounded cognition postulate that, in order to process and comprehend language, people mentally simulate the situations or concepts that it refers to (Barsalou, 2008; Hinojosa, Moreno, & Ferré, 2020; Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011; Zwaan, 2016) and studies on embodied language processing specifically emphasize the role of sensorimotor simulation during which neural traces to concrete experiences, such as the feeling of hugging or joy, are reactivated (Majid, 2012; Niedenthal, Winkielman, Mondillon, & Vermeulen, 2009; Winkielman, Niedenthal, Wielgosz, Eelen, & Kavanagh, 2015).

Research has shown that reading action words like 'lick', 'pick' and 'kick' activates certain regions of the (pre)motor cortex involved in performing these actions (Hauk, Johnsrude, & Pulvermüller, 2004). While the exact nature of mental simulation and its role in processes of (language) comprehension and cognition are still subject to debate, there seems to be a general consensus that there is a link between mental simulation and physiological responses. In that same vein, processing affective language, i.e., language referring to emotions or other emotionally relevant concepts, might also involve sensorimotor simulation and consequently elicit physiological responses associated with the emotion in question.

Numerous studies have used facial electromyography (fEMG) to measure facial muscle activation and indeed found muscle responses were congruent with the negative or positive valence of linguistic input: increased activity of the *corrugator supercilii* or 'frowning muscle' for negative emotions (e.g., 'angry') and increased activity of the *zygomaticus major,* responsible for lifting the corners of the mouth for smiling, for positive emotions (e.g., 'happy') (Foroni & Semin, 2009; Larsen, Norris, & Cacioppo, 2003). This effect has been found both for isolated affective words as well as for phrases and sentences containing affectively relevant information (Fino, Menegatti, Avenanti, & Rubini, 2016; Niedenthal et al., 2009). With phrases or sentences, simulation may occur on two different levels: the lexical level associated with retrieval of individual words and their meanings (e.g., 'woman', 'is', 'happy'), and the situational level associated with a more holistic representation of a phrase or sentence ('a happy woman'). For instance, your frowning muscles likely relax a little when reading about a loving embrace, a positive event, whereas they activate when reading about a child that went missing, a negative event.

But reading emotionally salient language within a given context rather than isolated words or phrases may also elicit some emotional evaluation, e.g., how do we feel about the emotional event. Reading about someone happily accepting a bonus check at work may make us feel happy too, but this happiness may not manifest if we previously read how this character casually continued walking upon seeing an elderly man trip on the sidewalk. The positive event of getting a bonus could be evaluated as negative as we may find it unfair or undeserved in the context, which can evoke a sense of indignation and subsequentially elicit frowning. Similarly, reading about this character discovering a dent in their parked car may elicit a sense of happiness or satisfaction (because the bad character had it coming), or even some *Schadenfreude*, i.e., joy resulting from someone else's misfortune (Ouwerkerk, Van Dijk, Vonkeman, & Spears, 2018; Smith & van Dijk, 2018). The effects of emotion evaluation also tend to be reflected in our facial expressions (Ekman & Oster, 1979; Ekman, 1993), and evaluation-based valence may modulate the corrugator in a similar way as simulation: an event that is evaluated as negative may elicit more activation than a positive event.

Some of our earlier studies have already explored what happens when we read about emotions in more complex contexts by using short morally loaded narratives, but it remains unclear what happens exactly when the effects from simulation and evaluation result in opposing valence (i.e., evaluating a positive event as negative or a negative event as positive) ('t Hart, Struiksma, van Boxtel, & van Berkum, Jos J. A., 2019; 't Hart, Struiksma, van Boxtel, & van Berkum, Jos J. A., 2018; 't Hart, Struiksma, van Boxtel, & van Berkum, Jos J. A., 2021). The following section will elaborate on the findings of these studies to clarify the issue of the opposing effects of simulation and evaluation before introducing how this current study aims to follow up the existing research by zooming in on the fairness-based evaluation effect.

Our prior studies

The combined effect of emotion simulation and evaluation of character affect has been explored in three prior fEMG studies ('t Hart et al., 2019; 't Hart et al., 2018; 't Hart et al., 2021). For the first two studies, participants were presented with short morally loaded narratives where we embedded phrases containing a character's moral or immoral behavior (e.g., Mark is helping someone out vs. calling someone names) followed by a phrase with a positive or negative adjective describing their affective state (e.g., 'Mark is happy' vs. 'Mark is angry'). The manipulation of a character's morality allowed us to promote both simulation and evaluation of the characters' affective state later on and to compare the fEMG response to positive and negative adjectives when preceded by moral or immoral behavior.

While reading, fEMG was recorded over the corrugator muscle since it has a robust relationship with emotional valence and its signal is bidirectional, therefore allowing for differential corrugator patterns for words or phrases with negative valence (increased activity, frowning) and positive valence (decreased activity, relaxation) (Larsen et al., 2003; Tan et al., 2012; van Berkum, Struiksma, & 't Hart, 2020). Interestingly, reading the adjective 'angry' in a phrase like 'Mark is angry' only lead to more corrugator activation than 'Mark is happy' when this Mark had been introduced as a moral character, but not when introduced as an immoral character. In fact, there was no difference in activation at all between positive and negative adjectives describing immoral characters' affective states.

't Hart et al. (2019) noted that this pattern of the corrugator response to the adjectives could not be explained by language-driven simulation only, which they dubbed the 'simulation-only' account: mentally simulating the lexical meaning of 'happy' or 'frustrated' or an imagined situation in which someone is happy or frustrated' would result in a differential corrugator pattern that reflects higher corrugator activation for negative than positive adjectives, but no influence from evaluation based on moral status. Likewise, an 'evaluation-only' account would predict a differential corrugator pattern strictly based on evaluation guided by a characters moral status, e.g., it is fair when moral characters are 'happy' and unfair when they are 'frustrated', whereas the opposite is true for immoral characters. This account would therefore predict that the corrugator pattern for immoral characters is 'flipped' in comparison to the pattern for moral characters, i.e., more corrugator activation in response to positively valenced adjectives and less in response to negatively valenced adjectives, but this is also not in line with the findings. They therefore proposed a third account that postulates that there are multiple forces that drive facial muscle activation, as readers may both simulate the character's affective state and evaluate the occurrence in the moral context of the narrative. This 'multiple-drivers' account was based on the facial Affective Language Comprehension model by Van Berkum et al. (2020). See Figure 1 for a simplified schematic overview of how reading affective language is postulated to invoke both simulation and evaluation processes which may in turn modulate facial muscle activation ('t Hart et al., 2021). Also note that the overview includes both the situational and the lexical level of simulation.

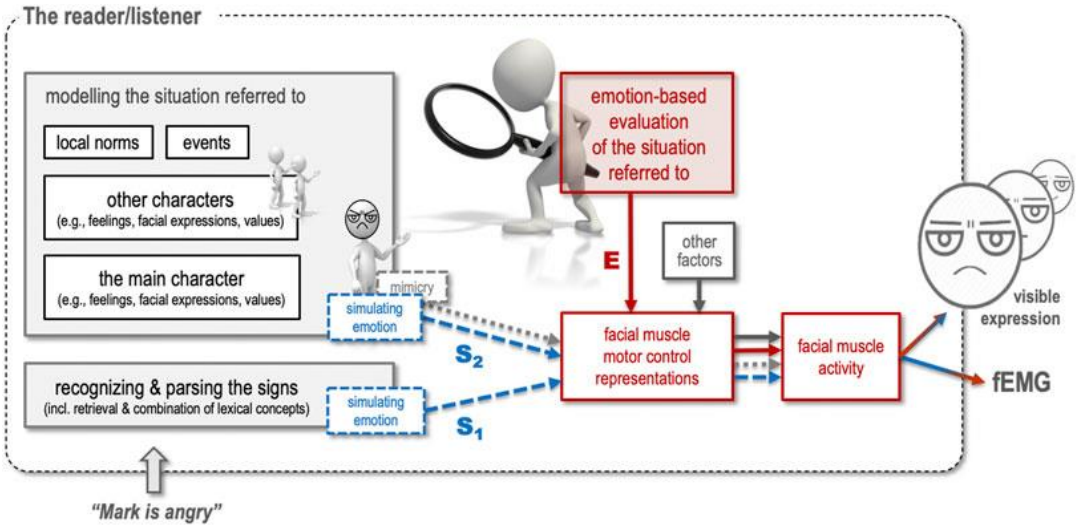


Figure 1 | The multiple-drivers account, adapted from the broader fALC model (see *(van Berkum et al., 2020)*), shows how facial muscle activity may be affected by both simulation and evaluation of affective language during processing. Additionally, the model recognizes that there may be other factors, including mimicry, that drive our facial expressions which will be further reviewed in the discussion.

The findings from 't Hart et al. (2019) could be explained under this third account: if simulation and evaluation act as counteracting forces in the case of immoral characters, the net result is no valence effect at all. 'Bad Mark – angry' would cause increased corrugator activity from simulating 'angry' but decreased activity from evaluating it as a positive thing as it may be viewed as a fair or deserved punishment. Likewise, reading that 'bad Mark' is 'happy' would cause decreased corrugator activity from simulating 'happy' but decreased activity from evaluating his happiness as negative since a bad character does not deserve it. The potentially opposing effect moral evaluation may in such cases reflect a general desire for immoral behavior to be punished and not rewarded, and perhaps even some *Schadenfreude* in the case of 'bad Mark - angry' and *Glückschmerz* (pain from another's fortune) in the case of 'bad Mark – happy' (Ouwerkerk et al., 2018).

Alternatively, the null effect of valence for immoral characters could also reflect that readers do not care about what happens to them or do not (want to) identify with them to the extent that they are less inclined to simulate or evaluate the affective states of the immoral characters (Hoeken & Sinkeldam, 2014). But, like the authors noted, it seems quite unlikely that readers are completely unaffected by adjectives in immoral conditions because they do not care about or dislike immoral characters ('t Hart et al., 2019). Phenomena such as gossiping and rumors are evidence that, at least in real life, we often do have opinions about those we do not like. Furthermore, they noted that the multiple-drivers account does not rule out other factors that may affect language-induced facial muscle activity, nor does it make presumptions regarding the magnitude of either the simulation or evaluation effect or whether the two processes occur simultaneously or consecutively.

With these findings in mind, our third and latest study already looked whether a weaker identification or association with a character would in turn evoke a weaker effect of evaluation by comparing morality-based in- and out-groups with personality-based minimal in- and outgroups ('t Hart et al., 2021). In short, telling participants before each trial that a character was either good or bad for the morality-based groups was expected to have the same effect that the descriptions of (im)moral behavior had in the prior two studies, and two newly created minimal groups were based on a fake personality test that participants filled in beforehand because personality type was expected to invoke a weaker in-/out-group effect than morality. The hypothesis of this study was that readers show a weaker evaluation effect for characters in minimal groups, which under the multiple-drivers account would change the net force of simulation and evaluation on the corrugator. Specifically for adjectives describing out-group characters, in which case forces of emotional evaluation and simulation are postulated to counteract, the results should now show a small effect of simulation following the adjective's valence. As expected, moral in- and outgroups replicated the corrugator pattern found before, however, the minimal in- and out-groups appeared to have no effect at all: there were no differences in corrugator responses to positive and negative adjectives describing regardless of personality type.

While this latest study from 2021 looked at reducing the evaluation force to see a larger simulation effect, our current study aimed to do the opposite by boosting evaluation.

The Current Study

The main goal of the current study was to increase the evaluation force in order to manipulate the potential interplay between the simulation and evaluation forces postulated by the multiple-drivers account. If the findings of our 2019 study ('t Hart et al., 2019) were indeed a result of simulation and evaluation forces outweighing each other in the case of immoral characters, rather than reflecting carelessness or an unwillingness to identify with such characters, then evaluation may also be able to override simulation as an influence on corrugator activity. In order to boost the postulated effect of evaluation, we inserted a rating task in the same composite short narratives used in 2019. Table 1 below shows how in these narratives, phrases such as "Mark is angry" or "Mark is happy" were embedded in a context where they always followed a segment in which the character displayed moral or immoral behavior and consecutively a rating task inquiring participants to rate this behavior on a 7-point scale from 'very bad' to 'very good'.

