Köln Z Soziol (2024) 76:603–625
https://doi.org/10.1007/s11577-024-00961-3

Kölner Zeitschrift für Soziologie und Sozialpsychologie

ABHANDLUNGEN

Conflict Detection in Language Processing: Using Affect Control Theory to Predict Neural Correlates of Affective Incongruency in Social Interactions

Gesche Schauenburg · Arash Aryani · Chun-Ting Hsu · Tobias Schröder · Markus Conrad · Christian von Scheve · Arthur M. Jacobs

Received: 23 May 2023 / Accepted: 11 June 2024 / Published online: 6 August 2024

G. Schauenburg
Department of Psychiatry and Psychotherapy, Clinical Neuropsychology Unit, University Medical Center Hamburg-Eppendorf
20246 Hamburg, Germany
E-Mail: g.schauenburg@uke.de

A. Aryani
Biological Psychology and Cognitive Neuroscience, Freie Universität Berlin
Habelschwerdter Allee 45, 14195 Berlin, Germany
E-Mail: arash.aryani@fu-berlin.de

T. Schröder
Institute for Applied Research Urban Futures (IaF), University of Applied Sciences Potsdam
Kiepenheuerallee 5, 14469 Potsdam, Germany
E-Mail: tobias.schroeder@fh-potsdam.de

M. Conrad
Universidad de La Laguna
38200 San Cristobal de La Laguna, Spain
E-Mail: maconrad@ull.es

✉ C. von Scheve
Institute of Sociology, Freie Universität Berlin
Garystr. 55, 14195 Berlin, Germany
E-Mail: christian.von.scheve@fu-berlin.de

A. M. Jacobs
Experimental and Cognitive Neuropsychology, Freie Universität Berlin
Habelschwerdter Allee 45, 14195 Berlin, Germany
E-Mail: ajacobs@zedat.fu-berlin.de

C.-T. Hsu
Department of Education and Psychology, Freie Universität Berlin
Habelschwerdter Allee 45, 14195 Berlin, Germany
E-Mail: chun-ting.hsu@riken.jp

Abstract Affect control theory (ACT) is a sociological theory of meaning processing in social interactions. Meaning, according to ACT, derives from cultural institutions and situational affordances, having denotative (declarative) as well as connotative (affective) properties. Mathematical formalizations of ACT allow predictions of affective incongruency (in the terminology of ACT, *deflection*), which arises from conflicting institutional and situational meanings in a given interaction context. Although ACT is theoretically consistent, its propositions regarding cognitive and affective processing have rarely been tested. The present study fills this gap by investigating the neural processing of affective incongruency in linguistic descriptions of social interactions. Following a neurosociological paradigm, the study draws on neurocognitive evidence on the effects of words' affective content on word processing as well as on a previous electroencephalography study that investigated processing of affective incongruency using event-related brain potentials. We hypothesized that affective incongruency is associated with activation in the anterior cingulate cortex (ACC), a brain area known for conflict processing. To test this hypothesis, we replicated the electroencephalography study using functional magnetic resonance imaging: We visually presented sentences describing social interactions to 23 participants in a silent reading task while measuring differences in the hemodynamic response in two conditions of affective congruency. Results show expected increases in neural activity for affectively incongruent sentences in the left ACC, supporting the assumption that affective language content influences meaning-making at very early semantic processing stages. The results also add to the emerging neuroscientific evidence for ACT's mathematical model of impression formation.

Keywords Neurosociology · Sociology of emotion · Affective meaning · Situational meaning · fMRI · Symbolic interactionism

Konflikterkennung in der Sprachverarbeitung: ein Test der Theorie der Affektsteuerung zur Prognose neuronaler Korrelate affektiver Inkongruenz in der sozialen Interaktion

