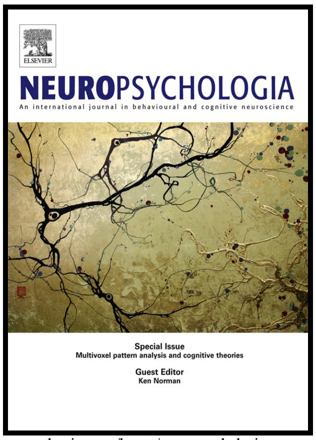
Author's Accepted Manuscript



www.elsevier.com/locate/neuropsychologia

Making sense of social interaction: Emotional coherence drives semantic integration as assessed by event-related potentials

Gesche Schauenburg, Markus Conrad, Christian von Scheve, Horacio A. Barber, Jens Ambrasat, Arash Aryani, Tobias Schröder

PII: S0028-3932(19)30001-6
DOI: https://doi.org/10.1016/j.neuropsychologia.2019.01.002
Reference: NSY6993

To appear in: *Neuropsychologia*

Received date: 21 February 2018
Revised date: 18 November 2018
Accepted date: 3 January 2019

Cite this article as: Gesche Schauenburg, Markus Conrad, Christian von Scheve, Horacio A. Barber, Jens Ambrasat, Arash Aryani and Tobias Schröder, Making sense of social interaction: Emotional coherence drives semantic integration as assessed by event-related potentials, *Neuropsychologia,* https://doi.org/10.1016/j.neuropsychologia.2019.01.002

Making sense of social interaction: Emotional coherence drives semantic integration as assessed by event-related potentials

Gesche Schauenburg[1], Markus Conrad[2], Christian von Scheve[3], Horacio A. Barber[4], Jens Ambrasat[3], Arash Aryani[1], & Tobias Schröder[5]

Author Note

Authors' affiliations: [1]Freie Universität Berlin, Department of Psychology, Berlin, Germany;

[2]Universidad de La Laguna, Department of Cognitive, Social and Organizational Psychology, San Cristobal de La Laguna, Spain;

[3]Freie Universität Berlin, Department of Sociology, Berlin, Germany;

[4]Instituto de Tecnologías Biomédicas, Universidad de La Laguna, & Basque Center on Cognition, Brain and Language (BCBL), Spain

[5]University of Applied Sciences, Potsdam, Germany.

.

This research was funded by a grant from the Cluster of Excellence "Languages of Emotion" at Freie Universität Berlin to C.v.S, T.S. and M.C. (project #411) as part of the Excellence Initiative of the German Research Foundation (DFG).

Declaration of interest: none.

Correspondence concerning this paper should be addressed to Gesche Schauenburg, email: gesche.schauenburg@gmail.com; phone: +49 (0)176 21540485

Abstract

We compared event-related potentials during sentence reading, using impression formation equations of a model of affective coherence, to investigate the role of affective content processing during meaning making. The model of Affect Control Theory (ACT; Heise, 1979, 2007) predicts and quantifies the degree to which social interactions deflect from prevailing social norms and values - based on the affective meanings of involved concepts. We tested whether this model can predict the amplitude of brain waves traditionally associated with semantic processing. To this end, we visually presented sentences describing basic subject-verb-object social interactions and measured event-related potentials for final words of sentences from three different conditions of affective deflection (low, medium, high) as computed by a variant of the ACT model (Schröder, 2011). Sentence stimuli were closely controlled across conditions for alternate semantic dimensions such as contextual constraints, cloze probabilities, co-occurrences of subject-object and verb-object relations. Personality characteristics (schizotypy, Big Five) were assessed to account for individual differences, assumed to influence emotion-language interactions in information processing. Affective deflection provoked increased negativity of ERP waves during the P2/N2 and N400 components. Our data suggest that affective incoherence is perceived as conflicting information interfering with early semantic processing and that increased respective processing demands – in particular in the case of medium violations of social norms - linger on until the N400 time window classically associated with the integration of concepts into embedding context. We conclude from these results that affective meanings influence basic stages of meaning making.

Making sense of social interaction: Emotional coherence drives semantic integration as assessed by event-related potentials

1. Introduction

As social beings, humans use language to communicate and represent socially relevant information. Significant communicative content is not only transferred via denotative meanings of concepts, but also via their affective meanings. Influential theoretical claims from the field of psychology posit an intimate link between general semantics and affective dimensions: Attempts to measure and describe the semantic meaning of a large amount of different concepts through a small set of scales via semantic differential ratings usually use the following canonical affective dimensions of *evaluation* (or *valence*), which refers to the hedonic dimension of pleasantness vs. unpleasantness, *potency* (or *control*), which characterizes the range of strength vs. weakness, and *activity* (or *arousal*), which specifies the extent of excitement vs. calmness (Osgood, Suci, & Tannenbaum, 1957; Osgood, May, & Miron, 1975). While initially thought to represent basic dimensions of meaning in general (Osgood et al., 1957), these measures later have become well-established in the research representing the dimensional view of emotion (e.g. Osgood et al., 1957; Wundt, 1896; Russel & Mehrhabian, 1977; Bradley, Greenwald, Petry, & Lang, 1992; Feldman Barrett & Russel, 1998); though different studies sometimes use different labels for these emotion dimensions (as indicated in parentheses above). Scholl (2013) even claims that these three emotion dimensions represent the fundamental socio-emotional basis of human communication. If this claim holds true, a close interconnection between neural processing of emotion and language as the primary tool for communication must be expected (see Koelsch et al., 2015, for a theoretical proposal).

1.1. Empirical effects of affective content on single word reading

Numerous studies from psycholinguistics and neuroscience show how words' affective content influences language processing – already at automatic processing stages and before conscious access (e.g. Bernat, Bunce, & Shevrin, 2001; Fischler & Bradley, 2006). For visual word recognition, such effects have been shown with behavioral measures such as response latencies (e.g. Kousta, Vinson, & Vigliocco, 2009; Võ, Jacobs, & Conrad, 2006), or memory performance (e.g. Kensinger & Corkin, 2003; Doerksen & Shimamura, 2001), with physiological measures such as pupil dilation (Kuchinke, Võ, Hofmann, & Jacobs, 2007; Võ et al., 2008), and eye fixations (Scott, O'Donnel, & Sereno, 2012), or with neural correlates of language processing using event-related potentials (ERP, e.g., Conrad, Recio, & Jacobs, 2011; Hofmann, Kuchinke, Tamm, Võ, & Jacobs, 2009; Kissler & Herbert, 2013; Recio, Conrad, Hansen & Jacobs, 2014; Schacht & Sommer, 2009; see Citron, 2012, and Jacobs et al., 2015, for reviews), transcranial magnetic stimulation (Weigand et al., 2013), or functional magnetic resonance imaging (e.g. Grimm, Weigand, Kazzer, Jacobs, & Bajbouj, 2012; Herbert et al., 2009; Kuchinke et al., 2005). Even in experimental tasks for which emotional aspects are per se irrelevant (such as the lexical decision task, e.g. Hofmann et al., 2009; or affective Simon task; e.g. Altarriba & Basnight-Brown, 2011) or where attention to emotional properties of stimuli interferes with efficient task resolution (emotional Stroop task: Sass et al., 2010; Malhi, Lagopoulos, Sachdev, Ivanovski, & Shnier, 2005), an influence of the affective content of visually presented single words was observed.

In every-day life, we encounter more complex linguistic structures than words in isolation. Hence, there has been an evolving research investigating how emotional word content in embedded language context influences language processing (e.g., Hsu et al., 2014; 2015a, b, c; Lüdtke & Jacobs, 2015).

1.2. ERP effects during sentence reading

Concerning sentence processing, the most studied component of the ERP signal is probably the N400 – proposed to reflect meaning activation and semantic integration processes (Kutas & Hillyard, 1980; Kutas & Federmeier, 2011). The amplitude of this negative-going brainwave peaking about 400ms after critical word onset is inversely correlated with the ease to integrate a stimulus in a given context – or with generally increasing processing demands (see Barber & Kutas, 2007, for a review). Accordingly, N400 amplitudes were shown to increase with semantic violations (e.g., Kutas & Hillyard, 1980; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Molinaro, Conrad, Barber, & Carreiras, 2010), expectancy violations (Kutas & Hillyard, 1984), world-knowledge violations (Hagoort, Hald, Bastiaansen, & Petersson, 2004), or discourse-induced expectation violations (Nieuwland, & Van Berkum, 2006a). Concerning affective language processing, studies using single word presentation consistently showed behavioral processing advantages for affective word content, typically accompanied by two ERP components: an early posterior negativity (EPN) – assumed to reflect attention allocation to emotionally relevant stimuli - and a late positive potential (LPC) – assumed to reflect more elaborate semantic processing of emotion-laden words – (see Citron, 2012, for a review). ERP studies using emotion-laden words embedded in sentences provided somewhat mixed results: Results from Holt, Lynn, & Kuperberg (2009), Martin-Loeches et al. (2012), as well as Delaney-Busch and Kuperberg (2013) converge on persisting LPC effects for emotion-laden words also in more natural sentence reading paradigms. But whereas the two latter studies observed attenuated N400 effects for emotion-laden words – potentially mirroring the behavioral processing advantage for such words in terms of more easy integration into sentence context, Holt, Lynn, & Kuperberg (2009), on the other hand, report larger N400 for emotion-laden as compared to neutral target words.

As emotion-laden words can be embedded into sentence contexts in very different ways, such heterogeneous pattern of results may be not so surprising after all. To investigate the role of affective language content during meaning making, it appears most pertinent to focus on how the affective content of words influences the way we integrate their meaning into a specific affective context. León, Díaz, de Vega, and Hernández (2010) reported an N400 and an early N100/P200 effect when participants read a target word within a sentence describing a protagonist's emotional state that was inconsistent with a preceding story. While their data show that typical N400 effects for violations of expectations or world-knowledge (Hagoort et al., 2004) extend to situations where such expectations relate to emotional states, we would like to raise the question of whether ongoing online affective evaluation of concepts and their respective consistency or congruency represents a mandatory feature of sentence processing. Empirical results, again, sum up to a heterogeneous pattern: For instance, Martín-Loeches et al. (2012) found no modulation of the N400 semantic anomaly effect depending on target word valence, thus, affective content seems not to interfere with general semantic processing. Delaney-Busch and Kuperberg (2013) focused more directly on congruency of affective valence between critical words across two different sentences: N400 effects were given only for semantic incongruency between neutral target words and preceding neutral context, no such effects were obtained for emotional incongruency. Apparently, the affective salience of an emotional word might override (Wang, Bastiaansen, Yang, and Hagoort; 2013) or overleap (Delany-Busch & Kuperberg, 2013) potential effects of affective congruency in the N400 time window. In a similar vein, Wang, Bastiaansen, and Yang (2015) found that affective content influences the time course of incongruency effects, i.e., emotion incongruent verbs following positive person names elicited a N400 effect, whereas incongruent verbs following negative names elicited no N400 but a P600 effect. Since the processing of emotional words requires the interaction of linguistic and emotional systems, contradictory responses from both of them would activate executive resources related to conflict detection and resolution. Previous data have demonstrated the influence of the emotional valence of words on conflict processing. These studies have focused on the modulation of the N200 component, which is associated to conflict detection processes and consistently reported using the flanker task (Eriksen & Eriksen, 1974). The N200 effect obtained with the flanker task is enhanced with emotional words compared to neutral words, showing that emotional information (positive or negative) directly affects these conflict-monitoring processes (Kanske & Kotz, 2010; 2011). Therefore, we can hypothesize that the processing of conflicting emotional information during sentence reading could trigger the same mechanisms of those involved in the flanker task, leading to a similar ERP effect in the time range of the N200.

The heterogeneous pattern of results of the above mentioned studies on emotional congruency effects (Holt, Lynn, & Kuperberg, 2009; Martín-Loeches et al., 2012; Delaney-Busch & Kuperberg, 2013; León et al., 2010) may in part be due to the fact that they tapped into different aspects of emotion and/or syntactic processing realizing affective congruency also mainly across different sentences.