We hypothesized that requesting participants to actively reflect on and rate the moral or immoral behavior would boost evaluation-driven corrugator activity due to emotional evaluation of the affective state adjective in the context of a character's moral status.

Since the first two segments were identical to those in the first two prior studies, we expected to replicate the pattern of corrugator activity for the character manipulation segment, i.e., a large decrease in corrugator activation when reading about moral behavior and a large increase in corrugator activation when reading about immoral behavior. Finding a clear differential corrugator pattern also verifies that the moral manipulation has affected the readers, which crucially modulates evaluation of the affective state adjective. Frowning activity in response to the state adjective is our main focus since the clearly valenced single-word segment allows for very precise time-locking and clean fEMG measurements.

In contrast with our previous results, we now hypothesized to find a differential corrugator pattern in the case of immoral characters that is evaluation-based. Following our multiple-drivers model, fairness-based evaluation of an immoral characters' affective state should have the inverse effect that valence-based simulation would have, e.g., reading about positive affect should in this case lead to more frowning than reading about negative affect as the first should be evaluated as unfair or undesirable and the latter as fair and desirable. Despite our expectation that the effect of evaluation would be larger than the supposed opposing effect of simulation in modulating corrugator activity, we did hypothesize that simulation may still occur and attenuate the differential response. The predicted net result for immoral characters is therefore a differential corrugator response that reflects an attenuated effect of evaluation-based valence.

For the moral characters, we expected to replicate the effect of simulation and evaluation joining forces again or potentially find a larger joint effect, e.g., the effect of simulation may be slightly smaller as our emotion system may have to compensate for any extra focus that is now required by the emphasized process of evaluation, or simulation may remain unaffected and be joined by the boosted evaluation.

The last segment is a subclause describing the reason for the affective state. This segment previously appeared to elicit renewed corrugator activity similar to that at the state adjective ('t Hart et al., 2019; 't Hart et al., 2021). Based on our previous findings, we again expected the corrugator pattern to be similar as the expected pattern for the state adjectives, i.e., a substantial differential pattern for moral characters and an attenuated evaluation-based valence effect for immoral characters. There is, however, not an obvious element in these clauses that carry valence, so time-locking it to the corrugator response is less precise and we should be cautious in our interpretation of these signals. This segment is therefore not a main focus in the current study.

In sum, our main expectations are a substantial differential corrugator response that is valence-based in the case of moral characters, i.e., more frowning for negative adjectives, and an attenuated and reverse corrugator response in the case of immoral characters, i.e., slightly more

frowning for positive adjectives. Due to the magnitude of the hypothesized joint and opposing forces, the differential corrugator pattern for immoral characters is expected to stay closer to a null-result and should lie in between the differential pattern for moral characters.

Table 1 | This table illustrates the structure and timing of a stimulus and the design of the stimuli is further elaborated below. For continuity, the affective state adjective and the affect reason always appeared in the same valence condition.

Baseline	*Neutral distractor image of a forest scene (always the same)*			2 s
Introduction	Mark is driving through the pouring rain, on his way to his mother. He's still in the inner city and big puddles have formed. It's been raining non-stop since yesterday. Some streets are practically flooded. There are few cars on the road and fewer bicycles and pedestrians still. Mark is headed for a giant puddle and spots a pedestrian on the sidewalk.			18 s
Character morality (moral/immoral)	Mark slows down to avoid the puddle, making sure he doesn't splash the pedestrian.	OR	Mark accelerates through the puddle on purpose to create a big splash and soak the pedestrian.	5 s
Rating task	**I find this behavior:** **very bad** **very good** **1 2 3 4 5 6 7**			5,25 s
Continuation	Once outside the city he is driving along on the freeway. There still isn't a lot of traffic and Mark is enjoying the landscape and the drive. He's got the radio on full blast and sings along loudly. When he glances at the dashboard to adjust the channel he spots a warning light. He forgot to put petrol in the car and has been running on empty for a while.			15 s
Transition	. . .			1 s
Name	Mark			0.75 s
Verb	is			0.75 s
Affective state adjective	happy	OR	frustrated	1 s
Neutral	when after a few minutes			2.5 s
Affect reason	he spots a petrol station in time and avoids being stranded.	OR	he runs out of petrol and becomes stranded by the roadside.	2.5 s
Comprehension question (after 12/64 narratives)	Mark listens to the radio while driving on the freeway YES [1] NO [7]			(max. 6 s)
	Press "space" to continue to the next story			

All segments were separated by a blank screen of 250 ms, except the rating task and continuation.

Methods

Participants

For the current study, we tested 51 participants (41 female, 9 male, and 1 person identified as another gender), with ages ranging from 18 to 29 (*M* = 21.65 year, *SD* = 2.56) who were recruited from the participant database of the Utrecht Institute of Linguistics (UiL) OTS/Institute for Language Sciences. All participants were native Dutch speakers and pursuing a university or HBO (applied higher education) degree. Furthermore, they had no Botox injections in the face and normal or corrected-to-normal vision. Those who had participated in any of the previous studies with (variants of) our stimuli or attended lectures or presentations about the purpose of this study were excluded from the

participant pool. Participants received a compensation of €20 for their participation in the two-hour experiment.

This study and its procedures were approved by the Faculty Ethics Assessment Committee – Humanities of Utrecht University (reference number 22-013-02). All participants read the information letter detailing the nature of the experiment and the procedures before the start of the experiment. The informed consent form emphasized that participants have the right to withdraw consent at any time during the experiment without having to explain why. The consent form was implemented in the coding of the software used to present our stimuli; the experiment only started running if participants pressed the "YES, I consent" button.

Materials and Design

For the current study we used the same sixty-four Dutch narrative stimuli as 't Hart et al. (2019) which had been pretested for 't Hart et al., (2018), but now with the addition of an embedded rating task (see Table 1). Like the previous experiments, the current experiment had a fully crossed 2 x 2 within-subjects design, crossing Character Morality (moral vs. immoral behavior) and Event Valence (positive vs. negative affective event).

Due to our crossed design, there were four versions of each narrative, i.e., 1. moral – positive (good Mark, happy), 2. moral – negative (good Mark, frustrated), 3. immoral – positive (bad Mark, happy), 4. immoral – negative (bad Mark, frustrated). We reused the four already existing pseudo-randomized lists of all sixty-four narratives in which: A) per list, each narrative occurs once in one of the four variants; B) all participants see 16 narratives in each of the four conditions (8 with a male and 8 with a female main character); C) average item properties in each list are similar in terms of pro-sociality and expectedness; D) two lists have the reverse order of the two other lists; E) each narrative occurs with both moral and immoral and both female and male main characters across the four different lists, with the exception of nine narratives that have fixed gender due to stereotypical behavioral expectations. Our main dependent variable was activity of the corrugator supercilii in response to the critical segments.

The newly added rating task consisted of the words "I find this behavior" and a rating scale consisting of the numbers 1 through 7 with the labels 'very bad' and 'very good' written above the 1 and 7 respectively. The task was presented for a maximum of 5.250 ms during which participants had to press one of the seven letters on the bottom row of a QWERTY-keyboard that were correspondingly labeled from 1 through 7 with stickers. The main goal of the embedded rating task was to boost evaluation of the affective state adjective segment later on, but it also functioned as a self-report measure of the readers' conscious evaluation of characters' behavior. To encourage participants to carefully read our narratives, we also distributed twelve comprehension questions over the

experiment which consisted of either true or false statements about the preceding trial. Participants had to indicate whether the statements were true by pressing a button for 'Yes' or 'No' and all statements were identical for all four stimuli lists. Six of the twelve statements were true and the correct answers were not reliant on experimental conditions.

We also included two additional tasks for exploratory purposes. In our Story Fairness Task, we presented our narrative stimuli again, now in their entirety, and asked participants to rate how fair they thought the ending of each narrative was on a horizontal 7-point rating scale with only the ends labeled 'very fair' and 'very unfair'. The Story Fairness Task aimed to explore conscious fairness-based evaluation of our stimuli to see to which extent it would correlate with the EMG responses to the state adjectives, since this was expected to be modulated by spontaneous or even unconscious evaluation. There were four versions of the Story Fairness Task corresponding to all four pseudo-randomized stimuli lists used in their EMG session.

The last task consisted of a Justice Sensitivity Trait questionnaire to obtain some insight into moral evaluation and moreover, to explore participants' sensitivity towards justice, for instance how much they cared about bad people being punished and good people being rewarded (see Supplementary Materials for the full questionnaire. Based on the ratings we were able to assign them a justice trait score as if their sensitivity were a personality trait. The first part consisted of two open questions asking participants for comments on their judgement process during the Story Fairness Task and on characters and situations in the narratives. The second and main part of the questionnaire consisted of 17 statements, e.g., "If someone behaves badly they also deserve to be treated badly by others" and "I think that it's unfair if something good happens to a bad person", and required participants to rate to which extent they (dis)agreed with them on a 5-point rating scale from 'Completely true' to 'Completely untrue'. All statements were written in such a way that a higher level of agreement corresponded to participants more strongly valuing the principle of 'wanting positive things for good people and negative things for bad people'. Due to the task being a self-report measure, responses may reflect both participants' personal values and beliefs concerning justice, as well what they think is socially sensible or desirable.

Experiment Procedure and Data Acquisition

Participants were welcomed and seated in a comfortable chair where they were asked to read the information letter unless they indicated that they had already read it beforehand. They were seated at approximately 60 cm from the laptop screen on which we presented our experimental stimuli and collected participants' answers indicated by button responses. After participants confirmed that they were properly informed and agreed to participate in the experiment, we started preparations for the

fEMG experiment. After participants had read the detailed instructions and performed two practice trials to familiarize themselves with the procedure, they had the opportunity to ask any remaining questions before starting the experimental stimuli.

The stimuli were presented in Times New Roman (font size 20) and segment presentation was automatically timed as indicated in Table 1. The presentation rate, i.e., starting the next trial by pressing the space bar, was self-timed. The experimental trials were divided into four roughly equal blocks with three breaks during which we checked how the participant was doing and offered them some water.

fEMG studies on affective valence generally include both the corrugator to measure frowning and the zygomaticus, involved in upwards movement of the mouth corners, is often measured an indication of smiling (Larsen et al., 2003). However, the zygomaticus appears to be a less reliable measure of valence as it has a relatively poor level of repeatability and it is more sensitive to crosstalk from neighboring muscles associated with negative emotions, for instance disgust (Larsen et al., 2003; Tan et al., 2012). Our main focus is therefore on the corrugator signal and although we did measure the zygomaticus for comparability reasons (‘t Hart et al., 2019; 't Hart et al., 2018; 't Hart et al., 2021), findings from the zygomaticus are not reported in this paper. fEMG was continuously measured on the right side of the face with reusable Ag/AgCl electrodes (2 mm contact area) over the standard recording sites for the corrugator supercilii and zygomaticus major (Van Boxtel, 2010). The raw EMG data was recorded with a sampling rate of 2048 Hz using the NeXus-10 MKII biofeedback device (Mind Media) in combination with BioTrace+ software (MindMedia). Including preparations and instructions, the fEMG experiment lasted approximately 90 minutes.

After the EMG part was finished, the electrodes were removed, and participants continued with the Story Fairness Task and Justice Sensitivity questionnaires on a lab computer for exploratory purposes followed by a short exit questionnaire. Altogether, this lasted around 30 minutes. When participants finished the questionnaires, they were debriefed, given their financial compensation, and thanked for their participation. One experiment session was approximately 2 hours in total.