Zusammenfassung Die Theorie der Affektsteuerung (Affect Control Theory) ist eine soziologische Theorie zur Verarbeitung von Bedeutung in der sozialen Interaktion. Bedeutung speist sich darin sowohl aus institutionalisierten als auch aus situativen Gegebenheiten und umfasst denotative (deklarative) wie konnotative (affektive) Dimensionen. Die mathematische Formalisierung der Theorie erlaubt Vorhersagen der affektiven Kongruenz oder Stimmigkeit („affective deflection") einer Interaktionssituation, die sich aus dem Zusammenspiel institutionalisierter und situativer Bedeutungen ergibt. Obgleich theoretisch konsistent, sind die Thesen der Theorie zum Zusammenspiel kognitiver und affektiver Bestandteile von Bedeutung nur selten empirisch überprüft worden. Die vorliegende Studie schließt diese Lücke, indem sie die neuronale Verarbeitung affektiver Inkongruenz in der sprachlichen Darstellung sozialer Interaktionen untersucht. Einem neurosoziologischen Paradigma folgend, stützt sich die Studie auf neurokognitive Befunde zur Rolle der affektiven Konnotation von Wörtern in der visuellen Wortverarbeitung sowie eigene Befunde zur neurophysiologischen Verarbeitung affektiver Inkongruenz mittels ereigniskorrelierter Hirnpotenziale. Wir testen die Hypothese, dass die affektive Inkongruenz sozialer Interaktion mit der Aktivierung des anterioren cingulären Cortex (ACC) korreliert, einem Hirnareal, das an der Verarbeitung konfligierender Informationen beteiligt ist. Dazu replizieren wir ein bestehendes Experimentaldesign unter Nutzung funktioneller Magnetresonanztomographie (fMRT): Im Rahmen einer Leseaufgabe präsentieren wir 23 Teilnehmerinnen und Teilnehmern visuell Sätze, die soziale Interaktionen beschreiben, und messen Unterschiede in der hämodynamischen Reaktion unter zwei Bedingungen affektiver Inkongruenz. Die Ergebnisse zeigen die erwartete Zunahme der neuronalen Aktivität im linken ACC in der Bedingung hoher affektiver Inkongruenz. Dies stützt die Annahme, dass die affektive Konnotation von Begriffen die Genese von Bedeutung in sozialen Interaktionen bereits in frühen semantischen Verarbeitungsstufen bedingt. Die Befunde können zudem als neurowissenschaftliche Evidenz für die mathematische Formalisierung der Theorie der Affektsteuerung gelten.

Schlüsselwörter Neurosoziologie · Soziologie der Emotionen · Affektive Bedeutung · Situative Bedeutung · fMRT · Symbolischer Interaktionismus

1 Introduction

Over the past decades, a number of scientific disciplines have initiated collaborations with the neurosciences, including economics, philosophy, and psychology. Sociology, although a latecomer to the show, is no exception. The label “neurosociology” (as in neurophilosophy or neuroeconomics) reflects this collaboration and points at efforts of sociologists to better understand the social world in terms of human neurobiology (Kalkhoff et al. 2016). The neurosociological paradigm originated in sociological social psychology, in particular from symbolic interactionism as well as from evolutionary sociology, but it is also partly reflected in the works of sociologists in the fields of, for example, epidemiology, demography, health, genetics, and stratification (e.g., Falk et al. 2013; Hatzenbuehler et al. 2024; Sheridan 2023; Conley et al. 2014; see also von Scheve 2018).

Typically, neurosociology conceives of neural processes and mechanisms as specific levels of social reality that can be investigated using microsociological methods. One of the basic tenets of neurosociology is that the social sciences should take into account research in the neurosciences to advance and refine primarily microsociological concepts such as experience, mind, self, cognition, and emotion. In this regard, neurosociology’s major interest is in those branches of the neurosciences that aim at understanding the physiological, in particular neural but also neuroendocrine, immunological, developmental, and genetic foundations of social behavior and social psychological processes, and especially social, cognitive, and affective neuroscience (e.g., Schutt et al. 2015; Banich and Compton 2018; Armony and Vuilleumier 2013). Neurosociology therefore argues that many studies and research paradigms in these fields, as well as their hypotheses and results, are directly relevant for understanding the social processes and mechanisms traditionally studied by sociology (Franks 2010; see also von Scheve 2018).

Taking this approach one step further, neurosociology not only suggests incorporating neuroscientific findings into sociological analysis but also advocates for an active collaboration in the sense of using neuroscientific methods (as microsociological forms of inquiry) to address genuinely sociological questions. Taking social, cognitive, and affective neuroscience as a landmark, one of the key paradigms is to combine experimental methods from the social and behavioral sciences with medical imaging (e.g., functional magnetic resonance imaging [fMRI] or positron emission tomography) and brain mapping (e.g., electroencephalography [EEG] or magnetoencephalography) techniques, which may provide insights into the foundations of human social behavior and mental processes (see also von Scheve 2018). As of yet, however, most of the neurosociological literature relies on theoretical work that is primarily concerned with interpreting and adopting neuroscientific insights to inform existing sociological accounts of, for example, social interaction, inequality, or aging.