1.3. Affective structure of language reflects social norms and values

According to Affect Control Theory (ACT; Heise, 1979, 2007; MacKinnon, 1994; Schröder, Hoey, & Rogers, 2016; Smith-Lovin & Heise, 1988), people use the affective structure of language as a source of information to infer whether a given social situation is in line with prevailing social norms and expectations. Thus, this social psychological emotion theory capitalizes on the role of language: Woven into the affective meanings of words, socially and culturally shared knowledge and experiences are transferred and, thus, influence the way people build impressions of linguistically represented social situations, that is, how people construct the meaning, understanding, regulation and further course of social events (Rogers, Schröder, & von Scheve, 2014). In this theoretical framework, normativity of social events is judged by the degree of perceived affective coherence transferred by the interplay of the affective meanings of the words used to describe a given social event. Prior studies have demonstrated the widespread consensus on ratings of affective word meanings across members of the same society – supporting the basic assumption of ACT that culturally shared knowledge is reflected in the affective structure of language (Ambrasat, von Scheve, Conrad, Schauenburg, & Schröder, 2014; Heise, 2010).

While affective meanings reflect the emotional connotations of concepts at the single-word level, the social psychological construct of affective coherence can be understood as "the mutual goodness of fit" (Schröder, 2011) of *all* the connotations of the concepts used to linguistically represent the relevant situation. Similar to other cognitive-consistency theories (e.g. Heider's Balance Theory, 1946), one basic assumption of ACT is that people strive to maintain the affective meanings of the concepts in their mental representations and actions. Thus, situations for which affective meanings of involved concepts match each other, easily integrate in our stream of perception and action; while we mentally stumble over events that are represented by concepts whose emotional connotations do not fit together. For instance, in the situation "A mother plays with a child", the affective meaning of the concept "mother" almost perfectly matches the emotional connotations of the other words "play" and "child". However, the emotional connotation of the concept mother may harmonize less with that of the concept "to beat somebody". Therefore, the situation "A mother beats a child" would strongly violate our general affective representation of the word mother - because of the incongruency between its common sentiment and its situational, transient affective meaning. This mismatch would encourage us to somehow "rebalance" the lack of perceived coherence – for instance by postulating that the child was badly misbehaving before (which of course is not an appropriate reason to beat a child, but in this way of representing the situation, e.g. as "A mother beats a naughty child", the initially perceived violation of affective coherence would not be that severe anymore). In the research framework of ACT, affective coherence can be modeled mathematically using impression-formation equations which were obtained in empirical studies by regressing the ratings of the evaluation, activity and potency dimensions of words in the context of a sample of given events on out-of-context semantic differential ratings of the same words (e.g., Averett & Heise, 1987; Schröder, 2011).

Behavioral and computational data provide empirical evidence for this mathematical model to accurately model social perception (Heise & MacKinnon, 1987; Schröder & Thagard, 2013). For a detailed description on how these impression equation formations were obtained in the German language see Schröder, 2011.

These ACT-based procedures simply aim at taking into account that the affective meaning of a given concept arises from the transient interplay of its fundamental sentiment (Heise, 2007) with its specific current context. Empirically validated impression-formation equations generated by ACT, thus, should provide a better or more comprehensive account of emotional congruency than, for instance, merely comparing maximized valence contrasts between word pairs. Because the coefficients of the equations capture "something about the normative process of impression formation in our culture" (Robinson, Smith-Lovin, & Wisecup, 2006, p. 185), the formal mathematical model of affective deflection, which corresponds to the squared Euclidean Distance of the fundamental and transient affective meanings, can be interpreted as a measure for the degree to which a given linguistically labeled social situation deviates from prevailing social norms and values (Heise, 2007). Hence, emotional congruency is not only determined by the interplay of the affective dimensions of all the words used to describe a given event, but at the same time reflects social norms and values of the mental representation of the relevant event.

Beyond the focus of language processing itself, the N400 paradigm has also become a common tool to investigate social cognition, particularly addressing the issue of social expectancy violations. For instance, in the case of stereotype processing, White, Crites Jr., Taylor, and Corral (2009) found larger N400 amplitudes for target words that were stereotypically incongruent with preceding gender words in a classic priming paradigm. Similar results concerning the N400 as an index for the accessibility of stereotypes were reported for social group stereotypes (Wang, Ma, Song, Shi, Wang, & Pfotenhauer, 2011), and racial stereotypes (Hehman, Volpert, & Simons, 2013). The N400 can also be used as a measurement for the conflict between individual and group judgments as has been shown by Huang, Kendrick, and Yu (2014) who reported larger N400 for the incongruency between participants' subjective ratings and group ratings on face attractiveness. Van Berkum, Holleman, Nieuwland, Otten, and Murre (2009) have demonstrated that the N400 is also sensitive to participants' individual value systems: They found that facing participants with moral statements disagreeing with their value systems elicited an N400 effect – along with an enhanced positivity in an early time window (200-250ms) and a more positive going deflection for value-inconsistent sentences in the late positive potential (500-650ms). These studies suggest that even at early stages, readers may already attempt to integrate the meaning of a sentence with cultural norms and values.

Furthermore, trying to reconcile these initial attempts on investigating emotion effects during sentence processing with the intriguing findings from the field of social cognition (White et al., 2009; Wang et al., 2011; Hehmanet al., 2013; Huang et al., 2014; Van Berkumet al., 2009) calls for a comprehensive theoretical framework to define and operationalize the affective coherence or congruency of sentences. The research question we wanted to tackle here by combining theoretical proposals from social psychology with a neuroscientific investigation of sentence reading is: How do readers continuously integrate affective information into meaning making during online sentence processing?

1.4. The present Study

We conducted the present study to explore on a neurophysiological level - using the high temporal resolution of EEG - at which point affective coherence influences the way we extract information from sentences. To this end, we recorded participants' EEG signals (focusing on ERPs to sentence final words) while they silently read semantically correct simple sentences (*Actor + Behavior + Object*) describing social interactions in three different conditions of deflections from affective coherence (low, medium, high), e.g., "*The schoolgirl admires the champion*" ("Das Schulmädchen bewundert den Champion", low deflection), vs.

"*The schoolgirl admires the rescuer*" ("Das Schulmädchen bewundert den Retter"; medium deflection), vs.

"*The schoolgirl admires the slacker*" ("Das Schulmädchen bewundert den Faulenzer", high deflection).

We applied the ACT-based mathematical model of impression formation to determine emotional congruency of sentences describing social events. We chose a silent reading paradigm to measure implicit processing of affective coherence – thus following ACT's basic assumption that people always use emotional connotations as a source for normativity judgments, not only when they are explicitly asked to do so.

Assuming that emotional consistency or affective coherence is a basic feature of semantic integration, the simplest hypothesis predicts a linear correlation between affective deflection and the N400 amplitude to sentence final words. Thus, N400 amplitude should increase systematically with the degree of emotional incongruency of sentences – determined by the relation between sentence final words and preceding context.

Moreover, as semantic context-integration effects in sentence reading were shown even at very short latencies (Penolazzi et al., 2007; Léon et al., 2010; Sereno, Brewer, & O'Donnell, 2003) one can expect ERP effects of affective coherence during sentence reading to arise in even earlier time windows. In particular, our manipulations of affective coherence deflection involve conflicting emotional connotations of words used to describe social interactions (see methods for details). Therefore, ERP effects may extend to the P2/N2 component as a correlate of conflict monitoring (Yeung, Botvinick, & Cohen, 2004; Folstein & Van Petten, 2008), with increased fronto-central negativity for affectively incoherent sentences.

Since we consider the present study as an interdisciplinary approach integrating scientific results related to language processing from the level of single words and sentence processing up to the level of social cognition, we also wanted to tackle individual differences concerning this subject at least in an exploratory mode. As personal traits such as schizotypy have been shown to modulate N400 amplitudes in language processing studies – probably due to a deficient use of context in integration processing – (Kutas & Federmeier, 2011; Kiang, Prugh, & Kutas, 2010, for a review on schizotypy and language, see Kiang, 2010) we administered the Schizotypy Personality Questionnaire (SPQ; Raine, 1991; German version: SPQ-G; Klein, Andresen, & Jahn, 1997) to investigate potential individual differences. With regard to the well-established personality-emotion relationships (Fossum & Barrett, 2000) on affective processing especially concerning the correlations of negative emotions with neuroticism and positive emotions with extraversion (e.g. McCrae & Costa, 1991) we assume that personality-associated emotion regulation might influence moral judgments (Athota, O'Connor, & Jackson, 2009) or correlate with evaluation (Fossum & Barrett, 2000). We therefore also measured the Big Five personality traits using the NEO Five Factor Inventory (NEO-FFI; Costa & McCrae, 1989; German version: Borkenau & Ostendorf, 1993) as a standard personality screening method to capture hints of potential personality effects on our ERP data.

2. Method

2.1. Participants

Forty-nine university students participated in this study. All participants were right-handed (Oldfield, 1971) German native speakers with normal or corrected-to-normal vision. Participation was monetarily rewarded (8 €/h). The data of two participants were excluded due to health issues and the data of further nine subjects were excluded due to bad signal-to-noise ratio (see EEG analyses). Thus, the data of thirty-eight subjects (mean age 25 years, range 20-31; 18 women) with no history of psychological or neurological diseases were analyzed. The whole experiment was designed and conducted according to the Declaration of Helsinki and was approved by the Ethics Council of the Freie Universität Berlin, Germany.

2.2. Materials and Design

Stimuli were 318 simple sentences describing social interactions in a basic actor-behavior-object manner, organized in three conditions of affective deflections (low, medium, high). All sentences had the following structure: determiner-subject-verb-(preposition[1])-determinant-object. Prior to final stimulus selection, we had generated a pool of sentences by permutating all possible sentence elements (654 subject words, 275 verbs, and 400 object words), for that affective meaning information on the three dimensions evaluation, arousal, and potency was available (The Berlin Affective Word List/BAWL, Võ et al., 2006; Affective Norms for German Sentiment/ANGST, Schmidtke et al., 2014; Ambrasat et al., 2014; Schauenburg, Ambrasat, Schröder, von Scheve, & Conrad, 2014) – assuring that object words consisted of no more than nine letters to avoid re-fixations on target words (Rayner, 1998). For the resulting approximately 72 Million sentences, we calculated emotional congruency in terms of affective deflection using regression equations for impression formations fitted for the German version (Schröder, 2011; see Appendix for equations). Based on their distribution we determined three conditions of affective deflection (low, medium, high) from which we selected our stimuli according to the following principles:

- Each "sentence context" formed by actor plus behavior had to provide very low contextual constraints and to be present in all conditions the same numbers of times.
- Each "target" (sentence-final object word) words had to represent a plausible but low cloze probability ending for a given sentence and to be present in all conditions the same numbers of times.

[1]Some German verbs demand a preposition between verb and object in transient sentences. In this stimuli set, some sentences included the preposition "mit" ("with") or "für" ("for).

Thus we ensured that our stimuli differed only in the emotional coherence of the whole sentences whereas all single elements (targets and preceding contexts) were identical across conditions (see Figure 1 for stimuli example) and that, furthermore, target words were neither predictable nor violating expectations generated by the context as both phenomena could have influenced the ERP components in undesired ways (Kutas & Hillyard, 1984; Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Federmeier et al., 2007).[2]

We determined cloze probabilities and contextual constraints (using the cloze method; Taylor, 1953) for our finally selected stimuli by presenting all sentence contexts to German native speakers (N = 41, mean age 28.65 years, range: 19-55; 26 women) who did not participate in the EEG-study. Participants were asked to write down the word coming first to their mind as an appropriate sentence ending. Overall, only twelve sentence contexts were ever completed by any participant with the preselected target word – but only in three cases respective identical responses were given by different participants (though each time N = 2). Thus, cloze probabilities were always very low and were controlled for across conditions (see Table 1). Overall contextual constraints were controlled for across conditions and also generally very low (see Table 1): On average, only five out of forty-one given responses to each sentence coincided. To further control for context dependent predictability of sentence-final words – as a function of merely cognitive rather than specifically emotional processes - we assessed and balanced semantic associations between sentence contexts and objects using subject-object and verb-object frequencies of co-occurrences in normal language based on the German corpus dlexDB (Heister et al., 2011).[3] In sum, the whole stimulus set included 318

[2] Strictly speaking, one could argue that it is impossible to fully separate affective/emotional content from cognitive/semantic content, as the conceptual distinction between cognition and emotion is likely only a phenomenological one that lacks neural substance (Duncan & Barrett, 2007; see also Thagard & Schröder, 2014). At the very least, our experimental procedures ensure that our empirical measure of affective coherence is not contaminated with facets of linguistic comprehension that have traditionally been interpreted as cognitive and/or semantic.