Data preparation and Analyses

Our planned analyses were preregistered during data collection but before looking at any of the data. A copy of the preregistration can be found here: EMG4 Preregistration

Analysis of fEMG

We used a custom made tool, EDF+Checker (https://github.com/UiL-OTS-labs/EDFPlusChecker), to check for time differences between the experiment triggers of the segments and the corresponding

markers in the EMG signals as these were used for time-locking. We then used BrainVision Analyzer 2 (Version 2.2.2) to filter the checked data with a band-pass filter between 20 and 500 Hz (48 dB/octave roll-off) and a notch filter at 50 Hz (Van Boxtel, 2010) and to perform signal rectification. Finally, we used segmentation in BrainVision to select the 2.000 ms epoch of baseline activity recorded during the pre-stimulus neutral distractor image (forest scene), and 58-s epochs each containing one full trial from baseline. These data were exported to MatLab for further segmentation.

For the baseline, we selected the mean corrugator activation during the last 1500 ms of the 2000 ms baseline signal since the first 500 ms tended to contain some noisy signal and artifacts. Corrugator activity from the whole trial epoch was segmented into three parts, time-locked to the pre-selected 1-s epoch of the character morality segment (3-4 seconds post-onset), the affective state adjective (0.5-1.5 second post onset), and for exploratory analyses the affect reason (1-2 seconds post-onset. Average corrugator activation in these 1-s epochs were reported as percentage of baseline activation. (See Supplementary Data for a graph with continuous average corrugator activity of all four conditions in 100 ms bins for the whole trial).

No participants were excluded from the main analyses based on amount of correct comprehension statements, however, one trial of one participant was excluded due to extreme outliers in corrugator activation when reading the morality segment with corrugator activity up to 46449% compared to baseline. This is a data loss of 0.03%.

We used R (version 4.2.1) (R Core Team, 2022) to analyze the fEMG data with Linear Mixed-Effects Models with lme4 (version 1.1-30) (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest packages (version 3.1-3) (Kuznetsova, Brockhoff, & Christensen, 2017). We then used the emmeans package (version 1.8.1-1) to look at average activation per condition and pairwise comparisons (Lenth, 2022). As preregistered, the dependent variable in our models was corrugator activity during the pre-selected epochs and the independent variables were character morality and event valence.

Furthermore, following the preregistration, all statistical models, including the intercept models, included subject and item as random intercepts. We did not include random slopes. For the morality segment we started with an intercept model with subject and item as random intercepts and corrugator activation during the selected 1s epoch as the dependent variable. We then added character morality as a fixed factor to create the mixed model and performed pairwise comparisons to look at significant effects. For both the affective state adjective segment and the reason segment, we again started with intercept models with subject and item as random intercepts and corrugator activity during the selected 1s epoch of the respective segments as the dependent variable. We were not interested in main effects of the morality or valence conditions, so we immediately added character morality and event valence as interacting fixed factors to create the mixed models and

performed pairwise comparisons to look at significant effects. The reason segment is discussed in the exploratory section as per our preregistration.

Exploratory analyses

Analysis of Embedded Rating Task

The embedded rating task is the task that requests participants for every narrative stimulus to rate the behavior of the main character on a 7-point scale (1-'very bad' to 7-'very good') immediately after they read the description of said character's moral or immoral behavior in the morality segment. We were curious to see to which extent the ratings on (im)moral behavior would correlate with participants' corrugator response when reading descriptions of said (im)moral behavior. We expected lower scores (badly-rated behavior) on the embedded rating task to be associated with higher corrugator activity (more frowning). We therefore used Spearman's product-moment correlation coefficient and looked at each participant's average corrugator response during the pre-selected 1-s epoch of the morality segment and their rating response regarding that same segment. For the correlation analysis we converted the corrugator signal to the log10 because the ratings were not normally distributed. This correlation analysis does not take into account that there are different subjects and different narratives with multiple ratings each.

Analysis of Story Fairness

Rating responses in our Story Fairness Task corresponded to values from -3 ('very unfair') to 3 ('very fair') with the median, 0, being regarded as a neutral rating. Thus the higher the score, the more fair the narrative's ending was judged. A combined amount of 87 ratings from 7 different participants were excluded from the analysis due to an error in the creation of the four Story Fairness Tasks. These issues were quickly fixed so the data of all other participants was complete. Furthermore, there was 1 participant whose ratings on all 64 narratives were excluded from analysis as everything was rated as neutral. This resulted in a total data loss of 4.6%. With the Story Fairness rating responses we were able to explore if and how the hypothesized effect of fairness-based evaluation would be reflected in a self-report measure as evaluation for this response is bound to occur on a more conscious level. Therefore, we built a linear mixed model again focused on the interaction between the fixed factors morality condition (moral, immoral) and event valence (positive, negative) and with subject and item as random effects, but now with the Story Fairness ratings as the outcome measure. Finally, we performed pairwise comparisons to look for significant effects.

Analysis of Justice Sensitivity Task

Rating responses on the Justice Sensitivity Trait questionnaire corresponded to values ranging from 2 to -2 (completely true/fair – completely false/unfair), 0 being neutral. Thus, in all cases a higher rating implied assigning more value to justice or fairness. We performed a correlation analysis of the rating responses to find out which statements best measured participants' attitudes toward justice and based on this we chose to use responses 14 of the 17 questions (Cronbach's Alpha $\alpha = 0.79$) for further analyses. All participants received a 'justice trait score' consisting of their mean ratings on the 14 selected statements ($M = 0.90$, $SD = 0.5$), which was then added to the statistical models from our analyses of the segments. Because the continuous justice scores did not allow for pairwise comparisons, we performed a median split to divide the participants into two groups, i.e., a 'high justice group' containing participants with scores above the median ($N = 25$) and a 'low justice group' with those who scored below the median ($N = 22$). Note that this excludes participants with the median as justice score ($N = 4$). The statistical analyses were performed using the raw justice trait scores so it includes all participants, whereas the pairwise comparisons and figures are based on the data after the split into the two groups.

Results

Character Morality Segment

As predicted, Figure 2 shows the effect of character morality (e.g., moral vs. immoral) on average corrugator activity: reading descriptions of immoral behavior clearly elicited more frowning than moral behavior. The 1-s epoch we pre-selected for analysis is indicated with vertical lines and also within this time window the average corrugator response differs substantially. The statistical analysis corroborates that descriptions of immoral behavior elicited higher corrugator activity than descriptions of moral behavior (respectively 162% and 106% of baseline activation, $difference_{immoral-moral} = 56.52$, $t(3150.45) = 10.48$, $p < 0.001$, 95% CI [45.95, 67.10]), meaning that we found a main effect of character morality. This demonstrates that the character manipulation in our stimuli was successful, replicating findings of the two previous studies where identical character morality segments were used ('t Hart et al., 2019; 't Hart et al., 2018).

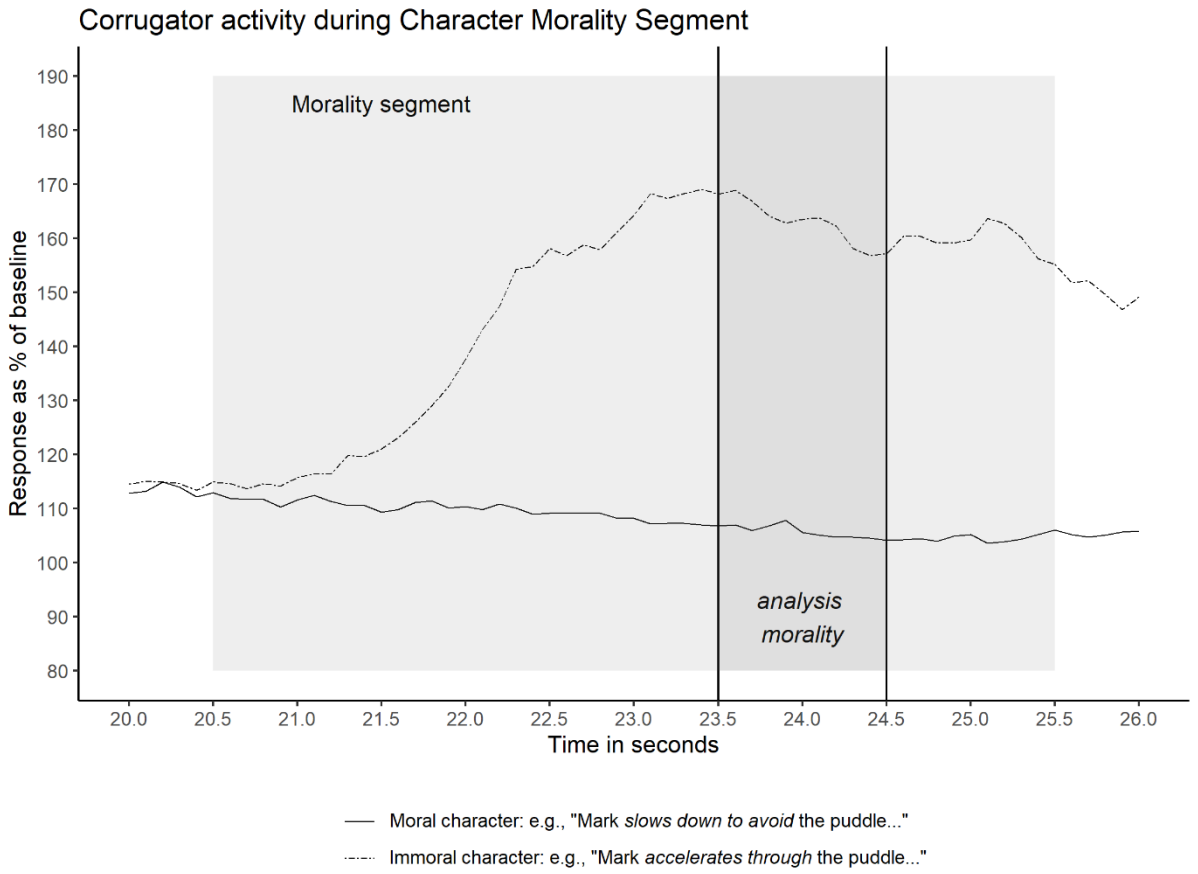


Figure 2 | Observed averages of corrugator responses to character morality. The timing of the entire character morality segment is highlighted as well as the pre-selected 1-s epoch that was used for analysis.

Affective State Adjective Segment

The affective state adjective segment consists of an isolated adjective referring to the main character's affective state (e.g., positive vs negative) as a result of something that is mentioned in the reason segment later on, and is the main focus of our study. Figure 3 shows the average corrugator activity during the last narrative segments with the two pre-selected 1-s epochs that were used for our analyses highlighted between the vertical lines. The figure also shows that the epoch pre-selected to analyze the corrugator activity for the state adjective again nicely frames the noticeable phasic response.