In this article, we build on the paradigm of neurosociology and use neuroimaging methods to investigate the neural correlates of affective conflict in the processing of social interactions. Our study draws on *affect control theory* (ACT), a sociological theory proposing that actors strive to maintain the basic affective meaning of social reality through their actions and behaviors. This theory provides formal models of the degree to which social interactions deviate from established routines based on the affective meanings of the concepts actors use to interpret (consciously or unconsciously) a social situation. Our study makes use of a dictionary we developed in previous work (Ambrasat et al. 2014; Schauenburg et al. 2015), which is—similar to established dictionaries for sentiment analysis—built from normative, out-of-context ratings of the affective meanings of a large number of words of the German language. These meanings allow us to build sentences describing expected (affective coherence) and unexpected (affective conflict) types of social interactions.

Using these sentences in experimental manipulations and employing fMRI, the study demonstrates that the affective meanings of concepts are essential in processing social interactions, above and beyond the denotative meanings actors use to interpret a situation. More specifically, our study shows that activity in the anterior cingulate cortex (ACC), a brain area commonly associated with automatic conflict processing, is correlated with affective conflict and coherence in processing sentences describing social interactions. In the following, we first provide a detailed account of ACT and the general approach of our study. We then provide detailed information on the methods, materials, and measures we used and present our results. We finally discuss our findings with respect to their broader implications.

2 Affect Control Theory

Very generally, ACT conceives of the social world as a web of institutions (e.g., the family, the economy, the educational system) and the various roles and identities that individuals occupy in relation to such institutions (MacKinnon and Heise 2010;

for the following description of ACT, see also Rogers et al. 2014). The institutional order is tied to individual minds through shared affective meanings of the concepts denoting roles and role identities (e.g., mothers, managers, teachers). Culture is thus understood as shared meaning-making within and across groups and institutions. Society shapes affective experiences when conceptual representations of social situations and their affective meanings are either confirmed or disconfirmed through specific actions or observations (Heise 2007).

Affect control theory has some of its intellectual roots in symbolic interactionist ideas about the creation, maintenance, and change of meaning through social interaction. According to this view, collective experiences with the social environment crystallize in shared linguistic symbols that structure individual minds. Thus, shared semantic structures reflect the institutional and symbolic order of societies (MacKinnon and Heise 2010). When individuals interact with each other in specific social situations, they rely on such semantic structures to generate a shared understanding of the situation. Linguistic categories thus provide the “common ground” (Clark 1996; see also Berger and Luckmann 1966) that allows for efficient communication and enables people to coordinate social action.

Affect control theory extends the classic symbolic interactionist paradigm in several ways: Conceptually, it assigns a central role to affect and emotion in the process of creating and maintaining meaning in social interaction. Methodologically, it holds that meaning can be quantified and measured, deviating from the predominantly interpretative and hermeneutic approaches in symbolic interactionism. Both of these amendments are rooted in Osgood’s (1962) proposition that the meanings of concepts can be measured along the perceptual dimensions of evaluation (good vs. bad), potency (strong vs. weak), and activity (active vs. passive); that these dimensions are universal across the human species; and that they reflect affective and connotative as opposed to denotative or declarative meanings. These dimensions constitute a three-dimensional “affective space” within which the affective meaning of any concept (e.g., persons, behaviors, situations, objects) can be located, and researchers have compiled numerous dictionaries of affective meanings in which thousands of concepts are represented by an evaluation–potency–activity profile acquired through rating studies (e.g., Võ et al. 2009; Ambrasat et al. 2014; Schauenburg et al. 2015).

Moreover, integrative theoretical work has shown how the principles of ACT can be applied to better understand the sociality of affect and emotion at multiple levels of analysis, based on individual differences and commonalities in the affective meanings of concepts. Rogers et al. (2014) have argued that ACT is, to a large extent, compatible with theories of affect and emotion focusing on these different levels, such as regarding cultural differences in affective meanings (Rogers 2019; Ambrasat et al. 2016), the role of affective meanings in social exchange (Clay-Warner et al. 2016), and their importance for the psychological (e.g., Lindquist and MacCormack 2014) and the neural (e.g., Pornpattananangkul and Chiao 2014) generation of emotion.