[3] Frequencies of co-occurrences of subject-object and of verb-object relations mutual information (MI) and t-scores (t) were calculated and kindly provided by Kay-Michael Würzner. One-way analyses of variances showed no differences between means of MI and t-scores across condition, i.e., across conditions, sentences did

sentences, 106 sentences for each of the three conditions of affective deflection. These were formed by 104 different sentence contexts (= *Actor* + *Behavior*) and 58 different target words (= *Object*). To realize a perfect control of both sentence context and target word identities for each participant, our design involves repetition of sentence contexts and targets in the following way: Across conditions, each participant saw each of 104 sentence contexts three times; two sentence contexts were used twice within each condition, i.e. six times in total. Concerning the total of 58 different target words, each participant saw each of 28 different target words three times (once per condition), while 18 target words were used twice, eight target words three, two target words four, and two other target words five times within each condition.

As illustrated in Figure 1, each sentence context and each object word entered each of the three conditions of affective deflection (low, medium, high) the same number of times; i.e., each condition included exactly the same sentence contexts on the one hand, and exactly the same object words on the other, sentences differed only in the specific combinations of those two elements. While our manipulation focuses on affective consistency of sentence, our selection procedure assures a perfect match of overall affective content across conditions. As we further closely controlled for co-occurrences and cloze probability across conditions, the 106 sentences of each condition were almost perfectly balanced concerning potential confounders.

<Figure 1 & Table 1 about here>

not differ with regard to their frequency of co-occurrences of subject-object and verb-object combinations (see Table 1).

Figure 1 Example of stimuli and schematic illustration of how all contexts and object words entered the condition of *low*, *medium*, and *high affective deflection*.

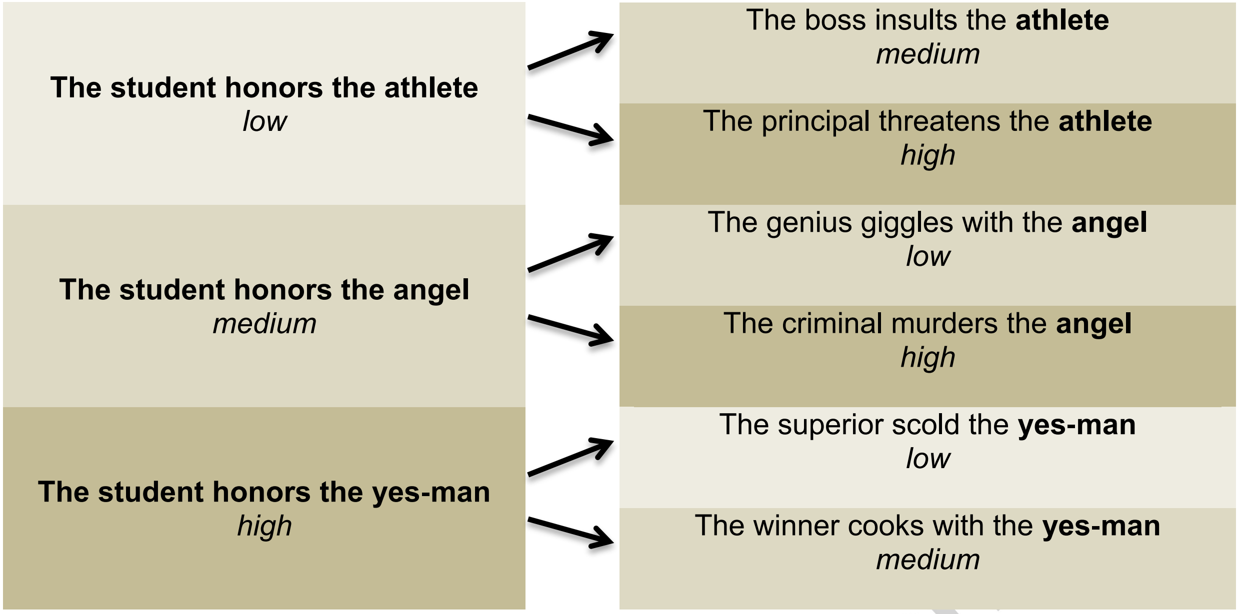


Table 1
Overview of stimulus characteristics and comparability across conditions.

	Condition of Affective Deflection			
	Low	Medium	High	Significance
Affective deflection	2.98	4.72	6.52	$F = 279.863$, $p < .0001$
	(0.10)	(0.11)	(0.10)	
N Targets	58	58	58	
N Contexts	104	104	104	
N Sentences	106	106	106	
Constraints	12.20 %	12.20 %	12.20 %	
Cloze probability	.1 %	.3%	.3%	
MI-SO	11.72	11.60	11.71	$F = .157$, $p = .855$
	(2.27)	(1.99)	(2.50)	
T-SO	9.11	8.30	7.23	$F = .087$, $p = .917$
	(33.46)	(20.10)	(17.05)	
MI-VO	11.33	11.30	11.44	$F = .145$, $p = .866$
	(1.94)	(1.75)	(2.10)	
T-VO	6.54	6.55	5.24	$F = .201$, $p = .818$
	(9.33)	(10.33)	(8.37)	

Means, standard deviations (in parentheses), and ANOVA results (F- and p-value) are shown. Affective deflection was calculated by impression formation equations fitted for the German language (Schröder, 2011). Constraints and Cloze probability are reported as weighted percentages (raters: N=41). MI-SO and T-SO report Mutual Information and smoothed t scores for subject-object frequencies of co-occurrences. MI-VO and T-VO report Mutual Information and smoothed t scores for verb-object frequencies of co-occurrences.

2.3. Personality Questionnaires

Participants completed the Schizotypy Personality Questionnaire (SPQ; Raine, 1991; German version: Klein et al., 1997) and the NEO Five Factor Inventory (NEO-FFI; Costa & McCrae, 1989; German version: Borkenau & Ostendorf, 1993). Both questionnaires are self-report scales. The SPQ assesses schizotypal personality based on the DSM-III-R criteria for schizotypal personality disorder (SPD; American Psychiatric Association, 1987). It includes 74 binary items (yes/no; "yes" coded as "1" and "no" coded as "0") encompassing nine subscales which refer to the nine different DSM-III-R schizotypal traits: Ideas of Reference (IR, 9 items), Social Anxiety (SA, 8 items), Magical Thinking (MT, 7 items), Unusual Perceptual Experience (UPE, 9 items), Eccentric Behavior (EB, 7 items), No Close Friends (NCF, 9 items), Odd Speech (OS, 9 items), Constricted Affect (CA, 8 items), and Suspiciousness (S, 8 items). The SPQ has a high internal ($\alpha = .88$) and high retest reliability ($r = .88$). Principal component analysis for the German version by Klein et al. (1997) yielded a two-factor solution of which both factors are comparable to the first two factors from the original version (Raine et al., 1991, found a three-factor solution; two-factor solution for the German version was also recommended by Dillmann, 2003, due to inconsistencies regarding the three-factor solution): the cognitive-perceptual factor (subscales: S, OS, UPE, EB, IR, MT) which is associated with "positive schizotypy" and the interpersonal factor (subscales: NCF, CA, SA) which is associated with "negative schizotypy". Factor scores were calculated by adding respective subscales' scores.

The NEO-FFI includes 60 items (5-point rating scale ranging from "strong disagreement" coded as "0" to "strong agreement" coded as "4") assessing five personality dimensions (12 items each): Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). NEO-FFI's subscales have good to high internal consistency ($\alpha = .72$ - $\alpha = .87$) and good to high retest reliabilities ($r = .71$ – $r = .82$).

2.4. Procedure

Before the experiment started, participants signed the written informed consent, completed questionnaires for demographic data and both psychometric questionnaires and were prepared for EEG recording. Participants were seated 80 cm in front of 17” computer monitor in a dimmed, electrically shielded and sound-attenuated room. They were asked to move as little as possible and to inhibit eye movements during sentence reading to prevent muscle artefacts in the EEG signal. Sentences were presented in a word-by-word manner using “Arial” font, size 30, in white letters on a black background in the center of the computer screen. Each trial started with a 500ms blank screen followed by a 500ms fixation cross announcing the start of the next sentence. Each word was presented for 250ms with a 450ms blank screen in between. After target word presentation a 1500ms blank screen was presented followed by 2000ms hash-tags screen which was included as a blinking pause for participants’ eyes. All 318 sentences were presented in randomized order to each participant including 36 attention questions following preselected sentences (twelve per condition with equivalent percentages of correct yes and no responses) to assure that participants attentively read the sentences. Participants were instructed to attentively and silently read the presented sentences and to answer occasionally inserted yes-no comprehension questions as correctly as possible by button press. Correct answers required the meaning of the entire sentence to be processed rather than only focusing on single words, e.g., “Das Schulmädchen tanzt mit dem Sieger“ (“The schoolgirl dances with the champion”) was followed by the question “Tanzt das Mädchen mit einem Gewinner?” (“Does the girl dance with a winner?”). Attention questions were presented until participants responded.

Ten initial practice trials each followed by an attention question ensured participants getting used to the word-by-word presentation and to how to answer the yes-no questions (respective feedback was provided only for practice trials). Three further practice trials were inserted to train to blink the eyes if necessary only during the blinking pause. Practice trials did not include critical words from the stimulus set. Duration of EEG recording was about one hour with four breaks in between.

2.5. Data acquisition and reduction

EEG-signal was recorded from 64 Ag/AgCl-electrodes. Four EOG electrodes were mounted to assess horizontal (HEOG) and vertical eye (VEOG) movements: Two electrodes on the outer canthi of both eyes and 2 electrodes on the infraorbital ridges of the right eye. 59 electrodes were affixed on the scalp using an elastic electrode cap (Easy GmbH Herrsching, Germany) at positions Fp1, Fp2, Fpz, AF3, AF4, F1, F2, F3, F4, F5, F6, Fz, FC1, FC2, FC3, FC4, FCz, FT7, FT8, C1, C2, C3, C4, C5, C6, Cz, T7, T8, CP1, CP2, CP3, CP4, CP5, CP6, CPz, TP7, TP8, P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, Pz, PO3, PO4, PO7, PO8, PO9, PO10, POz, O1, O2, Oz, and Iz of the international 10/10 system (Nuwer et al., 1998). The ground electrode was at position AFz. The raw EEG-signal was recorded using two 32-channel amplifiers (Brainamp, Brain Products, Germany) with amplifier's default filter settings (low cut-off time constant of 10 sec, which corresponds mainly to a 0.016 Hz high-pass filter and high cut-off frequency of 1000 Hz before digitization). After digitization with sampling rate of 5000 Hz and the signal was online filtered and down sampled to 500 Hz. Electrodes were online referenced to the right mastoid and impedances were kept below 5 kΩ. All channels were filtered offline using an IIR (Infinite Impulse Response) zero-phase shift Butterworth filters with a band pass filter 0.1 to 20 Hz (24 db / oct roll-off) and a notch filter of 50 Hz, and recalculated to average reference of left and right mastoids for the continuous data. Raw data were manually inspected and cleaned off noisy parts such as muscle artefacts, breaks and noisy electrodes (though all channels were kept for all subjects). Ocular artefacts were corrected using independent component analyses (Restricted Fast ICA, Zhou & Gotman, 2005; Jung et al., 1998, 2001) with Analyzer 2.0 software (Brain Products, Germany): ICA components were sorted in descending order according to their energy. Only the first up to six components were considered representing eye movement activity as categorized by scalp maps, which were removed before back transformation (*AM* = 2.40, *SD* = .97). EOG channels were then removed from further analyses. The continuous EEG signal was then segmented in 950 ms epochs starting 150 before target word (object word) onset, which served as pre-stimulus baseline. After baseline correction, segments containing artefacts (or those corresponding to incorrect responses to attention questions) were excluded from further analyses. Differences in values > 80 μV in intervals of 70ms as well as amplitudes >50 or <−50 μV were considered artifacts. Only segments free of artefacts were averaged per condition, participant and electrode, before grand averages were computed over all participants. Participants with less than 65 out of possible 106 segments per condition were excluded from analyses in order to optimize signal-to-noise ratio.