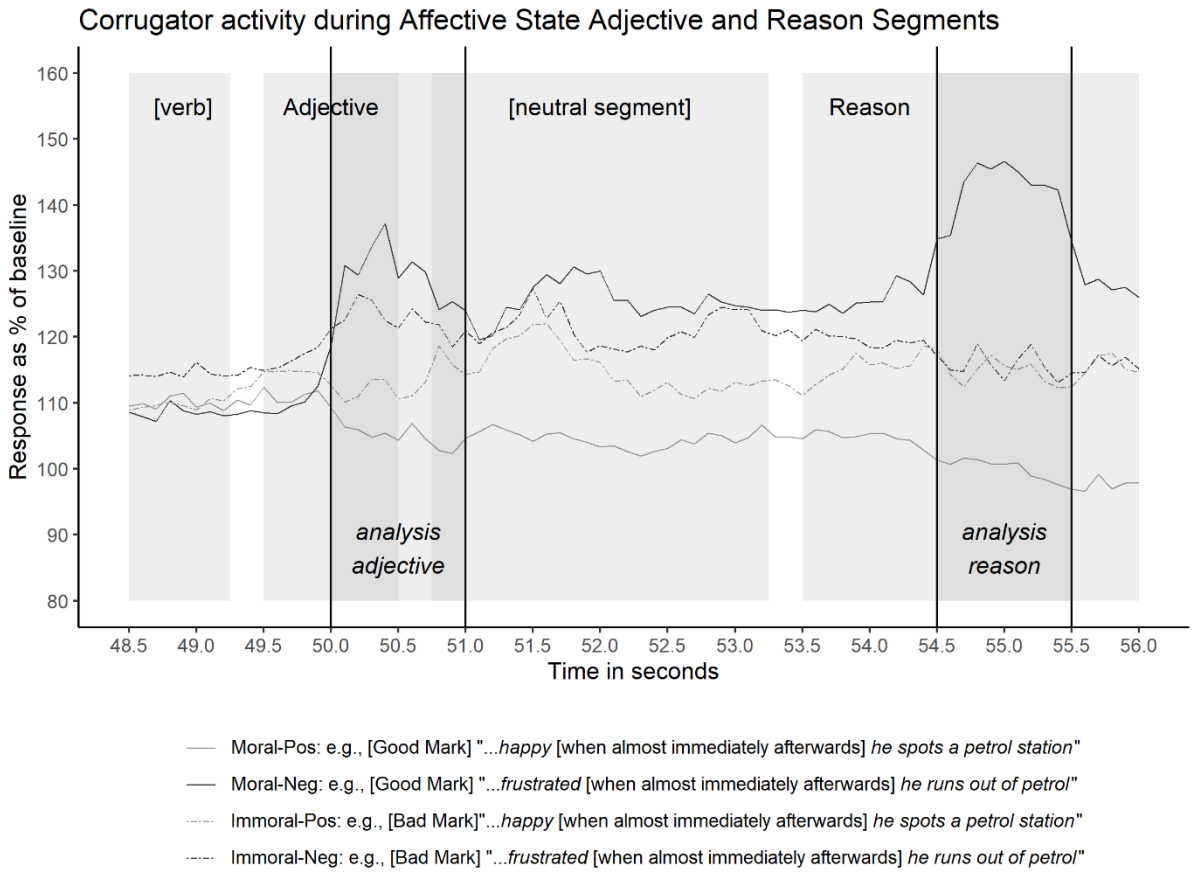


Figure 3 | Observed averages of corrugator responses during the narratives' last four segments (including the intersegmental 250 ms blank screen), the segments containing the adjective and reason being the critical segments used for analyses. The pre-selected 1-s epochs that we used to analyze the corrugator response to these segments are again highlighted between the vertical lines.

Our statistical analysis of the corrugator activity during the analyzed epoch regarding the state adjective found that the interaction between character morality (moral, immoral) x event valence (positive, negative) was significant ($F(1, 3150.4) = 6.32$, $p = 0.01$). We found that corrugator activity was higher when moral characters were described with negative adjectives than positive adjectives ($difference_{neg\text{-}pos} = 23.61$, $t(3153.48) = 5.65$, $p < 0.001$, CI [-34.02, -13.19]), which confirms our predictions and is in line with the postulated joint effect of simulation and evaluation.

An unanticipated finding was that corrugator responses to positive and negative adjectives describing immoral characters did not seem to differ ($difference_{neg\text{-}pos} = 8.76$, $t(3153.44) = 2.10$, $p = 0.14$, CI [-19.18, 1.66]), which is in contrast with our expectations of finding an evaluation-based valence effect. This means that we unexpectedly replicated the results for both moral and immoral characters from the 2019 study, i.e., a substantial adjective valence effect in the moral conditions and a null effect of valence in the immoral conditions ('t Hart et al., 2019).

Our exploratory results discussed below may give some additional insights into the process of reading our narratives including the corrugator response to the reason segment and conscious judgements of character morality and story fairness. We also looked at the impact of readers' attitude towards justice as a personality trait.

Exploratory results

Affect Reason Segment

The event reason segment is the last segment of the narrative and describes the reason for the main characters affective state, so logically, these two segments always appear in the same event valence condition. In two prior studies we found that reading the reason appeared to elicit renewed phasic corrugator activity, similar to the differential corrugator response at the adjective but of a larger magnitude ('t Hart et al., 2019; 't Hart et al., 2021). Similarly, we again look whether our expected corrugator pattern with boosted evaluation is visible for this segment. Figure 3 shows that the pre-selected time window for the reason segment again accurately targets the corrugator response to the event reason.

The interaction between character morality (moral, immoral) x event valence (positive, negative) was found to be significant $F(1, 3150.2) = 38.68$, $p < 0.001$). Pairwise comparisons revealed that corrugator activity was much higher for negative than positive event reasons in the case of moral characters ($difference_{neg\text{-}pos} = 42.27$, $t(3153.27) = 9.06$, $p < 0.001$, CI [-53.90, -30.63]) but this was not the case for immoral characters ($difference_{neg\text{-}pos} = 1.23$, $t(3153.23) = 0.26$, $p = 1.00$, CI [-12.87, 10.40]). This means that we indeed see the same valence-based differential pattern that we found for the state adjectives describing moral characters on a slightly larger scale, as the solid lines in Figure 3 are even further apart, and no differential pattern for adjectives describing immoral characters. Again, these results are a replication of the findings from the previous study ('t Hart et al., 2019).

Embedded Rating Task

We found a weak negative correlation between the embedded rating and corrugator activation during the pre-selected 1-s epoch of the morality segment, $r(3228) = -.21$, $p < 0.001$. This indicates that lower ratings, indicating worse behavior, were associated with higher average corrugator activity, which is in line with our expectations.

In obtaining Pearson's *r*, we looked at the relationship between corrugator activity and rating within every trial of all participants, regardless of the conditions or the various participants or trials. This may have contributed to the correlation coefficient being low. All things considered, this

preliminary result loosely supports the notion that participants frowning responses are indeed indicative of how they judge characters' behavior in terms of valence.

Story Fairness Ratings

Our mixed model with morality condition (moral, immoral) x event valence (positive, negative) and Story Fairness rating as the dependent variable showed that the interaction was significant $F(1, 3006.0) = 4992.98$, $p < 0.001$). The interaction is also shown in Figure 4 below and the average ratings, with positive numbers indicating 'fair' and negative indicating 'unfair', were as follows:, immoral – positive -1.27, immoral - negative 1.53, moral – positive 1.95, moral – negative -1.37.

Pairwise comparisons revealed that ratings were lower, i.e., indicating more 'unfair', for negative than positive events in the case of moral characters ($difference_{neg\text{-}pos} = -3.31$, $t(3008.81) = -54.11$, $p < 0.001$, CI [3.16, 3.47]), whereas they were higher for the immoral ones ($difference_{neg\text{-}pos} = 2.81$, $t(3006.79) = 45.77$, $p < 0.001$, CI [-2.96, -2.65]. The Story Fairness ratings also differed between morality conditions with ratings being lower for positive events in the immoral conditions than moral conditions ($difference_{immoral\text{-}moral} = -3.22$, $t(3016.90) = -52.48$, $p < 0.001$, CI [3.07, 3.38]) but higher for negative events in the immoral condition condition ($difference_{immoral\text{-}moral} = 2.90$, $t(3013.84) = 47.35$, $p < 0.001$, CI [-3.05, -2.74]).

Thus, the average ratings precisely show the evaluation-based valence patterns that we also expected to find in the EMG signal.

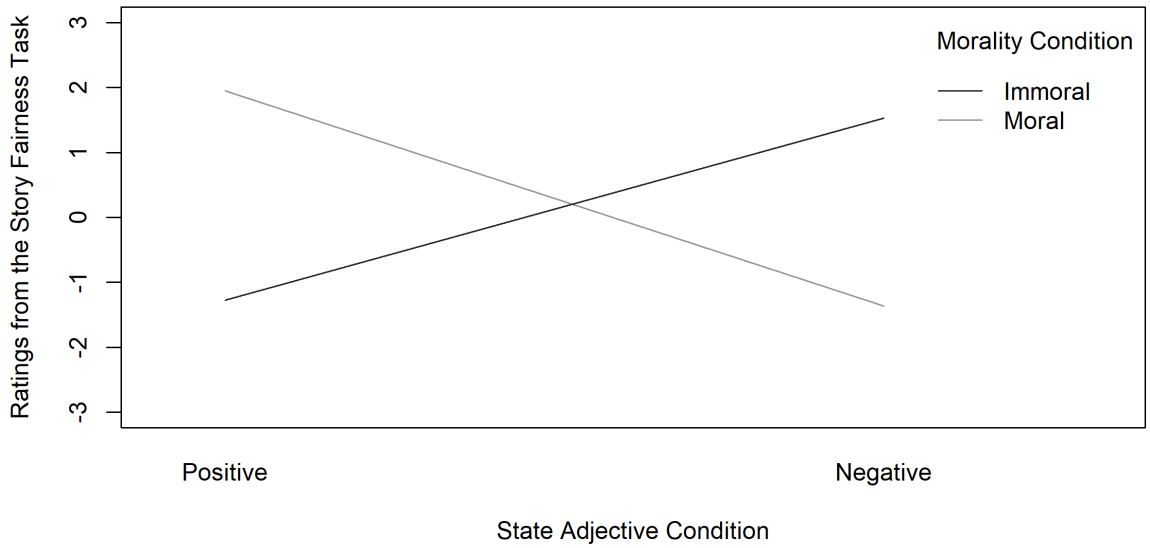


Figure 4 | The interaction between Morality Condition and State Adjective Condition on the ratings of the Story Fairness Task.

Justice Sensitivity as Personality Trait

The following statistical analyses with the justice trait scores were performed using the raw scores so it includes all participants. The pairwise comparisons and figures are based on the data after the median split into a high justice group and a low justice group consisting of participants with higher and lower justice trait scores.

Justice Sensitivity and Character Morality

We added participants' justice trait scores as an additional fixed factor in the analysis for the morality segment and found that the interaction between character morality (moral, immoral) x justice trait score (continuous numerical) was significant ($F(1, 2907.4) = 44.30$, $p < 0.001$). We then performed a median split to allow pairwise comparisons and to create the graph in Figure 5.

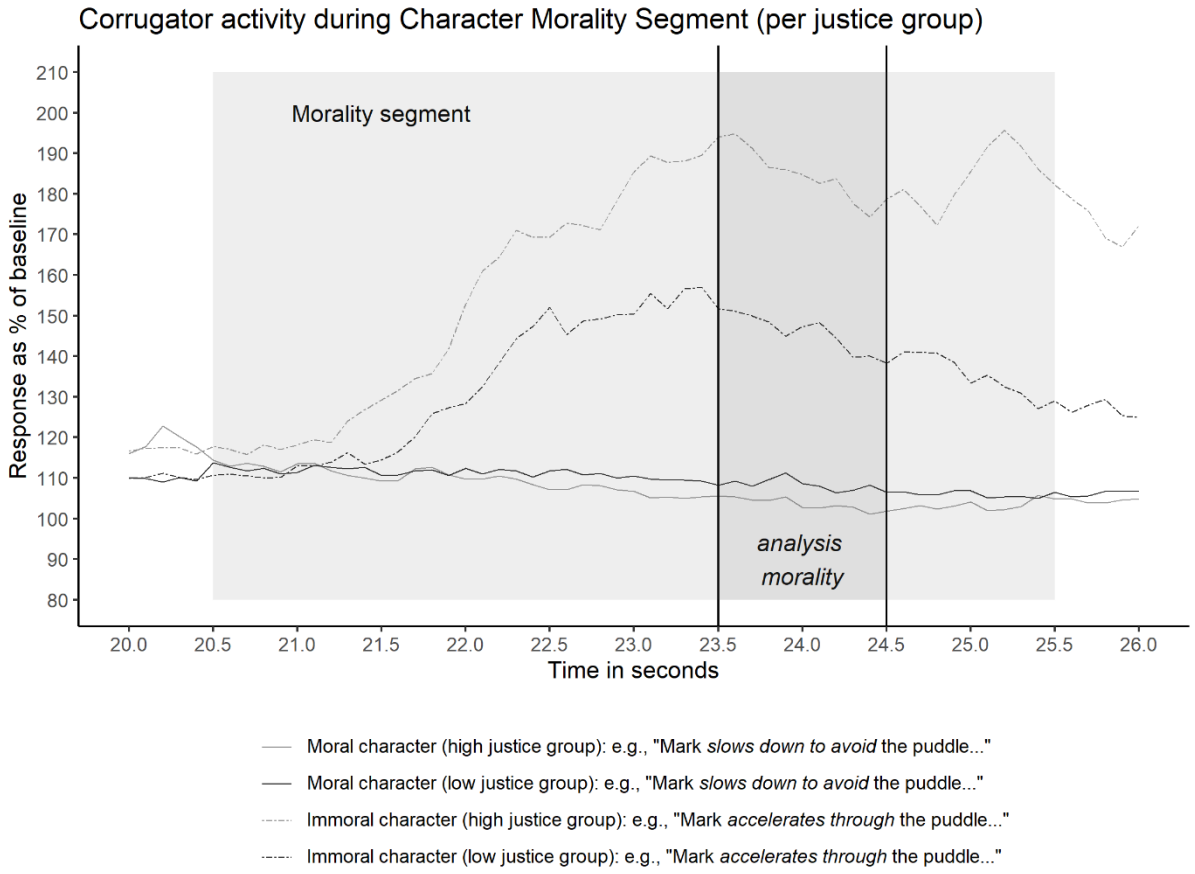


Figure 5 | Observed average corrugator responses to character morality in the high justice group. Again, the timing of the entire character morality segment is highlighted as well as the pre-selected 1-s epoch that was used for analysis.