2.1 From Affective Meanings to Social Interactions

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) in impression formation, that is, of all the connotations of the concepts used to linguistically represent a given social situation. Similar to other cognitive-consistency theories (e.g., Heider 1946), one basic assumption of ACT is that people strive to maintain the affective meanings of the concepts in their mental representations and social actions. Thus, situations for which affective meanings of involved concepts match each other easily integrate into our stream of perception and action, whereas we mentally stumble over events that are represented by concepts whose emotional connotations do not fit together and produce what ACT calls "deflection," a type of affective conflict. For instance, in the situation "A mother plays with a child," the affective meaning of the concept "mother" almost perfectly matches the affective connotations of the words "play" and "child." However, the affective 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 concept "mother" because of the incongruency between its common ("fundamental") sentiment and its situational, transient affective meaning.

This mismatch, or affective deflection, would encourage individuals to somehow "rebalance" the lack of perceived coherence—for instance, by postulating that the child was badly misbehaving before the beating or that the beating happened accidentally. Fundamental sentiments thus are trans-situative, socially shared affective meanings that are comparatively stable and resistant to change (Robinson et al. 2006, p. 182), whereas transient sentiments denote situational and dynamic affective meanings that may, but need not, deviate from fundamental sentiments (Robinson et al. 2006). In the framework of ACT, affective deflection can be modeled mathematically using impression-formation equations that 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 ratings of the same words (e.g., Averett and Heise 1987; Schröder 2011).

Although this mental stumbling over affectively incoherent social situations is theoretically well elaborated and has been empirically modeled (e.g., Schröder and Scholl 2009), its underlying principles and mechanisms, the way it is processed physiologically in situational meaning-making, and the way it relates to what we typically look at in analyses of social situations—i.e., conflict and incongruency in declarative meanings—are much less understood. In order to address this gap, we devised an experimental study in which we used brain-imaging methods to better understand the physiological—in particular, neural—bases of processing affectively incongruent social situations.

3 The Present Study

Existing neuroscientific research on the emotion–language relation suggests that affective language content influences the processing of meaning, perception, and behavior. Regarding isolated words in visual language processing, studies have demonstrated the relevance of words' affective meaning in a variety of tasks and on a range of behavioral and neuropsychological measures. These results support the general assumption that rapid, automatic neurophysiological responses differentiate the processing of words' affective meaning, for example regarding affectively pleasant and unpleasant words or emotion words (Bernat et al. 2001; Citron 2012; Zhang et al. 2014).

Investigating the effects of words' affective meanings at the level of sentences and phrases is significantly more challenging, not least because of a lack of theoretical models concerning the interplay of the affective meanings of multiple words in phrase or sentence contexts. In a previous study, we therefore investigated these questions, using ACT's mathematical model of impression formation (and deflection), by looking at event-related potentials (ERP), i.e., voltage fluctuations of certain brain structures in response to stimuli, measured using EEG (Schauenburg et al. 2019). Specifically, we focused on the coherence of words' affective meanings in sentence contexts describing social interactions.

Carefully controlled for confounding variables of semantic processing, this prior study found an increased negativity for affectively incongruent compared to congruent final words of sentences describing social interactions in the N2/P2 (130–270 ms) and N400 (300–450 ms after target word onset) components of the EEG signal. These components, or ERP, are known to be associated with conflict detection, for example in perceptual novelty and mismatch (N2/P2) and with semantic incoherence, i.e., semantic, expectancy, and world-knowledge violations (N400). These previous findings suggest that conflicting affective meanings (or *affective deflection*, in ACT terminology) are part of early semantic processing operating largely outside conscious awareness.

Although these results are informative regarding the temporal processing of affective conflict, they say little about the specific brain areas in which conflict is processed. This is because EEG signals have a high temporal but low spatial resolution. However, learning more about the specific brain areas involved in processing affective conflict is imperative for a more comprehensive understanding of the physiological foundations of affective conflict processing.

Existing neuroimaging research evaluating, for example, basic perceptual processing and the rather complex processing of moral judgment has shown that the ACC is related to error detection, conflict monitoring, expectancy-related processes, and affective experience (Botvinick et al. 2004; Braem et al. 2017; Carter et al. 1998; Greene et al. 2004; Mansouri et al. 2009). The ACC is well connected to limbic structures and prefrontal regions of the brain (Stevens et al. 2011). The former are predominantly associated with affective and automatic processing, while the latter are mainly associated with executive control, as in decision-making, attention allocation, and information integration and updating.