For the 38 participants included in the final data set mean numbers of segments per condition were: low affective deflection: *AM* = 96.63 (*SD* = 8.64), medium affective deflection: *AM* = 96.13 (*SD* = 9.35), and high affective deflection: *AM* = 96.37 (*SD* = 8.01) - not differing significantly between conditions [$F(2, 114) < 1$].

2.6. Analyses of EEG-Data

According to the literature on ERP effects in language processing involving emotion effects, ERP signals were segmented into three strategic time windows corresponding to P2/N2, N400 and P600 components – taking also into account the specific peaks and ranges of these overall waveforms in our data. In particular, for the early P2/N2 ERP component we chose a time window between 130 – 270 ms – meeting the general assumption that these components peak around 200 ms and the specific morphology of the waveform in our data (see Figure 2) where no single peak can be observed, but where the general positive going waveform around 200 ms seems to accommodate perfectly and as a whole into this time window of 140 ms length. As N2 effects reflecting conflict processing in emotion-laden words tend to have very localized frontocentral distributions (Kanske & Kotz, 2010; 2011; van Veen & Carter, 2002), we defined a region of interest (ROI) using the following six electrodes for analyses concerning the P2/N2 time window: F1, F2, Fz, FC1, FC2, FCz.

For the N400 component, we chose a standard time window of 300 – 450 ms, which contains the respective negative going waveform in our data almost entirely and is almost perfectly symmetrically distributed around the peak of this overall waveform in our data around 370 ms (see Figure 2).

Accordingly, we chose a standard time window for analyses of the LPC between 500 and 700 ms – matching the peak of this waveform at frontal electrodes in our data around 600 ms. We extended statistical analyses of ERP data to this later time window because some studies on emotional sentence (Holt et al., 2009, Delaney-Busch & Kuperberg, 2013; Wang, Bastiaansen, & Yang, 2015) and on moral statement processing (Van Berkum et al., 2009) reported late positivity effects around 600 ms.

Analyses of N400 and P600 components involved all electrodes: Each time six electrodes were grouped into clusters defining the two topographic factors *crosswise* (anterior, central, posterior) and *longitudinal* (left, middle, right): anterior left: Fp1, AF3, F3, F5, FC3, FT7; anterior middle: F1, F2, Fz, FC1, FC2, FCz; anterior right: Fp2, AF4, F4, F5, FC4, FT8; central left: C3, C5, T7, CP3, CP5, TP7; central middle: C1, C2, Cz, CP1, CP2, CPz; central right: C4, C6, T8, CP4, CP6, TP8; posterior left: P3, P5, P7, PO3, PO7, O1; posterior middle: P1, P2, Pz, POz, Oz, Iz; posterior right: P4, P6, P8, PO4, PO8, O2.

3 x 3 x 3 repeated measures ANOVAs were computed over mean amplitudes of clusters including the factors *crosswise* (3 levels: anterior, central, middle), *longitudinal* (3 levels: left, middle, right) and *affective deflection* (3 levels: low, medium, high) using Greenhouse-Geisser correction (Picton et al., 2000).

Post hoc-comparisons (using Benjamini Hochberg FDR correction for multiple comparisons) were conducted to further explore eventual main effects of the three level factor deflection or interactions of deflection effects with topographic factors.

2.7. Correlation Analyses: Rating Scales

To explore relationships of personality traits and ERP-signals elicited by emotional congruency of sentence final words, Pearson product moment correlation coefficients *r* were calculated between individual rating scale scores (SPQ-G and NEO-FFI) and individual differences between mean amplitudes per experimental condition for relevant time windows and topographic regions of interest (ROI). The anterior middle cluster of electrodes (= F1, F2, Fz, FC1, FC2, FCz) was used as ROI for the P2/N2 time window between 130-270ms and the central middle cluster (= C1, C2, Cz, CP1, CP2, CPz) was used for the N400 time window between 300-450ms[4].

3. Results

3.1. Behavioral data

All 38 subjects answered correctly at least two-thirds of the attention questions (M = 88.63%, SD = 8.10%). Response accuracy was not affected by deflection condition [$F(2, 111) = .081$, $p = .446$, $N = 38$; low deflection: $M = 10.39$ ($SD = 1.46$), medium deflection: $M = 10.76$ ($SD = 1.38$), high deflection: $M = 10.37$ ($SD = 1.67$)]. The behavioral data thus suggest that all participants attentively read and understood the presented sentences.

3.2. EEG data

Figure 2 shows the grand averaged ERPs elicited by target words in the conditions of low, medium, and high affective deflection on three representative electrodes (Fz, CPz, Pz). Target words in the low affective deflection generally elicited more positive ERPs across all three time windows of interest compared to target words completing sentences with medium or high affective deflection. Figure 3 shows topographies of contrasts between relevant conditions in different time windows. Contrasts between two conditions appear generally most pronounced between medium and low deflection conditions. Increasing negativity is

[4] Because ANOVAs on ERP data revealed no significant effects between 500-700ms, we did not correlate ERP data with questionnaire scores for this time window.

focused on fronto-central electrodes for the early and late time windows and rather centrally distributed for the N400 time window. To illustrate effects, Table 2 presents means and standard deviations for anterior, central, and posterior scalp regions for three conditions of affective deflection for the three relevant time windows.

< Figure 2 & 3 and Table 2 about here >

Figure 2 Results a). Grand Averages for electrode Fz, CPz, and Pz were effects were most pronounced.

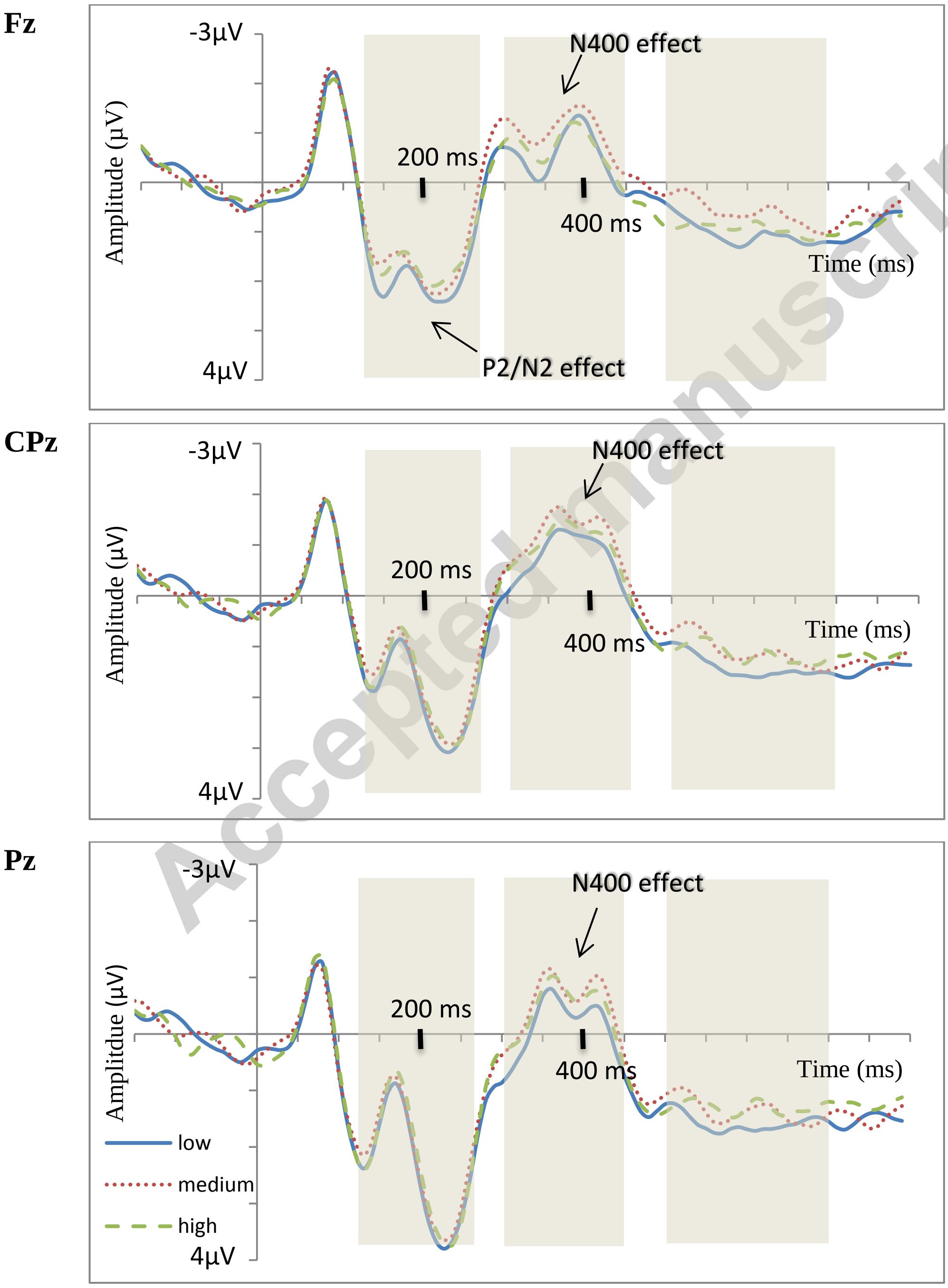


Figure 3 Results b). Topographies of contrasts and relevant time ranges.