In the high justice group mean corrugator activation compared to baseline activity was 184.3% for immoral characters and 102.7% for moral characters and in the low justice group this was 143.0%

for immoral characters and 108.5% for moral characters. Pairwise comparisons following the median split showed that corrugator activation differed between moral and immoral characters both in the high justice group (difference$_{immoral\text{-}moral}$ = 81.64, t(2944.35) = 9.88, p < 0.001, CI [-102.23, -61.05]) and the low justice group (difference $_{immoral\text{-}moral}$ = 34.56, t(2924.93) = 4.47, p < 0.001, CI [-53.83, -15.28]).

This means that the moral manipulation had the intended effect in both groups. While the two groups did not differ significantly in either morality condition, Figure 5 does suggest that the readers in the high justice group respond slightly stronger to our morality manipulation than those in the low justice group.

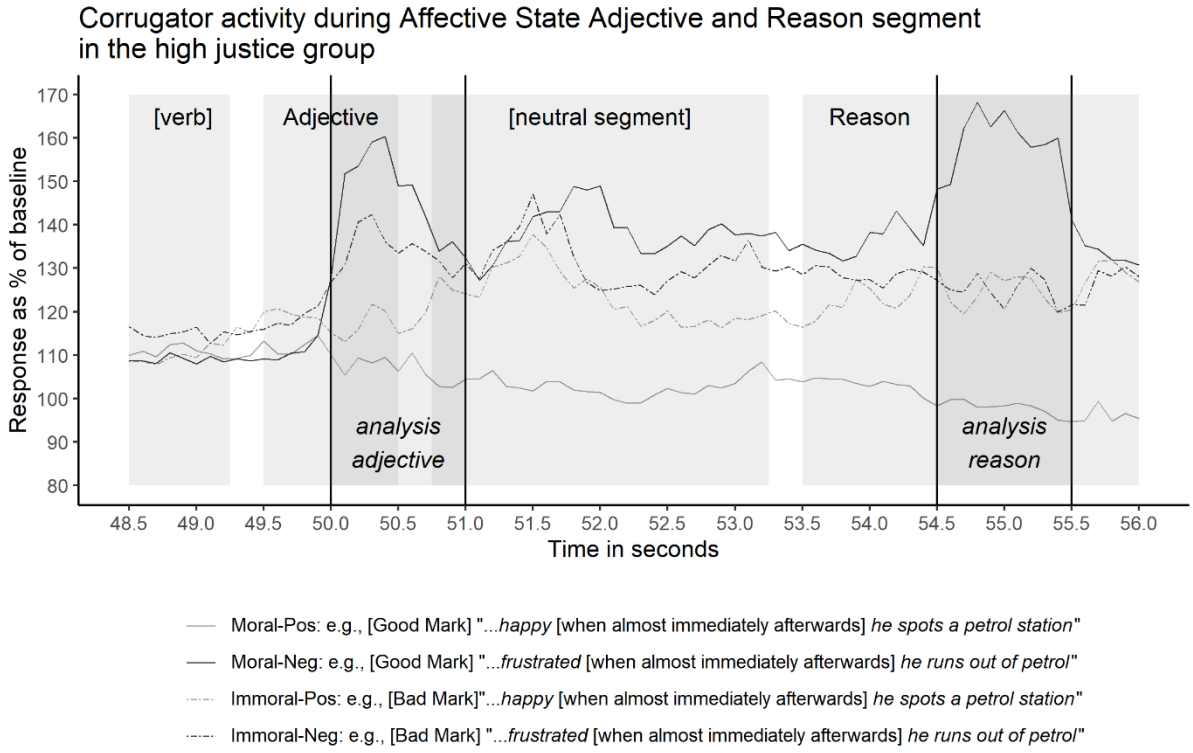


Figure 6 | Observed average corrugator responses during the narratives' last four segments (including the intersegmental 250 ms blank screen) for participants in the high justice group. The pre-selected 1-s epochs that we used to analyze the corrugator response to these segments are again highlighted between the vertical lines.

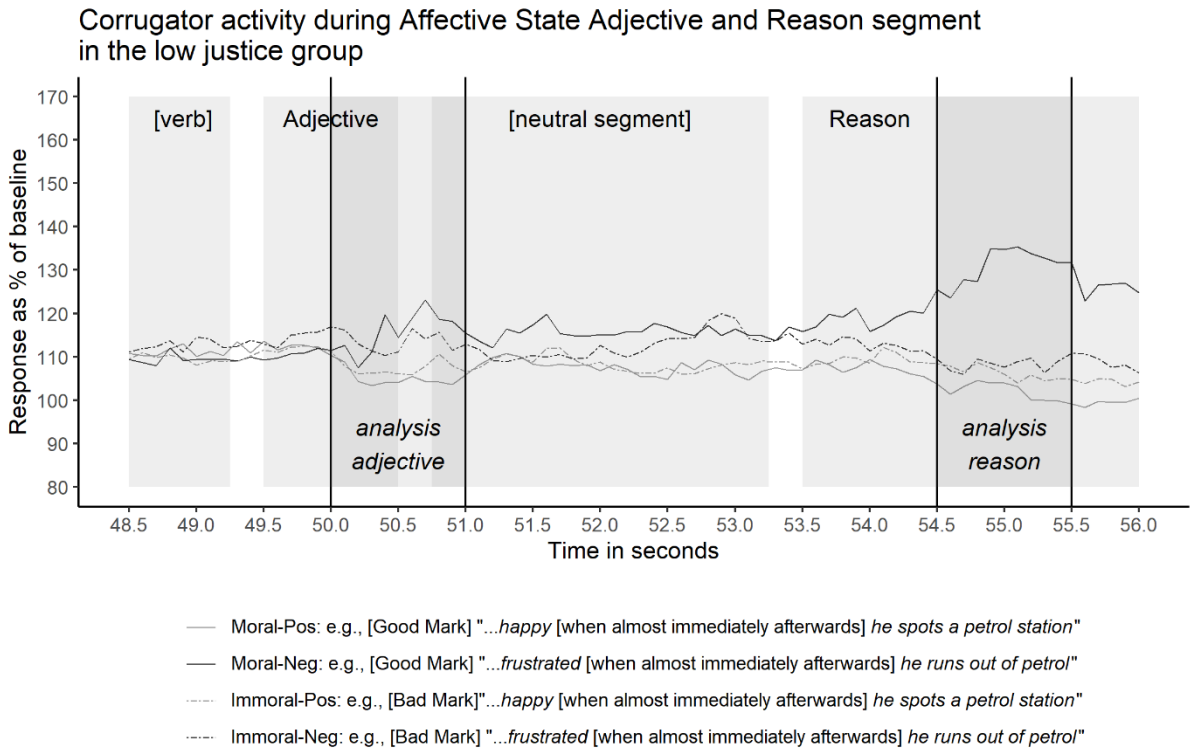


Figure 7 | Observed average corrugator responses during the narratives' last four segments (including the intersegmental 250 ms blank screen) for participants in the low justice group. The pre-selected 1-s epochs that we used to analyze the corrugator response to these segments are again highlighted between the vertical lines.

Justice Sensitivity to Affective State Adjective Segment

We added justice trait score as a factor in the analysis for the adjective segment and found that the three-way interaction between character morality (moral, immoral) x event valence (positive, negative) x justice trait score (continuous numerical) was significant ($F(1, 2927.2) = 3.89$, $p = 0.049$). Figure 6 and 7 show the average corrugator activation for the adjective segment across the conditions for respectively the high and low justice groups.

Interestingly, pairwise comparisons following the median split showed that corrugator activation of participants in the high justice group was significantly higher when a moral character's affective state was described with a negative adjective rather than a positive adjective ($difference_{neg\text{-}pos} = 38.01$, $t(2938.92) = 5.87$, $p < 0.001$, CI [-56.54, -19.49]) but this was not the case for participants in the low justice group ($difference_{neg\text{-}pos} = 9.73$, $t(2932.58) = 1.60$, $p = 0.75$, CI [-27.10, 7.60]).

While preliminary, these findings suggest that readers in the high justice group were affected more strongly by to our conditions. Nevertheless, there were still no significantly different corrugator responses between the immoral-negative and immoral-positive conditions either in the high ($difference_{neg\text{-}pos} = 13.46$, $t(2938.51) = 2.08$, $p = 0.37$, CI [-32.00, 5.08]) or low justice group

(difference$_{neg-pos}$ = 4.91, t(2932.58) = 0.81, p = 1.00, CI [-22.28, 12.46]) so we still did not find the expected evaluation-based valence effect.

Justice Sensitivity to Affect Reason Segment

Similar to the analysis for the adjective segment above, we added justice trait score as a factor in the model for the reason segment. Again, the three-way interaction between character morality (moral, immoral) x event valence (positive, negative) x justice group (continuous numerical) was found to be significant (F(1, 2825.0) = 17.49, p < 0.001).

Pairwise comparisons after we performed the median split showed that corrugator activation was significantly higher in the moral-negative condition than in the moral-positive condition both in the high justice group (difference$_{neg-pos}$ = 59.69, t(2807.54) = 8.07, p < 0.001, CI [-80.83, -38.54]) and the low justice group (difference$_{neg-pos}$ = 29.39, t(2803.79) = 4.07, p < 0.001, CI [-50.06, -8.71]).

This means that the corrugator responses of readers in both groups showed the expected valence effect in moral conditions despite it being on a much smaller scale in the low group. For the immoral conditions there were again no differences between corrugator responses to positive and negative adjectives in either the high justice group (difference$_{neg-pos}$ = 0.36, t(2807.18) = 0.05, p < 0.001, CI [-21.52, 20.80]) or the low justice group (difference$_{neg-pos}$ = 2.36, t(2803.80) = 0.33, p < 0.001, CI [-23.03, 18.31]).

Discussion

Our emotion systems become active when we process affective language, which consequently gives rise to facial muscle activation, but which cognitive processes does this reflect? Three of our previous studies explored how mental simulation and moral evaluation may influence corrugator activity and found evidence favoring the multiple-drivers account of language-induced emotion: simulation and evaluation acting as joint forces for moral characters and counteracting forces for immoral characters (‘t Hart et al., 2019; 't Hart et al., 2018; 't Hart et al., 2021). Nevertheless, even if the multiple-drivers account best explained the results, it still had to be interpreted with caution as the findings could not confirm that the corrugator responses was modulated by simulation and evaluation forces, nor could they verify the postulated interplay between the forces, i.e., joining or counteracting each other depending on contextual factors.

Our current study therefore aimed to boost effect of evaluation force as this should, under the multiple-drivers account, change the net effect on corrugator activation. To increase the effect of evaluation while keeping other factors the same, we used the same narratives as our 2019 study (‘t Hart et al., 2019) but now with an embedded rating task to stimulate readers to actively evaluate and

judge a characters' moral status. This allowed us to hypothesize how the corrugator patterns should subsequently respond to the narrative segments, with the time-locked segment of the isolated state adjective being the critical focus for our account.