Therefore, the ACC is well suited to monitor cognitive and affective information inflow from limbic structures and to recruit prefrontal areas for more deliberate processing (e.g., the dorsolateral prefrontal cortex; Greene et al. 2004), such as emotion regulation, social cognition, and decision-making when conflict is detected. The ACC might thus be particularly relevant for the neuropsychological foundations of ACT because recruitment of these (and other) prefrontal brain areas is essential for the reappraisal and reinterpretation of situations when conflicting information is detected. It could therefore be vital for a neuropsychological foundation of ACT's theoretical predictions. In the present study, we aimed to test the hypothesis that conflicting affective meanings in sentence processing correlate with increased activation of the ACC. To test this proposition, we used the materials from our previous EEG study (Schauenburg et al. 2019) and employed them in an event-related fMRI design. Functional MRI is a brain-imaging method that detects blood oxygen level–dependent changes arising from changes in neural activity in different brain areas, such as when the individual is engaged in a specific task or is presented with perceptual stimuli (e.g., Gore 2003).

4 Materials and Methods

4.1 Participants

Twenty-seven German native speakers gave written consent and participated in this study, which was conducted in accordance with the Declaration of Helsinki and had been approved by the ethics committee of Freie Universität Berlin. All participants were right-handed (German adaptation of Oldfield [1971]), with normal or corrected-to-normal vision and no history of psychological or neurological diseases. Participation was monetarily rewarded (8 €/hour). Data of four participants were excluded due to too much movement during fMRI data acquisition (more than 5 mm in any orientation) or due to too many wrong answers to attention questions (more than 16, i.e., one-third wrong answers). The final set included data from 23 subjects (mean age 24.70 years, range 19–34; 12 women, 11 men).

4.2 Stimuli

Experimental stimuli were sentences describing social interactions in a basic actor–behavior–object manner. That is, all sentences conformed to the structure *determinant–subject–verb–(preposition)–determinant–object*. These sentences were built from existing dictionaries that are commonly used in research on ACT and that include sentence elements (words) for which information on affective meaning is available on the dimensions of evaluation, arousal, and potency (654 subject words, 275 verbs, 400 object words; Võ et al. 2006; Schmidtke et al. 2014; Schauenburg et al. 2015, Ambrasat et al. 2014). We first permutated all possible sentence elements and then calculated the affective deflections for the approximately 72 million resulting sentences using regression equations for impression formation (Schröder 2011).

In our previous ERP study (Schauenburg et al. 2019), we allocated sentences to three categories of affective deflection: low, medium, high. To limit costly fMRI scanning time in the present study, we omitted sentences from the medium deflection category, only using sentences from the low and high deflection categories. The final stimuli sentences were then selected so that sentences in both categories included the same sentence contexts (i.e., combinations of subjects and verbs) and final words (i.e., objects). This implies, first, that sentences in the high and low deflection categories differed only in how the words were combined to form the sentences, resulting in either high or low average deflection. Second, this ensures that the contextual constraints (i.e., the expectancies induced by sentence contexts) were identical across conditions. In sentences with low contextual constraints, for example, subject and verb combinations can be followed by a variety of words in a meaningful way, whereas high contextual constraints limit the number of plausible (or expected) words to just a few.

Specifically, the stimuli sentences were balanced across high and low deflection conditions with regard to variables known to influence language processing, such as cloze probability and frequencies of subject-object as well as of verb-object of co-occurrences. A word's cloze probability refers to the proportion of subjects who would use this particular word to complete an unfinished sentence (Kutas and Hillyard 1984). This ratio can be low, moderate, or high, and the violation of the expectancy of certain responses is associated with specific neural correlates of semantic processing (Kutas and Hillyard 1984). This way, we ensured that our stimuli sentences differed only in the affective deflection of entire sentences and that sentences' final target words had the same expectancy across the two deflection conditions.