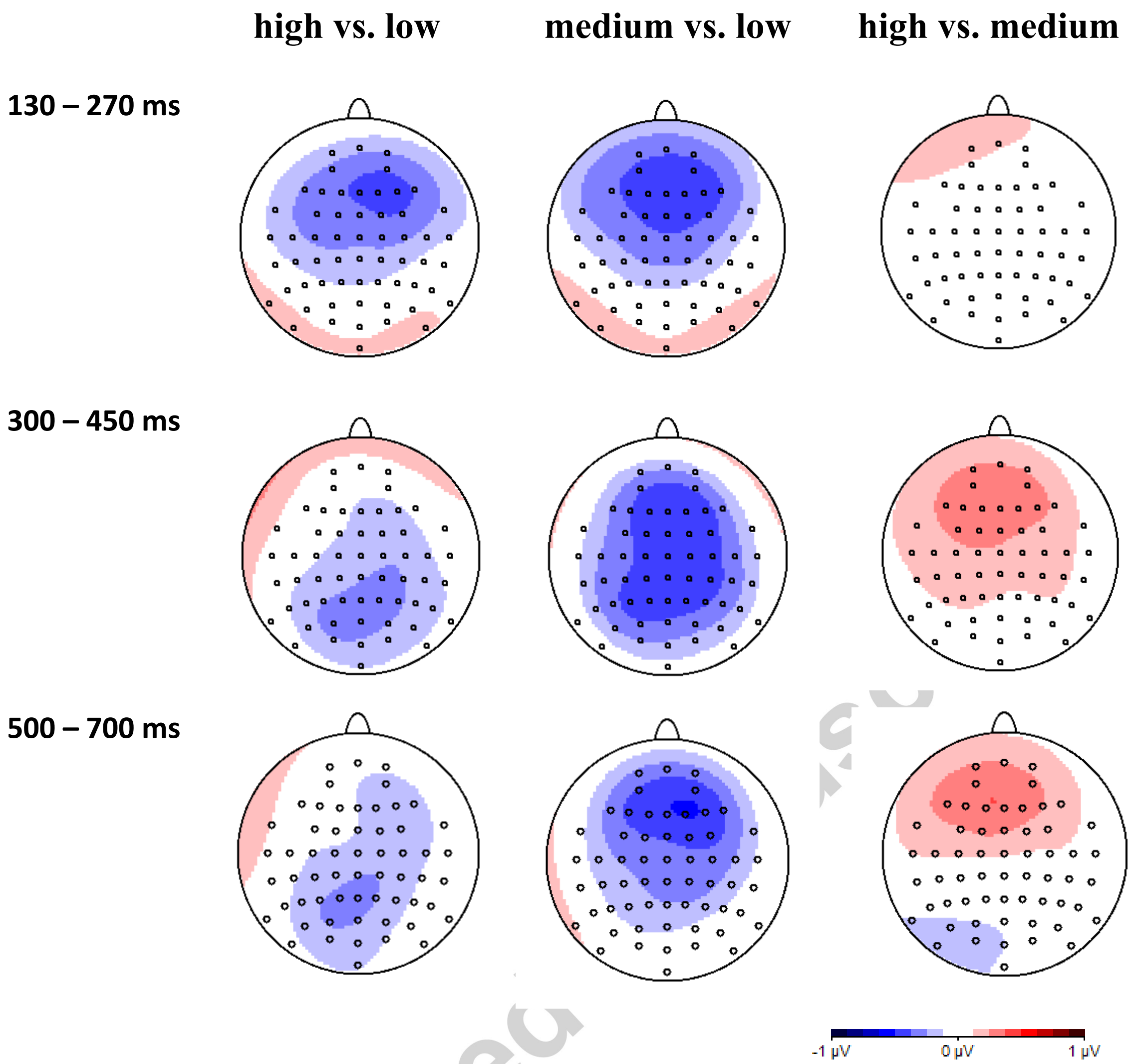


Table 2

Means in microvolt and standard deviations in parentheses for anterior, central, and posterior scalp regions for the three conditions of affective deflections for the relevant time windows (N = 38).

	Timewindow								
	130-270 ms			**300-450 ms**			**500 – 700 ms**		
	Condition of affective deflection								
Regions	**low**	**medium**	**high**	**low**	**medium**	**high**	**low**	**medium**	**High**
Anterior	1.68 (2.11)	1.44 (2.10)	1.43 (2.20)	-.55 (2.46)	-.90 (2.53)	-.62 (2.59)	.96 (1.45)	.59 (1.53)	.86 (1.69)
Central	1.59 (1.63)	1.43 (1.59)	1.42 (1.74)	-.26 (2.17)	-.58 (2.17)	-.40 (2.12)	1.35 (1.35)	1.13 (1.27)	1.22 (1.21)
Posterior	2.88 (2.18)	2.81 (2.27)	2.83 (2.27)	.85 (1.65)	.54 (1.83)	.62 (1.70)	1.32 (1.09)	1.21 (1.21)	1.14 (.98)

3.2.1. Early effect: 130-270 ms. The P2 component can be identified in our grand averages between 100 and 300 ms at frontal electrode sites. However, the waveform of the P2 component is modulated by an overlapping negativity which peaks around 170 ms. This negativity could be related to the N2 component itself or to any other effect of the presentation/processing of the stimuli in this specific task. In any case, we based our analysis on the average amplitude values of a time window adapted from previous literature of the N200 effect (e.g. Kotz, 2010; 2011). At these latencies, visual inspection shows an early effect of increased negativity for medium and high vs. low deflection conditions on fronto-central electrode sites, which was confirmed by statistical analyses: The ANOVA for the ROI on anterior electrodes with deflection as a three-level factor (low, medium, high) revealed a main effect of deflection [$F(2, 74) = 3.384$, $p = .046$, $\eta^2 = .084$]. Further comparisons showed increased negativity for high deflection condition [$AM_{\text{high\_deflection}} = 1.749$ µV, $SD_{\text{high\_deflection}} = 2.358$) $F(1, 37) = 5.813$, $p = .021$, $\eta^2 = .136$] and medium deflection condition [$AM_{\text{medium\_deflection}} = 1.748$ µV , $SD_{\text{medium\_deflection}} = 2.553$ µV, $F(1, 37) = 6.819$, $p = .013$, $\eta^2 = .156$] as compared to low deflection condition ($AM_{\text{low\_deflection}} = 2.058$ µV, $SD_{\text{low\_deflection}} = 2.432$). No significant difference was given for the comparison of high vs. medium deflection, $F(1, 37) < 1$.

3.2.2. N400: 300-450 ms. Visual inspection of the ERP data revealed a classical distribution of the N400 being most pronounced on centro-parietal electrodes and peaking at 370ms. ANOVA revealed a main effect of deflection [$F(2, 74) = 3.505$, $p = .038$, $\eta^2 = .087$]. Neither the three way interaction [$F(8, 296) < 1$], nor interactions between the factors crosswise [$F(4, 148) < 1$] or longitudinal and deflection [$F(4, 148) < 1$] were significant. Further comparisons showed that only the difference between medium and low deflection with increased negativity for the medium deflection condition was significant [$F(1, 37) = 7.153$, $p = .011$, $\eta^2 = .162$], while other comparisons revealed no significant effects [$F_{low\text{-}high}$

$_{deflection}(1, 37) = 1.898$, $p = .177$, $\eta^2 = .049$; $F_{medium\text{-}high\ deflection}(1, 37) = 1.639$, $p = .208$, $\eta^2 = .042$].

3.2.3. Late Positive Complex. 500-700 ms. Visual inspection indicated the classical positive going waveform over all electrodes for this time window with more negative averages for conditions of medium or high compared to the condition of low affective deflection. However, statistical analyses revealed no significant main effect of deflection [$F(2, 74) = 1.720$, $p = .187$, $\eta^2 = .044$], no three-way interaction [$F(8, 296) = 1.484$, $p = .191$, $\eta^2 = .039$] and no interaction between the factors crosswise [$F(4, 148) = 2.120$, $p = .121$, $\eta^2 = .054$] or longitudinal and deflection [$F(4, 148) < 1$].

3.3. Rating Scales

Please see Table 3 and 4 for descriptive statistics and resulting correlations. Significant correlations between scores on the SPQ and ERP data were restricted to the contrast between high and medium deflection conditions: The size of this somewhat counterintuitive contrast in our ERP data – less negativity for the high as compared to the medium condition - during both the 300-450 ms and the 130-270 ms time windows was associated with an increase in total schizotypy scores and in the Subscale of Eccentric Behavior. Furthermore, for the N400 time window, a similar positive association of the counterintuitive ERP contrast was found with the Cognitive Perceptive Factor.

NEO-FFI scores displayed additional correlations with ERP components: During the N400 interval, increasing Neuroticism scores were associated with stronger negativity for the high as compared to the medium condition (representing the canonically expected ERP pattern). Further, consistent significant correlations between ERPs from the two time windows were observed concerning Agreeableness: the higher individual Agreeableness scores, the less pronounced the hypothesized relative negativity of ERP waves for high vs. low affective deflection condition..

< Table 3 & 4 about here >

Table 3
Means, standard deviations, and ranges of rating scale scores for the study sample (N = 38).

Scale	*Mean*	*SD*	*Range*	
SPQ				*Klein et al. (1997)*
SPQ total	16.68	9.97	1 – 46	21.6 (.28)
SPQ IR	2.55	2.32	0 – 11	3.2 (.36)
SPQ SA	2.26	2.04	0 – 9	2.2 (.27)
SPQ MT	0.87	1.4	0 – 6	1.4 (.20)
SPQ UPE	1.37	1.63	0 – 5	2.4 (.27)
SPQ EB	1.68	2.00	0 – 7	2.2 (.31)
SPQ NCF	1.29	2.10	0 – 10	1.7 (.19)
SPQ OS	3.79	2.55	0 – 9	3.8 (.42)
SPQ CA	1.53	1.74	0 – 6	2.2 (.27)
SPQ S	1.34	1.63	0 – 7	2.4 (.30)
SPQ F1	11.61	7.28	1 – 28	
SPQ F2	5.08	4.61	0 – 18	
NEO-FFI				*Körner et al. (2008)*
N	2.34	0.64	0.42 – 3.5	1.62 (.62)
E	1.52	0.54	0.58 – 2.75	2.2 (.50)
O	1.11	0.46	0.25 – 2	2.05 (.46)
A	1.33	0.51	0.25 – 2.25	2.54 (.47)
C	1.45	0.54	0.58 – 2.58	2.71 (.55)

SPQ total Schizotypy Personality Questionnaire total score, *IR* Ideas of Reference, *SA* Social Anxiety, *MT* Odd Beliefs of Magical Thinking, *UPE* Unusual Perceptual Experience, *EB* Odd or Eccentric Behavior, *NCF* No Close Friends, *OS* Odd Speech, *CA* Constricted Affect, *S* Suspiciousness, *SPQ F1* Cognitive-Perceptual Factor, *SPQ F2* Interpersonal Factor, *N* Neuroticism, *E* Extraversion, *O* Openness, *A* Agreeableness, *C* Conscientiousness
In the right columns scale scores from Klein et al. (1997) and Körner et al. (2008) are reported for comparison.

Table 4
Pairwise correlations for amplitude differences between grand averages of different conditions of affective coherence and subscales and total score of Schizotypal Personality Questionnaire (SPQ) and NEO-FFI for early (130 - 270ms) and N400 (300-450ms) time window.

Effect	Early			N400		
Cluster	Middle Anterior			Middle Central		
	Amplitude Difference for conditions of Affective Deflection					
Scale	high – low	medium – low	high – medium	high – low	medium - low	high – medium
SPQ						
SPQ total	.261	-.154	.320*	.172	-.223	.356*
SPQ IR	-.034	-.303	.195	-.042	-.242	.201
SPQ SA	.127	-.075	.156	-.024	-.107	.085
SPQ MT	.182	.013	.135	.193	.006	.153
SPQ UPE	.312	-.152	.360	.0133	-.204	.209
SPQ EB	.232	-.052	.223*	.126	-.285	.379*
SPQ NCF	.160	.052	.088	.171	-.062	.200
SPQ OS	.130	-.006	.109	.258	-.015	.226
SPQ CA	.256	-.112	.286	.210	-.101	.270
SPQ S	.043	-.150	.144	-.046	-.121	.080
SPQ F1	.214	-.178	.301	.142	-.233	.342*
SPQ F2	.255	-.052	.217	.147	-.114	.230
NEO-FFI						
N	-.045	.313	-.265	-.097	.268	-.340*
E	.206	.259	-.026	.111	.017	.074
O	-.104	-.120	.005	-.138	.133	-.242
A	.343*	.037	.246	.349*	.082	.206
C	.030	.126	-.070	.080	-.088	.151

SPQ total Schizotypy Personality Questionnaire total score, *IR* Ideas of Reference, *SA* Social Anxiety, *MT* Odd Beliefs of Magical Thinking, *UPE* Unusual Perceptual Experience, *EB* Odd or Eccentric Behavior, *NCF* No Close Friends, *OS* Odd Speech, *CA* Constricted Affect, *S* Suspiciousness, *SPQ F1* Cognitive-Perceptual Factor, *SPQ F2* Interpersonal Factor, *N* Neuroticism, *E* Extraversion, *O* Openness, *A* Agreeableness, *C* Conscientiousness
*uncorrected $p < .05$

4. Discussion

The present EEG study investigated implicit affective information processing in sentences: We employed a mathematical model of impression formation based on Affect Control Theory (ACT; Heise, 2007) to generate three conditions of emotional congruency for semantically correct sentences describing social interactions, and explored how the interplay between emotional connotations of multiple words influences semantic processing. Thus, the aim of the present study aim was twofold: the study served (a) to contribute new insights to the research field of emotional language processing and (b) to provide, for the first time, neuroscientific evidence with regard to the social psychological model of ACT. Our highly controlled design - regarding variables known to generally influence visual word recognition - allows for ascribing differences between conditions in the EEG-signal to the manipulation of affective coherence alone. Such effects could be revealed in the P2/N2 latency range (130-270 ms) on frontal electrode sites with more negative amplitudes for affectively incongruent (both medium and high deflection) vs. congruent sentences and in the N400 component (300-450 ms) with a centro-parietal scalp distribution, where, in particular, the condition of medium deflection provoked more negative ERP amplitudes compared to congruent sentences. In order to achieve perfect control of basic semantic content across conditions each participant had seen the same target words repeatedly across conditions – only the combination with different preceding sentence contexts (also repeated across conditions in a perfectly balanced manner) determined the experimental manipulation of affective coherence. Our results, in general, support the assumption that affective connotations of words influence semantic sentence processing - even in absence of an explicit emotion processing task. More specifically, the interplay of emotional connotations of different words combined in a sentence seems to provide an initial, basic frame for meaning making in terms of affective coherence and congruency. Furthermore, our data make a strong point for an implicit processing of affective content to shape meaning making already during very early phases of processing. Note that a clear-cut distinction between affective language processing on the one hand and general semantic processing on the other appears difficult in general. Already the origin of the widely used affective scales of valence and arousal makes this very clear: they emerged as dimensions accounting for the greatest amount of variance of semantic differentials (Osgood et al., 1957). Our ERP effects appear well in line with previous findings on semantic congruency effects. Respective effects in our data result from a manipulation of affective coherence via a mathematical model using affective ratings. Rather than assuming affective vs. semantic processing as a dichotomy, our data may best be understood as evidence that affective connotations at the level of single words and affective coherence at the sentence level influence the way our brain processes semantics.