Crucially, and in contrast to previous findings, we predicted to find more corrugator activation for positive than negative adjectives in the case of immoral characters. Furthermore, we expected to replicate the previously found corrugator patterns for the morality manipulation segments and for adjectives describing moral characters, respectively more frowning for descriptions of immoral behavior or negative adjectives than for descriptions of moral behavior or positive adjectives. Our core results will be discussed per EMG-segment followed by our exploratory findings, and a general discussion.

Reading About Moral and Immoral Behavior

The character morality segment consists of a sentence in which the main characters' moral or immoral behavior is described, thereby setting the stage for morality-based evaluation of fairness.

As expected, we replicated the corrugator patterns found in our prior studies ('t Hart et al., 2019; 't Hart et al., 2018), namely more frowning for reading immoral behavior than moral behavior. This reaffirms that a description of (im)moral behavior can be used as an emotionally salient stimulus even if it is a relatively long text when compared to the frequently used single words or short phrases (Aziz-Zadeh, Wilson, Rizzolatti, & Iacoboni, 2006; Fino et al., 2016; Foroni & Semin, 2009; Larsen et al., 2003). It also confirms that the corrugator muscle is susceptible to the effect of moral valence. It should be noted that the previous analyses of the morality segment looked at corrugator activation during the entire 5-s epoch that the sentence was presented, whereas the current analyses only selected a predetermined 1-s based on the earlier findings ('t Hart et al., 2019; 't Hart et al., 2018).

Reading About Character Affect

The affective state adjective segment contains a single isolated adjective describing the affective state of the main character, e.g., "happy" or "angry" because of some reason mentioned later in the narrative. The main focus of this study was the corrugator response to state adjectives since a boosted effect of evaluation was expected to cause two differential corrugator patterns, a crucial difference from previous findings. Finding such a pattern while replicating the corrugator pattern for moral characters would argue in favor of both simulation and evaluation driving the corrugator response under the multiple-drivers account.

We now expected to find a small differential and evaluation-based valence effect in the case of immoral characters: moral evaluation should override the counteracting mental simulation, causing

more frowning for positive rather than negative adjectives. Contrary to our expectations, we did not find any difference at all between corrugator responses for immoral characters' positive and negative affective states. Moreover, descriptively speaking the signals actually resemble a regular valence-based response rather than the hypothesized inverse pattern based on evaluation (see Figure 4).

For moral characters, we expected to replicate the valence-based differential corrugator pattern because mental simulation and moral evaluation are again predicted to join forces in terms of valence under the multiple-drivers account, though the differential pattern could potentially arise on a larger scale due to enhanced evaluation. We indeed replicated the differential corrugator pattern for moral characters as we found substantially higher corrugator activity in response to negative than positive adjectives. The differential pattern of average corrugator activation appeared to be similar in size as in the prior studies, but this was not further investigated here.

The pattern for immoral characters, which was the focus of this study, does not fit with our hypothesis of evaluation overriding simulation but it does largely replicate previous findings. This may be a result of the rating task not having the expected effect of boosting evaluation, but another possible explanation is that the multiple-drivers account, or at least our current understanding of it, is incorrect. In any case, our findings at the adjective once more argue against a simulation-only or evaluation-only account.

Finally, the fact that the analyzed 1-s epoch was based on our prior findings and precisely captured the peak of the phasic corrugator response suggests that the timing of the corrugator response to adjectives appears to be consistent and robust, at least in the context of our narratives.

Exploratory findings

Reading About Reasons for Character Affect

The reason for a character's affective state is described in a subclause in a separate segment, e.g., "he spots a petrol station in time and avoids being stranded". This segment appeared to elicit a renewed phasic corrugator response in our prior studies ('t Hart et al., 2019; 't Hart et al., 2021), indicating that this segment actually covers additional affectively-salient information despite not consistently containing an element with obvious valence in a similar position in the subclause.

Based on the multiple-drivers account, we again expected to find a small differential corrugator pattern reflecting a small evaluation-based valence effect in the case of immoral characters, i.e., more frowning for positive than negative events, however, we did not find a differential corrugator response here. Figure 3 shows that the corrugator patterns for both positive and negative events are in fact right on top of each other. Moreover, it appears that the frowning

pattern does not reflect any valence at all, as the average corrugator activation was barely higher than the baseline.

For moral characters, the multiple-drivers model predicted a large differential corrugator response reflecting a standard valence effect, essentially replicating the findings from the previous studies. Our results indeed showed substantially more frowning for negative than positive events in the moral condition, which together with the findings for immoral characters argues in favor of the multiple-drivers account. This is a replication of the findings from 't Hart et al. in 2019. A slight difference, though this comparison should be interpreted with caution, is that our frowning patterns for reading bad people's reasons for positive and negative affect seem to approach each other whereas they appeared to diverge at this segment in the 2019 study ('t Hart et al., 2019).

Again, our analyzed 1-s epoch, preselected based on prior corrugator patterns, precisely captured the phasic corrugator response. This suggests that the timing of the corrugator response to our affect reason segments is quite consistent and robust in the context of our narratives, despite the lack of a consistent and distinctly valenced element at this location in the narrative.

Conscious measures of morality, fairness, and justice sensitivity trait

We suspected that lower ratings on the embedded rating task might be associated with higher corrugator activation for descriptions of immoral behavior. Indeed, we found that there was a weak negative association which supports the notion that both rapid physiological responses such as frowning as well as conscious judgments regarding immoral behavior reflect negative valence. The link may be quite weak due to interference from the correlation analysis not distinguishing between the different morality conditions, participants, or trials.

A striking result from the Story Fairness ratings is that they precisely show the pattern that we hypothesized to find in the EMG signals under the multiple-drivers account with boosted evaluation, i.e., for narratives involving moral characters there was a large fairness-based difference between endings rated 'fair' and 'unfair' and for narratives about immoral characters there was a similar but smaller fairness-based difference in ratings. This finding corroborates that our narrative stimuli had the capacity to manipulate fairness-based evaluation in a moral context. Furthermore, the fact that these self-reported judgements of fairness precisely reflect our expectations whereas the fEMG patterns do not could also imply that there is a discrepancy between the postulated evaluation involved in physiological responses such as frowning and conscious evaluation as indicated by self-reported judgements on the fairness of the narratives' endings.

We should also not forget that self-report measures are susceptible to a sociality bias, e.g., participants may adjust their personal opinions based on what they think is a socially desirable or

politically correct answer. This could be considered as a re-evaluation and may reflect a drastically different process compared to the postulated evaluation that we expect to occur rapidly to be noticeable in fEMG.

Lastly, using the Trait Scores from our newly-created Justice Sensitivity Trait questionnaire, we found that readers with lower Trait Scores appeared on average to be less affected by our moral manipulations than readers with higher Trait Scores. Despite character morality having a significant effect on corrugator responses in both groups, responses of readers in the Low-Justice group to positive and negative state adjectives or affect reasons did not differ in either morality condition. Our statistical analyses and Figures 5, 6, and 7 showed similar attenuated average corrugator responses in the low justice group for all three of our analyzed segments. Although the findings with Trait scores are preliminary and did not provide new evidence in favor of our hypotheses under the multiple-drivers account, they do suggest that the trait of justice sensitivity should not be overlooked in fEMG research.

General discussion

Which implications do our results have for our multiple-drivers model? While we did not find the hypothesized boosted effect of evaluation in corrugator responses to adjectives describing immoral characters, we do not think that our results provide clear cause for rejecting the multiple-drivers model either. Rather, the fact that our results largely replicated the findings of our previous three studies (‘t Hart et al., 2019; 't Hart et al., 2018; 't Hart et al., 2021), on which the authors based the initial decision to adopt the multiple-drivers model, points quite strongly to the possibility that our rating task did not achieve the expected effect of boosting moral evaluation. Rejecting the model seems too drastic in the light of our findings, however, it is possible that our interpretation of the multiple-drivers model is incorrect in its current form. Again, the multiple-drivers model did not make assumptions regarding the proportion of either force relative to the other.

An adjusted multiple-drivers model where simulation is a more powerful force than evaluation may be considered, as descriptively speaking, we found that frowning responses to adjectives appeared to reflect simulation-based valence more than evaluation-based valence for all conditions, in spite of our aim to boost evaluation and contrary to our expectations (Figure 4). From a linguistic perspective, it is arguably plausible that simulation of the lexical meaning of words and phrases may to have a stronger influence on processing written language whereas evaluation of their meanings in a given context is more secondary. Based on this assumption, simulation can have a more a primary role in language processing.

A striking finding of the Story Fairness ratings is that the responses here exactly reflect the patterns we expected to find in corrugator responses to morality-based evaluation of fairness even if we did not find it in our fEMG signals. While this confirms that our narratives contain the necessary elements to find an effect of evaluation, it remains an open question why frowning behavior appears to contrast the rating behavior. A discrepancy between conscious and unconscious evaluation of fairness would be able to explain this finding. The process of deliberate and conscious evaluation required for the self-report measure may be fundamentally different from the rapid morality-based evaluation we expected to see reflected in the corrugator responses (Barrett, Niedenthal, & Winkielman, 2007), as effects from personal values and beliefs or social desirability may also interfere with conscious outcome measures that include reflective aspects (Izard, 2009).

It should also be noted that Story Fairness Ratings presented the exact same narratives that had been read during the fEMG beforehand, so an effect of re-evaluation was unavoidable. A future study may counterbalance the order of the fEMG and Story Fairness Task to see how this affects both ratings and frowning behavior. The notion of multiple 'layers' of evaluation can have crucial implications for our multiple-drivers model as well as research on processes of (emotional) evaluation and judgements in general and requires further research.

Additionally, as discussed by (Hoeken & Sinkeldam, 2014; Szanto, 2015) and supported by our preliminary findings with Justice Trait scores, readers may not care as much about bad people and perhaps do not (want to) identify with them to the extent that they are less inclined to simulate and evaluate their affective states. As a result, the effect of evaluation in the context of immoral characters' emotions may be even smaller and perhaps even negligible for our narratives, and thus invisible in our frowning responses. What remains would be an attenuated effect of simulation-only on the frowning response, which would fit the findings of our study at the adjectives descriptively. This is however a preliminary data-driven suggestion that requires more research. We see that there is already a difference between low and high justice groups within our relatively heterogenous sample of participants, implicating that studies with a more homogenous group of participants may have to deal with even more variance between subjects.

A final implication of our study, while not related to our main aim but useful nonetheless, is that our findings demonstrate that the insertion of a rating task or occasional comprehension questions did not obstruct the reading process. Future studies aiming to manipulate simulation and evaluation processes during language processing may benefit from using such tasks and questions, because more research is definitely required to shed light on the many remaining open questions.

Limitations

Our study also has some limitations that we will briefly address. First of all there are some practical limitations, for instance the fact that 41 of our 51 participants were female. Secondly, we did not hide the purpose of the facial electrodes as participants were informed that their facial expressions were being measured nor did we include dummy electrodes that might distract the participants from guessing which muscles we targeted (Fridlund & Cacioppo, 1986). This may have affected our participants either consciously or unconsciously in the responsiveness of their facial expressions. In any case, we also had to be somewhat cautious in comparing our current results to those of previous studies due to some fundamental differences in data preparation and analyses. For instance, we selected the same 1.5-s of baseline activation for all trials of all participants rather than adjusting the selection based on visual inspection of the signal, which may consequently have affected our computed average corrugator responses that were reported as percentages compared to this baseline.

As for our stimuli, there was no obvious or direct link between a character's (im)moral behavior in first part of the narrative and the character's affect and its reason later on, i.e., Mark does not necessarily deserve a car damaged by a hailstorm because he behaved poorly towards someone earlier in the story. Some participants also remarked this in response to our open questions of our questionnaires and said that at times they experienced difficulties in deciding how fair the outcome of a story was because of this. One participant even rated all endings as neutral during the Fairness Task.