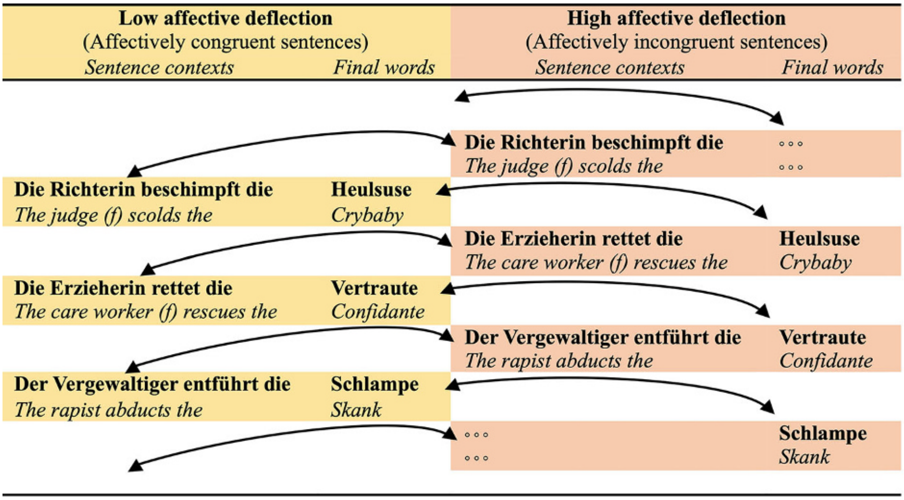

Fig. 1 Illustration for a selection of stimuli sentences showing how every sentence context (subject and verb combinations) and final word (object) are present in the high and low deflection conditions (as indicated by the bidirectional arrows). Original stimuli are presented in bold, with translations in italics

Table 1 Overview of study design and stimuli example

	Experimental sentences		Control threads	
	Low deflection	High deflection	Without @	With @
Stimulus example	The hussy chokes the beast	The hussy chokes the fairy	### ###### ###### ### #########	### ###### ##@### ### #########
N per run	26 (27)	27 (26)	10 (11)	11 (10)
N total	*106*	*106*	*42*	*42*
Question example	Did the hussy strangle the beast? (yes)	–	Was there an @? (no)	–
N yes, total	6	6	6	6
N no, total	6	6	6	6
N total, total	12	12	12	12

Applying all these criteria resulted in 212 stimulus sentences (106 per condition of high vs. low affective deflection). To ensure attentive reading, we included an attention question. To later analyze whether brain areas associated with semantic processing were actually active in the experimental condition, we also added a non-reading control condition in which we presented participants with hash signs (#). A total of 84 hash-sign threads were matched with the experimental conditions, ensuring that the length of single hash-sign strings matched the word length, and hash-sign threads were matched to sentence length, typically containing five or six words. To ensure better comparability with the experimental sentence reading condition, we also included an attention check in the control condition: Participants were instructed that, periodically, they would be asked if an "@" sign had been presented within the hash-sign strings. Consequently, the control nonreading condition included hash-sign threads alternately with and without "@" signs, accompanied by attention questions that were matched with the experimental sentence reading condition (see Fig. 1 for examples, and Schauenburg et al. 2019 for full details on stimulus generation). Table 1 offers an overview of the general study design and characteristics of stimulus sentences.

Since it is well established that an interaction's perceived likelihood is correlated with affective deflection (Heise 2010), we also collected likelihood ratings of the social interactions described in the experimental sentences for further analyses and as a manipulation check. After fMRI scanning, participants were asked to rate the likelihood of the described social interactions of each of the 212 sentences presented in random order on a seven-point Likert scale (from "highly unlikely" to "highly likely").

4.3 Procedure

Sentences and control conditions (hash-sign threads) were visually presented in the same manner. A trial started with the 250-ms presentation of a fixation cross followed by a 100-ms blank screen. Sentences and control conditions were then presented in a word-by-word fashion: Each word or hash-sign string with less than nine letters (hash signs) was presented for 250 ms, and words with more than nine letters (hash signs) were presented for 300 ms to ensure complete perception (controlled across runs). Since both experimental conditions included the same word stimuli, the presentation time of every word was perfectly balanced across conditions. There were 100-ms blank screens between words (hash signs), which improved smooth reading perception (as indicated by a pilot test). Each stimulus was presented in the center of the screen.