Our results provide, further, novel neuroscientific evidence supporting the model of affect control theory from social psychology. According to this theory, reducing the affective deflection between one's conceptual representation of a social situation and one's actions is the core motivational principle that drives human social interaction (Heise, 2007), ensuring compliance of individuals with prevailing cultural norms as the result of an automatic information-processing mechanism (cf. Schröder & Thagard, 2013). Previous empirical tests of that bold claim have shown affective deflection as computed with the mathematical ACT model to predict variables such as likelihood judgments and behaviors (e.g., Heise & MacKinnon, 1987; Schröder & Scholl, 2009). The EEG results reported here add to the body of evidence linking the deflection parameter to real-world observations, buttressing the claim that affect control theory is a genuine multi-level theory of social interaction and emotion (cf. Rogers et al., 2014).

Affective coherence violation, operationalized by deflection, elicited an early effect on anterior electrode sites in the time range of the P2/N2 components. Considering the latency, scalp distribution and experimental manipulation, this effect can be related with previously reported N2 effects triggered by conflict detection. The amplitude of the N2 component is modulated by different tasks and manipulations, like for example, the Go/No-Go task, oddball paradigm, or sequential matching task. Consequently, the N2 component has been linked to different cognitive processes such as response inhibition, target probability, perceptual novelty, and mismatch detection (for a review see Folstein & Van Petten, 2008). Of special interest for our study is that the N2 effect obtained with the flanker task is enhanced with emotional words compared to neutral words, showing that emotional information can modulate conflict processing (Kanske & Kotz, 2010, 2011). This type of N2 effect has been related to activity in the Anterior Cingulated Cortex (ACC), a brain area implicated in conflict monitoring (van Veen & Carter, 2002; Yeung et al., 2004) involving emotional and non-emotional distractors (Egner et al., 2008; 2010b). Although we are aware of the differences between the flanker task and the comprehension task of our experiment, we believe that similar cognitive control is necessary in order to resolve the conflict posed by our sentences in the high and medium deflection conditions. A recent study has also described an N2 effect in a sentence reading experiment, and its authors have made a similar claim in relation with conflicting predictions during reading comprehension (Payne & Federmeier, 2017). Accordingly, and because our manipulation of affective coherence can be understood as a manipulation of the degree to which the emotional connotations of the words in a given sentence harmonize or, respectively, clash with each other, we suggest that the present effect is a correlate of conflict detection in terms of affective coherence violation occurring already before lexical processing is completed. Rapid processing of emotion features has been shown to affect the P1, N1, P2 on the level of single words (Bayer, Sommer, & Schacht, 2012; Bernat, Bunce, & Shevrin, 2001; Hofmann et al., 2009) as well as on sentence level for which an emotion effect was observed as early as in the 90-200 ms latency range (Wang et al., 2013), where inconsistent emotions elicited larger N100/P200 (Léon et al., 2010), or where personal disagreement evoked ERP differences in the 200-250 ms latency range (Van Berkum et al., 2009). Therefore, we expected and found violations of affective coherence to impact semantic sentence processing at early stages as a correlate of tracking conflicting emotional information induced by sentence final words representing social norm violations. This early effect was significant only for the contrasts low vs. medium and low vs. high, respectively, but not for the contrast medium vs. high deflection of affective coherence. We assume that this might be due to some kind of cut-off mechanism during early stages of semantic processing, i.e., until some degree of affective deflection the brain perceives a linguistically represented social interaction as emotionally congruent whereas all deflections larger than this will be judged as affectively incoherent without any further differentiation. Such a categorical processing style of emotional features in language processing has also been suggested by Estes and Adelman (2008) in the case of lexical decision times for differently valenced words. In the framework of ACT, the P2/N2 effect can be described, accordingly, as a correlate of the "mental stumbling" when we encounter linguistic representations of situations which are affectively incoherent; likely reflecting a rapid reciprocal link of cultural norms and values and contextual affective meaning in language processing.

In general terms, the present N400 effect with increased negativity in response to increasing affective deflection when comparing medium and low deflection conditions corresponds to canonical incongruency effects reported to sentence final words, which violate preceding sentential context in terms of semantic correctness or predictability (e.g. Kutas & Hillyard, 1980; Kutas & Federmeier, 2011; Hagoort et al., 2004; Lau, Holcomb, & Kuperberg, 2013). But note that our data does neither involve perfectly gradual N400 deflection effects, nor a shared pattern of effects for high and medium deflection conditions as the high deflection condition did not produce significant differential effects in the N400 time window. On the other hand, our results suggest that the N400 is sensitive to subtle violations of affective congruency (as represented by sentences in the medium deflection condition), which go beyond superficial semantic processing. In other words, our data suggest that affective meaning making is a basic constituent of semantic processing as – at least certain forms of - affective incongruence seem to affect what is generally understood as the most prominent marker of semantic processing in sentence reading. While Wang et al. (2011) and Hehman et al. (2014) found that the N400 is sensitive to social expectancy and Van Berkum et al. (2009) revealed it to be sensitive to individual value systems, our study extends these findings to more general prevailing cultural norms and values which are not bound to individual political beliefs or delimited social subject matters such as stereotypical convictions. Moreover and importantly, our data show that these neuronal correlates of meaning making and their sensitivity to cultural norms can partly be predicted using general affective meaning ratings and an ACT-based mathematical formalization of impression formation.

Going into more detail, we found that while our design including three different conditions of affective coherence seems a priori well suited for capturing potential gradual effects of this measure, the N400 effect of affective coherence was significant only for the contrast of low versus medium deflection. We consider two potential reasons for why increasing affective incoherence does not necessarily lead to gradual, linear effects across all conditions: A material-based one concerning the mental representation of deviant social events, and a personality-based one, which emphasizes N400 sensitivity to specific trait characteristics, for which we correlated our ERP data with participants' scores on personality questionnaires in an exploratory approach. Please note that both are post-hoc attempts to explain a potentially interesting but a priori unexpected detail of our findings, which might encourage future research.

As affective deflection – according to the ACT theory - represents violations of social norms, the a priori hypothesis of linearly increasing effects of the deflection manipulation may not be met in the case of particularly high deflection, because – at least at a conscious processing level – also an explicit "breaking the rules" attitude appears predictable to some degree in any norm system. One might expect such "paradoxical coherence" effects to be restricted to comparably late and more conscious processing stages as reflected by the N400 – as opposed to earlier more automatic processing stages. Whereas both medium and high deflection conditions had significantly differed from the low deflection condition at early processing stages during the P2/N2 time window – an effect we interpret as "mental stumbling" involving a cut-off mechanism impeding further differentiation between medium and strong norm violations, now, at later, more elaborate processing stages, the integration of very strong norm violations into elaborate mental schemas may be more easy when meeting prototypical patterns. In other words, the absence of an N400 effect for the high vs. low deflection contrast might be explained by mental schemas or prototypical representations of immoral events for which people form mental representations during the process of growth and acculturation. Acquiring knowledge about which acts are allowed to do ("pet the cat") and which are not ("don't slap the cat") is probably among the first learning experiences of children becoming active parts of society; thus, it is not so far off the mark to assume that people hold mental representations for strong norm violations which serve as source for normativity judgments of social events in everyday life. Empirically, this approach is supported, for instance, by Ask and Fransson (2001) who reported faster correct reaction times in a morality judgment (immoral/moral) task to sentences describing prototypical immoral events as compared to sentences describing non-typical immoral events; providing evidence for prototypic representations of norm deviations. Prototypical representations, hence, should allow for faster, i.e., easier integration processes reflected by a decreased N400 effect. Accordingly, although facing a null effect here, we carefully hypothesize post-hoc that sentence-final words with a strong affective deflection representing strong norm violations might be more easily integrated in sentence processing than weaker norm violations, because their accessibility is facilitated by existing schemas. The fact that comparing the high deflection condition to the "baseline" of low deflection resulted in the above described null effect only concerned the relatively late N400 window - whereas conflict monitoring in the earlier time window resulted in increased negativity for both medium and high deflection conditions – may support the idea of a 'conscious locus' of the discussed null effect for strong norm violations at relatively late processing stages.

Moreover, note that such more or less conscious attitudes toward violation of social norms presumably vary considerably across individuals. While resulting inter-individual variance represents an interesting topic for personality research, it may also prevent effects for the condition of high deflection from reaching significance in analyses treating participants as a homogenous group. Personality traits have the potential to influence language processing and related brain waves, e.g. schizotypy and the N400 (Kiang, 2010; Kutas, 2006). Our correlation analyses of personality characteristics for the early and the N400 effect may indeed offer some potentially interesting results: all correlations emerging as significant – though uncorrected for multiple testing - in either of both processing stages (note their general consistency across early and late ERP effects) involve the high affective deflection condition : Higher scores on *Agreeableness* seem to be associated with decreased respective ERP effects in the early as well as in the N400 time window, potentially due to an *Agreeableness*-related judgment style more liberally semantically integrating even immoral events. On the other hand, increasing *Neuroticism* appeared to be associated with an increased N400 for the comparison of high vs. medium affective deflection, that is, Neuroticism may go along with a tendency to pay close attention to existing social norms. Higher schizotypy appears correlated with a decreased effect only for the comparison of high vs. medium affective deflection in the early and the N400 time window. Besides a potential general alteration of semantic processing associated with schizotypy, this might reflect a less strict normativity judgment style concerning obvious or high deflection norm violations apparently due to high schizotypal persons' own tendency to show more *Eccentric Behavior*..

Together, these correlation results (uncorrected for multiple comparisons) emerge from a clearly exploratory approach on our data connecting some particular and rather unexpected findings among our results from the field of neuroscience with personality differences. Still, these findings may suggest that the failure to obtain an N400 modulation for the sentences representing particularly strong affective norm violations at the group level may partly be due to personality differences. Further, our correlation data, may still offer an interesting perspective on the way personality traits influence how people represent social interactions regarding their norm compliance, particularly in the case of strong norm violations and inspire future research - in particular as separate personality characteristics seemed to have differential influences on early, automatic, and later N400, i.e. more controlled, processing stages.

But note also that all of our sentences were of extremely low cloze probability and contextual constraints, which may represent another reason for why our N400 effects are in general less robust compared to studies involving really sharp contrasts between regular and irregular sentences.