Lastly, there may still be other possible factors such as mimicry, sometimes also referred to as emotional contagion that modulate the corrugator response in processing written affective language as mentally simulating a situation of a character experiencing emotion might be able to elicit such vivid imagery that it prompts readers to mimic the corresponding facial expressions (Hess, Kappas, McHugo, Lanzetta, & Kleck, 1992; Hess & Fischer, 2013; Hess, 2021).

All in all, while our study has found some interesting results, studying processes involved in affective language processing is complex and there are many open questions that remain to be explored.

Acknowledgements

I would like to express my gratitude to Jos van Berkum and Marijn Struiksma for giving me the opportunity and guidance to conduct this fEMG study. Additionally, would like to thank Marijke Beulen for being the fourth team member for this project and offering additional guidance and support.

References

't Hart, B., Struiksma, M. E., van Boxtel, A., & van Berkum, Jos J. A. (2019). Tracking affective language comprehension: Simulating and evaluating character affect in morally loaded narratives. *Frontiers in Psychology, 10* doi:10.3389/fpsyg.2019.00318

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology, 16*(18), 1818-1823.

Barrett, L. F., Niedenthal, P. M., & Winkielman, P. (2007). *Emotion and consciousness* Guilford Press.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617-645.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01

Fino, E., Menegatti, M., Avenanti, A., & Rubini, M. (2016). Enjoying vs. smiling: Facial muscular activation in response to emotional language. *Biological Psychology, 118*, 126-135. doi:10.1016/j.biopsycho.2016.04.069

Foroni, F., & Semin, G. R. (2009). Language that puts you in touch with your bodily feelings: The multimodal responsiveness of affective expressions. *Psychological Science, 20*(8), 974-980.

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology, 23*(5), 567-589.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*(2), 301-307.

Hess, U. (2021). Who to whom and why: The social nature of emotional mimicry. *Psychophysiology, 58*(1), e13675. doi:10.1111/psyp.13675

Hess, U., & Fischer, A. (2013). Emotional mimicry as social regulation. *Personality and Social Psychology Review, 17*(2), 142-157.

Hess, U., Kappas, A., McHugo, G. J., Lanzetta, J. T., & Kleck, R. E. (1992). The facilitative effect of facial expression on the self-generation of emotion. *International Journal of Psychophysiology, 12*(3), 251-265.

Hinojosa, J. A., Moreno, E. M., & Ferré, P. (2020). Affective neurolinguistics: Towards a framework for reconciling language and emotion. *Language, Cognition and Neuroscience, 35*(7), 813-839.

Hoeken, H., & Sinkeldam, J. (2014). The role of identification and perception of just outcome in evoking emotions in narrative persuasion. *Journal of Communication, 64*(5), 935-955.

Izard, C. E. (2009). Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annual Review of Psychology, 60*, 1-25.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1-26. doi:10.18637/jss.v082.i13

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. doi:10.1111/1469-8986.00078

Lenth, R. V. (2022). Estimated marginal means, aka least-squares means. Retrieved from https://CRAN.R-project.org/package=emmeans

Majid, A. (2012). Current emotion research in the language sciences. *Emotion Review, 4*(4), 432-443.

Niedenthal, P. M., Winkielman, P., Mondillon, L., & Vermeulen, N. (2009). Embodiment of emotion concepts. *Journal of Personality and Social Psychology, 96*(6), 1120.

Ouwerkerk, J. W., Van Dijk, W. W., Vonkeman, C. C., & Spears, R. (2018). When we enjoy bad news about other groups: A social identity approach to out-group schadenfreude. *Group Processes & Intergroup Relations, 21*(1), 214-232.

R Core Team. (2022). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org

Smith, R. H., & van Dijk, W. W. (2018). Schadenfreude and gluckschmerz. *Emotion Review, 10*(4), 293-304.

Szanto, T. (2015). Collective emotions, normativity, and empathy: A steinian account. *Human Studies, 38*(4), 503-527.

't Hart, B., Struiksma, M. E., van Boxtel, A., & van Berkum, Jos J. A. (2018). Emotion in stories: Facial EMG evidence for both mental simulation and moral evaluation. *Frontiers in Psychology, 9* doi:10.3389/fpsyg.2018.00613

't Hart, B., Struiksma, M., van Boxtel, A., & van Berkum, Jos J. A. (2021). Reading about us and them: Moral but no minimal group effects on language-induced emotion. *Frontiers in Communication, 6* doi:10.3389/fcomm.2021.590077

Tan, J., Walter, S., Scheck, A., Hrabal, D., Hoffmann, H., Kessler, H., & Traue, H. C. (2012). Repeatability of facial electromyography (EMG) activity over corrugator supercilii and zygomaticus major on differentiating various emotions. *Journal of Ambient Intelligence and Humanized Computing, 3*(1), 3-10.

van Berkum, J., Struiksma, M., & 't Hart, B. (2020). *Using facial EMG to track emotion during language comprehension: Past, present, and future.* Unpublished manuscript. Retrieved Nov 12, 2021,

Van Boxtel, A.Facial EMG as a tool for inferring affective states. Paper presented at the *Proceedings of Measuring Behavior, , 7* 104-108.

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia, 49*(5), 1105-1127.

Winkielman, P., Niedenthal, P., Wielgosz, J., Eelen, J., & Kavanagh, L. C. (2015). Embodiment of cognition and emotion.

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review, 23*(4), 1028-1034.

Supplementary Materials

Model summaries

intercept exploratory

Model	-2 LL	nr of parameters	p model fit (chi-square distribution)	Model comparison	Predictor (added)
EMG Morality segment					
0	42404	4			Random Subject and Random Item
1	42296	5	0,0000	better	Fixed factor Morality Condition
EMG Adjective segment					
0	38416	4			Random Subject and Random Item
1	38380	7	0,0000	better	Interaction between Morality Condition and Adjective Valence
EMG Reason segment					
0	39170	4			Random Subject and Random Item
1	39088	7	0,0000	better	Interaction between Morality Condition and Adjective Valence
Story Fairness Task rating					
1	10059,238	7,0000	0,0000		Interaction between Morality Condition and Adjective Valence with Random Subject and Random Item
EMG Morality (after median split)					
1	39088	5			Fixed factor Morality Condition with Random Subject and Random Item
2	39088	7	0,0000	better	Interaction between Morality Condition and Justice score (continuous)
3	38994	7	0,00013		Interaction between Morality Condition and Justice group (high/low)
EMG Adjective (after median split)					
1	35470	7			Interaction between Morality Condition and Adjective Valence with Random Subject and Random Item
2	35456	11	0,00731	better	Interaction between Morality Condition, Adjective Valence, and Justice score (continuous)
3	35458	11	0,0138		Interaction between Morality Condition, Adjective Valence, and Justice group (high/low)
EMG Reason (after median split)					
1	34712	7			Interaction between Morality Condition and Adjective Valence with Random Subject and Random Item
2	34674	11	0,00000	better	Interaction between Morality Condition, Adjective Valence, and Justice score (continuous)
3	34702	11	0,03879		Interaction between Morality Condition, Adjective Valence, and Justice group (high/low)

Pairwise Comparisons State Adjective Segment

Emmeans

MoralityCondition	StateAdjectiveCondition	emmean	SE	df	lower.CL	upper.CL
Moral	Positive	104.69	7.03	72.81	90.68	118.69
Immoral	Positive	113.54	7.03	72.81	99.54	127.54
Moral	Negative	128.29	7.03	72.81	114.29	142.30
Immoral	Negative	122.30	7.03	72.84	108.29	136.30

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

Emmeans

StateAdjectiveCondition	MoralityCondition	contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
Positive	.	Moral - Immoral	-8.85	4.18	3153.48	-19.27	1.56	-2.12	0.13
Negative	.	Moral - Immoral	6.00	4.18	3153.54	-4.42	16.41	1.43	0.48
.	Moral	Positive - Negative	-23.61	4.18	3153.48	-34.02	-13.19	-5.65	0.00
.	Immoral	Positive - Negative	-8.76	4.18	3153.44	-19.18	1.66	-2.10	0.14

Pairwise Comparisons Affect Reason Segment

Emmeans

MoralityCondition	StateAdjectiveCondition	emmean	SE	df	lower.CL	upper.CL
Moral	Positive	99.24	6.7	85.49	85.92	112.56
Immoral	Positive	114.52	6.7	85.49	101.20	127.85
Moral	Negative	141.51	6.7	85.49	128.19	154.83
Immoral	Negative	115.76	6.7	85.54	102.43	129.08

Degrees-of-freedom method: asymptotic

Confidence level used: 0.95

Emmeans

StateAdjectiveCondition	MoralityCondition	contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
Positive	.	Moral - Immoral	-15.28	4.67	3153.27	-26.91	-3.65	-3.27	0
Negative	.	Moral - Immoral	25.75	4.67	3153.32	14.11	37.39	5.52	0
.	Moral	Positive - Negative	-42.27	4.67	3153.27	-53.90	-30.63	-9.06	0
.	Immoral	Positive - Negative	-1.23	4.67	3153.23	-12.87	10.40	-0.26	1

Pairwise Comparisons Story Fairness Task ratings

Emmeans

MORALITYCONDITION	STATEADJECTIVECONDITION	EMMEAN	SE	DF	LOWER.CL	UPPER.CL
Moral	Positive	1.95	0.05	254.96	1.85	2.05
Immoral	Positive	-1.27	0.05	253.26	-1.38	-1.17
Moral	Negative	-1.37	0.05	250.46	-1.47	-1.26
Immoral	Negative	1.53	0.05	253.48	1.43	1.63

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

Emmeans

StateAdjectiveCondition	MoralityCondition	contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
Positive	.	Moral - Immoral	3.22	0.06	3016.90	3.07	3.38	52.48	<.0001
Negative	.	Moral - Immoral	-2.90	0.06	3013.84	-3.05	-2.74	-47.35	<.0001
.	Moral	Positive - Negative	3.31	0.06	3008.81	3.16	3.47	54.11	<.0001
.	Immoral	Positive - Negative	-2.81	0.06	3006.79	-2.96	-2.65	-45.77	<.0001

Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

Conf-level adjustment: sidak method for 4 estimates

P value adjustment: sidak method for 4 tests

Pairwise Comparisons Morality Segment per Justice Group (after median split)

Emmeans

MoralityCondition	J_group	emmean	SE	df	lower.CL	upper.CL
Moral	0	108.45	15.48	58.01	77.45	139.44
Immoral	0	143.00	15.48	58.01	112.01	174.00
Moral	1	102.67	16.49	57.82	69.66	135.68
Immoral	1	184.31	16.49	57.84	151.30	217.32

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

Emmeans

J_group	MoralityCondition	contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
0	.	Moral - Immoral	-34.56	7.73	2924.93	-53.83	-15.28	-4.47	0.00
1	.	Moral - Immoral	-81.64	8.26	2944.35	-102.23	-61.05	-9.88	0.00
.	Moral	J_group0 - J_group1	5.78	22.43	56.22	-51.95	63.50	0.26	1.00
.	Immoral	J_group0 - J_group1	-41.31	22.43	56.23	-99.04	16.42	-1.84	0.25

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Conf-level adjustment: sidak method for 4 estimates
P value adjustment: sidak method for 4 tests

Pairwise Comparisons State Adjective Segment per Justice Group (after median split)

Emmeans

MoralityCondition	StateAdjectiveCondition	J_group	emmean	SE	df	lower.CL	upper.CL
Moral	Positive	0	104.72	10.26	66.72	84.25	125.20
Immoral	Positive	0	107.60	10.26	66.72	87.13	128.08
Moral	Negative	0	114.46	10.26	66.72	93.98	134.93
Immoral	Negative	0	112.51	10.26	66.72	92.04	132.98
Moral	Positive	1	106.36	10.93	66.64	84.54	128.18
Immoral	Positive	1	120.57	10.93	66.64	98.75	142.38
Moral	Negative	1	144.37	10.93	66.64	122.56	166.19
Immoral	Negative	1	134.02	10.93	66.71	112.20	155.85