Taken together, our findings make a strong case for emotional congruency to automatically influence meaning making during online sentence processing. Future studies should further investigate the mental and neural reactions to strong norm violations, while especially taking into account potential relationships between personality characteristics and moral judgment style during automatic vs. more controlled processing stages. Our preliminary findings suggesting such associations might serve to promote a holistic approach to the investigation of the role of affective processing in the process of meaning making. In order to investigate how our findings obtained for a manipulation of affective coherence of sentences relate to emotion processing in the brain in a stricter sense, i.e., involving the activation of brain structures typically associated with affect, like, e.g., the limbic system, future studies might employ neuroscientific methods allowing for a more profound "insight" into the brain, e.g. fMRI, as recent studies have already shown that the overall affective content of words (e.g., Kuchinke et al., 2005) and sentences (e.g., Hsu et al., 2015) triggers activation in emotion specific brain areas during the reading process.

.

References

Athota, V. S., O'Connor, P. J., & Jackson, C. (2009). The role of emotional intelligence and personality in moral reasoning. *European Journal of Personality Research, 11*, 453–470.

Altarriba, J., & Basnight-Brown, D. M. (2011). The representation of emotion vs. emotion-laden words in English and Spanish in the Affective Simon Task. *International Journal of Bilingualism, 15*, 310–328.

Ambrasat, J., von Scheve, C., Schauenburg, G., Conrad, M., & Schröder, T. (2014). Consensus and stratification in the affective meaning of human sociality. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 8001-8006.

American Psychiatric Association (1987). *DSM-III-R: Diagnostic and statistical manual of mental disorders. Washington*, DC: The Association.Averett, C. P., & Heise, D. R. (1987). Modified social identities: Amalgamations, attributions, and emotions. *Journal of Mathematical Sociology, 13*, 103-132.

Ask, K. & Fransson, N. (2001). *Mental representation of immoral events: The effect of event typicality on judgement time* (Göteborg Psychological reports No. 4, Vol. 31). Sweden: Göteborg University, Department of Psychology.

Averett, C. & Heise, D. (1987). Modified social identities: Amalgamations, attributions, and emotions. *The Journal of Mathematical Sociology, 13*, 1-2. doi: 10.1080/0022250X.1987.9990028

Barber, H. A. & Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Research Reviews, 53*, 98-123.

Bayer, M., Sommer, W., & Schacht, K. (2012). P1 and beyond: Functional separation of multiple effects in word recognition. *Psychophysiology, 49*, 959-569.

Bernat, E., Bunce, S., & Shevrin, H. (2001). Event-related brain potentials differentiate positive and negative mood adjectives during both supraliminal and subliminal visual processing. *International Journal of Psychophysiology, 42*, 11-34.

Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae.* Göttingen: Hogrefe.

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 379-390.

Citron, F. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language, 122*, 211–226.

Conrad, M., Recio, G., & Jacobs, A. M. (2011). The time course of emotion effects in first and second language processing: across cultural ERP study with German–Spanish bilinguals. *Frontiers in Psychology, 2*, 1-16.

Costa, P. T., Jr. & McCrae, R. R. (1989). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research, 1084*, 89-103.

Delaney-Busch, N. & Kuperberg, G. R. (2013). Friendly Drug-dealers and Terrifying Puppies: Affective primacy can attenuate the N400 effect in emotional discourse contexts. *Cognitive, Affective, and Behavioral Neuroscience, 13*, 473-90.

Dillmann, J. (2003). *Negative Priming als kognitiver Vulnerabilitätsmarker bei Personen mit Schizotypie auf Basis faktor- und clusteranalytisch ermittelter multidimensionaler Merkmale* (unpublished doctoral dissertation). Friedrich-Schiller-Universität Jena, Germany.

Doerksen, S. & Shimamura, A. P. (2001): Source Memory Enhancement for Emotional Words. *Emotion, 1*, 5-11.

Duncan, S., & Barrett, L. F. (2007). Affect is a form of cognition: A neurobiological analysis. *Cognition and emotion*, *21*(6), 1184-1211.

Egner, T., Monti, J. M., Trittschuh, E. H., Wieneke, C. A., Hirsch, J., & Mesulam, M. M. (2008). Neural integration of top-down spatial and feature-based information in visual search. *The Journal of Neuroscience, 28*, 6141-6151.

Eriksen, B. A. & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*, 143-149.

Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words in lexical decision and naming: Comment of Larsen, Mercer, and Balota (2006). *Emotion, 8*, 441–444. doi:10.1037/1528-3542.8.4.441.

Federmeier, K. D., Wlotko, E., De Ochoa-Dewald, E., and Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146*, 75-84.

Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74*, 967–984.

Fischler, I. & Bradley, M. (2006). Event-related potential studies of language and emotion: words, phrases, and task effects. *Progress in Brain Research, 156*, 185-203.

Folstein, J. R. & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*, 152-170.

Fossum, T. A. & Feldman Barrett, F. (2000). Distinguishing Evaluation from Description in the Personality-Emotion Relationship. *Personality and Social Psychology Bulletin, 26*, 1027-1035.

Grimm, S., Weigand, A., Kazzer, P., Jacobs, A. M., & Bajbouj, M. (2012). Neural mechnisms underlying the integration of emotion and working memory. *NeuroImage, 61*, 1188–

1194.

Hagoort, P., Hald, L. A., Bastiaansen, M. C. M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*, 438-441. doi:10.1126/science.1095455.

Hehman, E., Volpert, H. I., & Simons, R. F. (2014). The N400 as an index of racial stereotype accessibility. *Social Cognitive and Affective Neuroscience, 9*, 544-552.

Herbert, C., Ethofer, T., Anders, S., Junghofer, M., Wildgruber, D., Grodd, W., & Kissler, J. (2009). Amygdala activation during reading of emotional adjectives—An advantage for pleasant content. *Social Cognitive and Affective Neuroscience, 4*, 35–49.

Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology: Interdisciplinary and Applied, 21*, 107-112.

Heise, D. R. (1979). *Understanding events: Affect and the construction of social action*. New York, NY: Cambridge University Press.

Heise, D. R. (2007). *Expressive order: Confirming sentiments in social action*. New York, NY: Springer.

Heise, D. R. (2010). *Surveying cultures: Discovering shared conceptions and sentiments*. Hoboken, NJ: Wiley.

Heister, J., Würzner, K.-M. Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB – eine lexikalische Datenbank für die psychologische und linguistische Forschung. *Psychologische Rundschau, 62*, 10-20.

Hofmann, M. J., Kuchinke, L., Tamm, S., Võ,M. L.-H., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective, & Behavioral Neuroscience, 9*, 389–397.

Holt, D. J., Lynn, S.K., & Kuperberg, G. R. (2009). Neurophysiological correlates of comprehending emotional meaning in context. *Journal of Cognitive Neuroscience, 21*, 2245-2262.

Hsu, C.-T., Conrad, M., and Jacobs, A. M. (2014). Fiction feelings in Harry Potter: Haemodynamic response in the mid-cingulate cortex correlates with immersive reading experience. *Neuroreport* 25, 1356–1361. doi: 10.1097/WNR. 0000000000000272

Hsu, C.-T., Jacobs, A. M., Altmann, U., and Conrad, M. (2015c). The magical activation of left amygdala when reading harry potter: an FMRI study on how descriptions of supranatural events entertain and enchant. *PLoS One* 10:e0118179. doi: 10.1371/journal.pone.0118179

Hsu, C.-T., Jacobs, A. M., Citron, F., and Conrad, M. (2015b). The emotion potential of words and passages in reading harry potter: an fMRI study. *Brain Lang.* 142, 96–114. doi: 10.1016/j.bandl.2015.01.011

Hsu, C.-T., Jacobs, A. M., and Conrad, M. (2015a). Can Harry Potter still put a spell on us in a second language? An fMRI study on reading emotion-laden literature in late bilinguals. *Cortex* 63, 282–295. doi: 10.1016/j.cortex.2014. 09.002

Huang, Y., Kendrick, K. M., & Yu, R. (2014). Social conflicts elicit an N400-like component. *Neuropsychologia. 65*, 211–220.

Jacobs AM, Võ ML-H, Briesemeister BB, Conrad M, Hofmann MJ, Kuchinke L, Lüdtke J and Braun M (2015) 10 years of BAWLing into affective and aesthetic processes in reading: what are the echoes? Front. Psychol. 6:714. doi: 10.3389/fpsyg.2015.00714

Jung, T.-P., Humphries, C., Lee, T.-W., Makeig, S., McKeown, M.J., Iraguid, V., Sejnowski, T.J. (1998). Extended ICA Removes Artifacts from Electroencephalographic Recordings. In M. I. Jordan , M. J. Kearns, & S. A. Solla (Eds*.), Advances in Neural Information Processing Systems 10* (pp. 894-900), Cambridge/London : MIT Press/Bradford Books.

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J. (2001). Analysis and Visualization of Single-Trial Event-Related Potentials, *Human Brain Mapping, 14*, 166-185.

Kanske, P. & Kotz, S. A. (2010). Modulation of early conflict processing N200 responses to emotional words in a flanker task. *Neuropsychologia, 48*, 3661-3664.

Kanske, P. & Kotz, S. A. (2011), Conflict processing is modulated by positive emotion: ERP data from flanker task. *Beavioural Brain Research, 219*, 382-386.

Kensinger, E. A. & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition, 31*, 1169-1180.

Kiang, M. (2010). Schizotypy and language: A review. *Journal of Neurolinguistics, 23*, 193-203. doi: 10.1016/j.jneuroling.2009.03.002

Kiang, M., Prugh, J., & Kutas, M. (2010). An event-related brain potential study of schizotypal personality and associative semantic processing. *International Journal of Psychophysiology, 75*, 119-126. doi: 10.1016/j.ijpsycho.2009.10.005.

Kim, Y. Y., Lee, B., Shin, Y. W., Kwon, J. S., & Kim, M.-S. (2006). Activity of left inferior frontal gyrus related to word repetition effects: LORETA imaging with 128-channel EEG and individual MRI. *NeuroImage, 29*, 712-720.

Kissler, J. & Herbert C. (2013). Emotion, Etmnooi, or Emitoon? - Faster lexical access to emotional than to neutral words during reading. *Biological Psychology*. doi:pii: S0301-0511(12)00195-0. 10.1016/j.biopsycho.2012.09.004.

Klein, C., Andresen, B., & Jahn, T. (1997). Erfassung der schizotypen Persönlichkeit nach DSM-III-R. *Diagnostica, 43*, 347-369.

Koelsch, S., Jacobs, A. M., Menninghaus, W., Liebal, K., Klann-Delius, G., von Scheve, C., & Gebauer, G. (2015). The quartet theory of human emotions: an integrative and neurofunctional model. *Physics of life reviews*, *13*, 1-27.

Körner, A., Drapeau, M., Albani, C., Geyer, M., Schmutzer, G., & Brähler, E. (2008). Deutsche Normierung des NEO-Fünf-Faktoren-Inventars (NEO-FFI). *Zeitschrift für Medizinische Psychologie, 17*, 133-144.

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition, 112*, 473–481. doi: 10.1016/j.cognition.2009.06.007

Kuchinke, L., Jacobs, A. M., Grubich, C., Võ, M. L.-H., Conrad, M., & Herrmann, M. (2005). Incidental effects of emotional valence in single word processing: An fMRI study. *NeuroImage, 28*, 1022–1032.

Kuchinke, L., Võ, M. L.-H., Hofmann, M., Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. International Journal of Psychophysiology, 65 (2), 132-140.Kutas, M. & Federmeier, K.D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647.

Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Kutas, M. & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*,161-163.

Kutas, M. (2006). One Lesson Learned: Frame Language Processing – Literal and Figurative – as a Human Brain Function. *Mataphor and Symbol, 21*, 285-325.

Kutas, M. & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review Psychology*, 62, 621-647. doi: 10.1146/annurev.psych.093008.131123

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single word contexts. *Journal of Cognitive Neuroscience, 25*, 484-502.

León, I., Díaz, J. M., de Vega, M., & Hernández, J. A. (2010). Discourse-based emotional consistency modulates early and middle components of event-related potentials. *Emotion, 10*, 863-873. doi: 10.1037/a0019983.