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95

Emmeans

StateAdjectiveCondition	J_group	MoralityCondition	contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
Positive	0	.	Moral - Immoral	-2.877	6.074	2932.645	-20.249	14.496	-0.474	1.000
Negative	0	.	Moral - Immoral	1.944	6.074	2932.651	-15.428	19.316	0.320	1.000
Positive	1	.	Moral - Immoral	-14.206	6.482	2953.496	-32.745	4.334	-2.191	0.293
Negative	1	.	Moral - Immoral	10.350	6.487	2953.110	-8.203	28.902	1.596	0.755
.	0	Moral	Positive - Negative	-9.731	6.074	2932.576	-27.103	7.642	-1.602	0.751
.	0	Immoral	Positive - Negative	-4.910	6.074	2932.582	-22.282	12.462	-0.808	0.999
.	1	Moral	Positive - Negative	-38.014	6.477	2938.922	-56.538	-19.490	-5.869	0.000

.	1	Immoral	Positive - Negative	-13.458	6.481	2938.512	-31.996	5.079	-2.076	0.371
Positive	.	Moral	J_group0 - J_group1	-1.636	14.940	65.919	-45.866	42.594	-0.110	1.000
Positive	.	Immoral	J_group0 - J_group1	-12.965	14.940	65.919	-57.195	31.265	-0.868	0.997
Negative	.	Moral	J_group0 - J_group1	-29.919	14.940	65.919	-74.149	14.311	-2.003	0.455
Negative	.	Immoral	J_group0 - J_group1	-21.514	14.942	65.955	-65.748	22.721	-1.440	0.867

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Conf-level adjustment: sidak method for 12 estimates
P value adjustment: sidak method for 12 tests

Pairwise Comparisons Affect Reason Segment per Justice Group (after median split)

Emmeans - mor x val x Justice group

MoralityCondition	StateAdjectiveCondition	J_group	emmean	SE	df	lower.CL	upper.CL
Moral	Positive	0	101.05	10.41	74.07	80.30	121.80
Immoral	Positive	0	106.69	10.41	74.07	85.94	127.44
Moral	Negative	0	130.44	10.41	74.07	109.69	151.18
Immoral	Negative	0	109.05	10.41	74.07	88.30	129.80
Moral	Positive	1	97.81	10.64	73.96	76.61	119.02
Immoral	Positive	1	123.92	10.64	73.96	102.72	145.13
Moral	Negative	1	157.50	10.64	73.96	136.30	178.71
Immoral	Negative	1	124.28	10.65	74.06	103.07	145.49

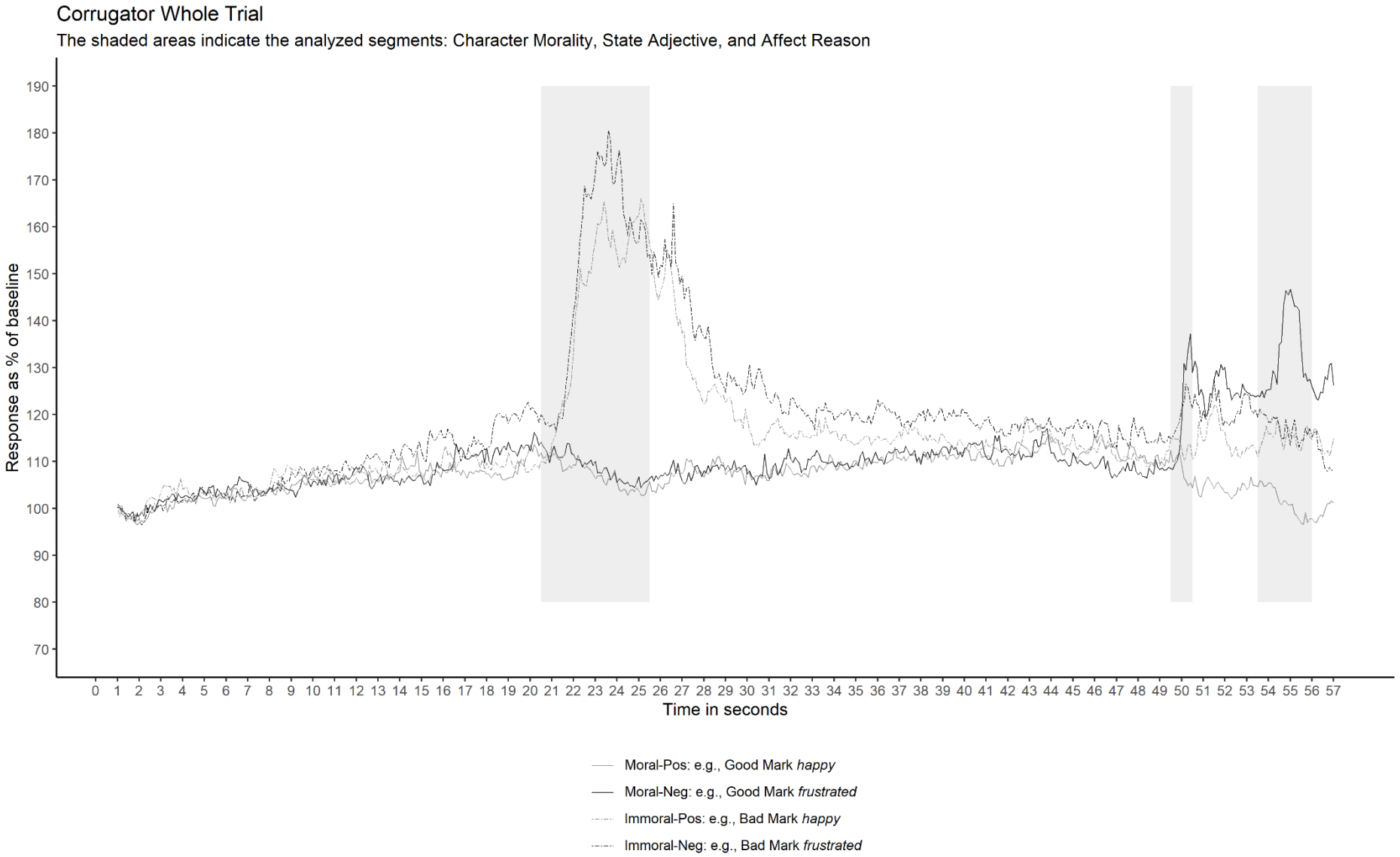
Degrees-of-freedom method: kenward-roger

Confidence level used: 0.95

StateAdjectiveCondition	J_group	MoralityCondition	contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
Positive	0	.	Moral - Immoral	-5.64	7.24	2822.05	-26.35	15.07	-0.78	1.00
Negative	0	.	Moral - Immoral	21.39	7.24	2822.06	0.68	42.09	2.95	0.04
Positive	1	.	Moral - Immoral	-26.11	7.40	2821.45	-47.28	-4.94	-3.53	0.01
Negative	1	.	Moral - Immoral	33.22	7.41	2821.06	12.03	54.41	4.48	0.00
.	0	Moral	Positive - Negative	-29.39	7.23	2803.79	-50.06	-8.71	-4.07	0.00
.	0	Immoral	Positive - Negative	-2.36	7.23	2803.80	-23.03	18.31	-0.33	1.00
.	1	Moral	Positive - Negative	-59.69	7.39	2807.54	-80.83	-38.54	-8.07	0.00
.	1	Immoral	Positive - Negative	-0.36	7.40	2807.18	-21.52	20.80	-0.05	1.00

Positive	.	Moral	J_group0 - J_group1	3.24	14.76	71.84	-40.33	46.81	0.22	1.00
Positive	.	Immoral	J_group0 - J_group1	-17.23	14.76	71.84	-60.80	26.34	-1.17	0.97
Negative	.	Moral	J_group0 - J_group1	-27.06	14.76	71.84	-70.64	16.51	-1.83	0.59
Negative	.	Immoral	J_group0 - J_group1	-15.23	14.76	71.89	-58.81	28.35	-1.03	0.99

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
Conf-level adjustment: sidak method for 12 estimates
P value adjustment: sidak method for 12 tests


Corrugator Whole Trial
The shaded areas indicate the analyzed segments: Character Morality, State Adjective, and Affect Reason
Response as % of baseline
190
180
170
160
150
140
130
120
110
100
90
80
70
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57
Time in seconds
Moral-Pos: e.g., Good Mark *happy*
Moral-Neg: e.g., Good Mark *frustrated*
Immoral-Pos: e.g., Bad Mark *happy*
Immoral-Neg: e.g., Bad Mark *frustrated*

UiL OTS
UTRECHT INSTITUTE
OF LINGUISTICS

Block 1

0.
Nu volgt een korte vragenlijst over jouw perspectief op eerlijkheid en juistheid.

Proefpersoonnummer (dient te worden ingvuld door de experimentleider):

Block2

1. Wat nam je mee in je beoordeling toen je moest aangeven hoe eerlijk je de afloop van de verhaaltjes vond? Waar let je op? Schrijf je antwoord in een paar zinnen (max. 5).

2. Heb je nog andere opmerkingen over de karakters en gebeurtenissen in de verhaaltjes?

Block 3

.. Geef aan in hoeverre je het eens bent met de volgende algemene uitspraken. Dit laatste deel staat dus los van de verhaaltjes die je hebt lezen.

1. Ik kan medelijden hebben met goede mensen.

Helemaal waar				Helemaal niet waar
○	○	○	○	○

2. Ik kan medeleven hebben met slechte mensen.

Helemaal waar				Helemaal niet waar
○	○	○	○	○

3. Ik vind het minder vervelend als een slecht persoon iets slechts overkomt dan een goed persoon.

Helemaal waar ○ ○ ○ ○ ○ Helemaal niet waar

4. Ik vind het minder fijn als een slecht persoon iets goeds overkomt dan een goed persoon.

Helemaal waar ○ ○ ○ ○ ○ Helemaal niet waar

5. Als iemand zich slecht gedraagt dan is het diens verdiende loon als anderen zich ook slecht gedragen tegen die persoon.

Helemaal waar ○ ○ ○ ○ ○ Helemaal niet waar

6. Als iemand zich goed gedraagt dan heeft diegene het verdiend om goed te worden behandeld door anderen.

Helemaal waar ○ ○ ○ ○ ○ Helemaal niet waar

7. Boontje komt om z'n loontje, je krijgt wat je verdient.

Helemaal waar

Helemaal niet waar

8. Ik vind het oneerlijk als een slecht persoon iets goeds overkomt.

Helemaal waar Helemaal niet waar

9. Ik vind het eerlijk als een slecht persoon iets slechts overkomt.

Helemaal waar Helemaal niet waar

10. Ik vind het oneerlijk als slechte mensen niet bestraft worden.

Helemaal waar Helemaal niet waar

11. Ik vind het oneerlijk als goede mensen niet beloond worden.

Helemaal waar Helemaal niet waar

12. Té zware straffen zijn oneerlijk.

Helemaal waar Helemaal niet waar

13. Ik vind het oneerlijk als slechte mensen te veel gestraft worden.

Helemaal waar ○ ○ ○ ○ ○ Helemaal niet waar

14. Als Mark met opzet Maarten slaat dan verdient hij misschien een klap terug van Maarten, maar **niet** van een ander persoon (die niks van Marks gedrag weet).

Helemaal waar ○ ○ ○ ○ ○ Helemaal niet waar

15. Ik kijk er meer van op als een goed mens iets slechts overkomt dan een slecht mens.

Helemaal waar ○ ○ ○ ○ ○ Helemaal niet waar

16. Ik verwacht over het algemeen dat een goed persoon vaker iets goeds overkomt dan een slecht persoon.

Helemaal waar ○ ○ ○ ○ ○ Helemaal niet waar

17. Mensen die zich fatsoenlijk gedragen verdienen geluk en voorspoed.

Helemaal waar				Helemaal niet waar
○	○	○	○	○

Block 2

EIND.

Deelname aan dit onderdeel opslaan?

- ○ Ja
- ○ Nee

Verhalen en emotie 4: De buurman is een klootzak