Lüdtke, J. & Jacobs, A. (2015). The emotion potential of simple sentences: additive or interactive effects of nouns and adjectives? *Frontiers in Psychology, 6*, Article 1137. doi: 10.3389/fpsyg.2015.01137

MacKinnon, N. J. (1994). *Symbolic interactionism as affect control*. Albany: State University of New York Press.

Martín-Loeches, M., Fernández, A., Schacht, A., Sommer, W., Casado, P., Jiménez-Ortega, L., & Fondevila, S. (2012). The influence of emotional words on sentence processing: electrophysiological and behavioral evidence. *Neuropsychologia, 50*, 3262-3272. doi:

10.1016/j.neuropsychologia.2012.09.010.

McCrae, R. R. & Costa, P. T., Jr. (1991). Adding Liebe und Arbeit: The Full Five-Factor model and Well-Being. *Personality and Social Psychology Bulletin, 17*, 227-232.

Molinaro, N., Conrad, M., Barber, H., & Carreiras, M. (2010). On the functional nature of the N400: Contrasting effects related to visual word recognition and contextual semantic integration. *Cognitive Neuroscience, 1*, 1-7, doi:10.1080/17588920903373952.

Nieuwland, M.S. & Van Berkum, J.J.A. (2006a). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*, 1098-1111.

Osgood, C. E., May, W. H., Miron, M. S. (1975). *Cross-cultural universals of affective meaning*. Urbana: University of Illinois.

Osgood, C. E., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning.* Urbana: University of Illinois.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*, 97–113. doi:10.1016/0028-3932(71)90067-4.

Payne, B. R., & Federmeier, K. D. (2017). Pace yourself: Intraindividual variability in context use revealed by self-paced event-related brain potentials. *Journal of Cognitive Neuroscience*. 29(5), 837–854.

Penolazzi, B., Hauk, O. & Pulvermüller, F. 2007. Early lexical access and semantic context integration as revealed by event-related brain potentials. *Biological Psychology*, ***74***, 374-388.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Jr., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publications criteria. *Psychophysiology, 37*, 127-152.

Raine, A. (1991). The SPQ: a scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophrenia Bulletin, 17*, 555–564.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin, 142*, 372-422.

Recio, G., Conrad, M., Hansen, L. B., & Jacobs, A. M. (2014). On pleasure and thrill: the interplay between arousal and valence during visual word recognition. *Brain and language*, *134*, 34-43.

Robinson, D. T., Smith-Lovin, L. & Wisecup, A. K. (2006). Affect control theory. In J. E. Stets & J. H. Turner (Ed.), *Handbook of the Sociology of Emotions* (pp. 179-202). US: Springer.

Rogers, K. B., Schröder, T. & von Scheve, C. (2014). Dissecting the sociality of emotion: A multi-level approach. *Emotion Review, 6*(2), 124-133.

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality, 11*, 273-294.

Schacht, A., & Sommer, W. (2009). Time course and task dependence of emotion effects in word processing. Cognitive, *Affective & Behavioural Neuroscience, 9*, 28–43. doi:10.3758/CABN.9.1.28.

Sass, S. M., Heller, W., Stewart, J. L.,LevinSilton,R., Egdar, J.C.,Fisher, J. E., & Miller, G. A. (2010). Time course of attentional bias in anxiety: Emotion and gender specificity. *Psychophysiology, 47*, 247–259.

Schauenburg, G., Ambrasat, J., von Scheve, C., Schröder, T., & Conrad, M. (2014). Emotional Connotations of Words related to Authority and Community. *Behavior Research Methods*. Advance online publication. doi: 10.3758/s13428-014-0494-7.

Schmidtke, D., Schröder, T., Jacobs, M., & Conrad, M. (2014). ANGST: Affective Norms for German Sentiment Terms derived from the Affective Norms of English Words. *Behavior Research Methods*. Advance online publication. doi: 10.3758/s13428-013-0426-y.

Scholl, W. (2013). The socio-emotional basis of human interaction and communication: How we construct our social world. *Social Science Information, 52*, 3-33.

Schröder, T., Hoey, J., & Rogers, K. B. (2016). Modeling dynamic identities and uncertainty in social interaction: Bayesian affect control theory. *American Sociological Review, 81*, 828-855.

Schröder, T. & Thagard, P. (2013). The affective meanings of automatic social behaviors: Three mechanisms that explain priming. *Psychological Review, 120*, 255-280.

Schröder, T. (2011). A model of language-based impression formation and attribution among Germans. *Journal of Language and Social Psychology, 30,* 82-102.

Scott, G. G., O'Donnell, P. J., Sereno, S. C. (2012). Emotion words affect eye fixations during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 783–792.

Sereno, S.C., Brewer, C.C., & O'Donnell, P.J. (2003). Context effect in word recognition: evidence for early interactive processing. *Psychological Science, 14*, 328–333.

Smith-Lovin, L. & Heise, D. R. (1988). *Analyzing social interaction: Advances in affect control theory*. New York, NY: Gordon and Breach Science.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly, 30*, 415-433.

Thagard, P. & Schröder, T. (2014). Emotions as semantic pointers: Constructive neural mechanisms. In L. F. Barrett & J. Russell (Eds.), *The psychological construction of emotion* (pp. 144-167). New York: Guilford.

Van Berkum, J. J. A., Holleman, B., Nieuwland, M. S., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological*

Science, 20, 1092 -1099. doi:10.1111/j.1467-9280.2009.02411.x.

van Veen, V. & Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology and Behavior, 77*, 477–482.

Võ, M. L.-H., Jacobs, A. M., Kuchinke, L., Conrad, M., Schacht, A., & Hutzler, F. (2008). Emotion/cognition coupling in the eye: Introducing the pupil old/new effect. *Psycho-physiology, 45*, 130-140.

Võ, M. L.-H., Jacobs, A. M., & Conrad, M. (2006*).* Cross-validating the Berlin Affective Word List. *Behavior Research Methods, 38*, 606-609.

Wang, L., Ma, Q., Song, Z., Shi, Y., Wang,Y., & Pfotenhauer, L. (2011). N400 and the activation of prejudice against rural migrant workers in China. *Brain Research.1375*, 103–110.

Wang, L., Bastiaanse, M. C. M., & Yang, Y. (2015). ERP respo0nses to person names as a measure of trait inference in person perception. *Social Neuroscience, 10*, 89-99. doi:10.1080/17470919.2014.944995.

Wang, L., Bastiaansen, M. C. M., Yang, Y., & Hagoort, P. (2013). ERP evidence on the interaction between information structure and emotional salience of words. *Cognitive, Affective and Behavioral Neuroscience, 13*, 297-310. doi:10.3758/s13415-012-0146-2.

Weigand, A., Grimm, S., Astalosch, A., Guo, J. S., Briesemeister, B. B., Lisanby, S. H., & Bajbouj, M. (2013). Lateralized effects of prefrontal repetitive transcranial magnetic stimulation on emotional working memory. *Experimental Brain Research, 227*, 43–52.

White, K. R., Crites, S. L. Jr., Taylor, J. H., & Corral G. (2009). Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Social Cognitive and Affective Neuroscience, 4*, 191-198.

Wundt, W. (1896). *Gundriss der Psychologie*. Leipzig, Germany: Entgelmann.

Yeung, N., Botvinick, M. M. & Cohen, J. D. (2004). The neural basis of error-detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*, 931-959.

Zhou, W. & Gotman, J. (2005). Removing eye-movement artifacts from EEG during intracarotid amobartial procedure. *Epilepsia, 46*, 409-414.

Appendix

Calculation of affective coherence

For calculating the degree of affective coherence of each sentence stimulus, we used impression-formation equations fitted for the German language by Schröder (2011). Affective coherence is reversely determined as the deflection (D) between the basic without-context sentiments (basic EPA-profile) and the transient affective meaning (transient EPA-profile) of involved concepts (actor, behavior, object) in a given context:

$$D = (A'_e - A_e)^2 + (A'_p - A_p)^2 + (A'_a - A_a)^2 + (B'_e - B_e)^2 + (B'_p - B_p)^2 + (B'_a - B_a)^2 + (O'_e - O_e)^2 + (O'_p - O_p)^2 + (O'_a + O_a)^2$$

where A, B, and O label basic without-context EPA-ratings of Actor, Behavior, and Object. A', B', and O' label the transient EPA-ratings of the same concepts within a given context. The indeces e, p, and a label the ratings in each dimension of evaluation, potency, and activity.

Transient EPA-profile of an actor (A')

$$A'_e = -.38 + .42*A_e - .11*A_a + .47*B_e + .11*O_e + .05*A_e*B_e + .06*A_e*O_a + .09*A_a*O_e + .09*A_a*O_a + .04*B_e*O_e - .07*B_e*O_a - .13*B_p*O_e - .03*A_e*B_e*O_p + .02*A_e*B_p*O_e - .02*A_p*B_p*O_e + .03*A_p*B_e*O_a$$

$$A'_p = -.03 + .39*A_p + .08*A_a - .07*B_e + .57*B_p - .20*O_p + .16*O_a - .04*A_p*B_a - .07*A_a*O_p + .03*B_a*O_e + .06*B_a*O_p + .02*A_e*B_p*O_a + .02*A_p*B_a*O_a$$

$$A'_a = .10 + .39*A_a - .13*B_e + .14*B_p + .52*B_a - .03*A_p*B_a - .03*A_p*O_e - .06*A_a*B_a + .04*A_a*O_p + .07*B_p*O_p - .04*A_a*B_a*O_p$$

Transient EPA-profile of a behavior (B')

$$B'_e = -.72 + .23*A_e + .51*B_e + .20*O_e + .06*A_e*B_e + .08*A_e*B_p + .04*A_e*O_e - .04*A_e*O_p + .05*A_a*O_p + .09*A_a*O_a + .06*B_e*O_e - .09*B_e*O_a - .10*B_p*O_e + .03*A_e*B_p*O_a - .05*A_a*B_a*O_p$$

$B'_p = -.05 + .17*A_p + .10*A_a + .66*B_p + .02*A_e*B_a + .04*A_e*O_a - .09*A_a*B_p$

$- .05*B_e*O_a + .02*B_p*O_e - 01.*A_e*B_a*O_p + .02*A_p*B_p*O_a +$

$.03*A_a*B_e*O_a$

$B'_a = .18 + .28*A_a - .06*B_e + .62*B_a - .02*A_e*B_e - .03*A_p*O_e - .07*A_a*B_a +$

$.04*B_e*O_p + .04*B_e*O_a + .08*B_a*O_p + .02*A_e*B_a*O_e + .02*A_p*B_a*O_e -$

$.03*A_a*B_a*O_p + .03*A_a*B_a*O_a$

Transient EPA-profile of an object (*O*')

$O'_e = -.15 + .10*A_p + .13*B_e + .38*O_e + .06*A_e*B_e + .03*A_e*O_e - .04*A_p*B_p$

$- .03*A_a*B_e + .04*A_a*O_p - .06*B_p*O_a$

$O'_p = -.26 - .28*A_p + .17*B_e - .54*B_p + .15*B_a + .40*O_p + .03*A_e*O_p +$

$.08*A_p*B_a + .09*A_a*O_e + .06*A_a*O_p - .06*B_p*O_e - .03*A_p*B_a*O_a$

$O'_a = -.57 - .18*A_p + .28*O_a + .05*A_p*B_a + .05*A_p*O_p + .03*A_a*B_e -$

$.08*B_p*O_e + .08*B_p*O_p + .01*Ae*Be*OA + .01*Ae*Ba*Oe -$

$.03*Ap*Bp*Op - .03*Aa*Be*OA - .02*Aa*Ba*Oa$

Higlights

- Implicit processing of affective consistency during sentence processing
- Affect Control Theory used for mathematical prediction of sentence reading ERPs
- Affective consistency of sentences influences P200 and N400
- Affective meanings are basic constituents of meaning making
- Affective inconsistency of phrases is discovered already after about 200ms