The interstimulus interval between sentences and hash-sign threads was optimized with software, jittering[1] a mean 2000-ms interval. To keep participants attentive following the trials, yes/no questions (balanced for type of correct answer) were randomly assigned and had to be answered by a button press (index finger = yes, middle finger = no). There were 12 attention questions for each experimental condition, which focused on the preceding sentence and used synonyms to make answering sufficiently challenging. In control trials, 12 attention questions were randomly presented. Jittering order and stimulus presentation order across conditions (low affective deflection, high affective deflection, control condition, attention question) were optimized to ensure a maximal signal-to-noise ratio.

To ensure that participants were familiar with the experimental setting and task, they began with a practice run. This run consisted of eight trials, five sentence trials, and three control trials, each followed by an attention question. Emoticons gave feedback to participants' responses: Frowning emoticons indicated a wrong answer, and smiling ones indicated a right answer. Feedback was given only during practice trials. We used the software Presentation (Neurobehavioral Systems, Berkeley, CA, USA) to present stimuli and record participants' answers to attention questions. Practice runs were followed by four experimental blocks, in between which participants could take a break. After scanning, participants gave likelihood ratings on each sentence.

4.4 Design

Consistent with our previous study (Schauenburg et al. 2019), we employed an event-related fMRI design. The experiment consisted of four runs, each containing 74 trials (26 resp. 27 sentences of low affective deflection; 26 resp. 27 sentences of high affective deflection, 21 control trials) and 12 attention questions; assigned trials were randomized within each run, and the running order was randomized across subjects. Total scanning time took about 60 min. Table 1 presents an overview, and Fig. 2 illustrates the sequencing of a sentence reading trial.

4.4.1 fMRI Data Acquisition

Functional data were acquired on a Siemens Tim Trio 3 T MR scanner (Siemens Healthineers AG, Forchheim, Germany). Four runs of 282 volumes were measured

---

[1] Jittering refers to random variations in the time between successive stimuli and responses (https://surfer.nmr.mgh.harvard.edu/optseq/).

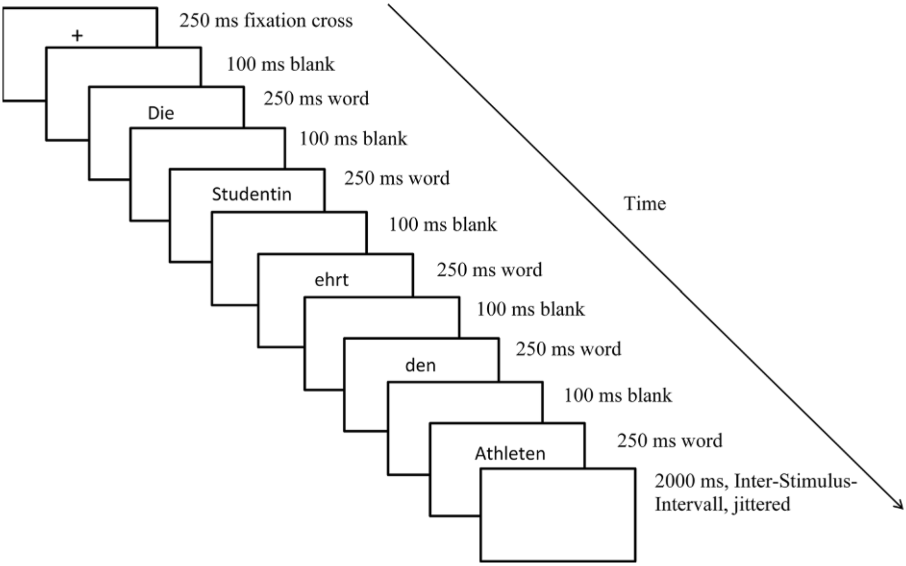

Fig. 2 Schematic presentation of a sentence reading trial

using a T2*-weighted echo-planar sequence (slice thickness 3 mm, no gap; 37 slices; repetition time [TR] 2 s; echo time [TE] 30 ms; flip angle 70°; matrix 64 × 64; field of view [FOV] 192 mm; voxel size 3.0 mm × 3.0 mm × 3.0 mm), and individual high-resolution T1-weighted anatomical data (MPRAGE sequence) were acquired (TR 1.9 s; TE 2.52 s; FOV 256 mm; matrix 256 mm × 256 mm; sagittal plane; slice thickness 1 mm; 176 slices; resolution 1.0 mm × 1.0 mm × 1.0 mm).

4.4.2 fMRI Data Preprocessing

Imaging data were preprocessed and analyzed using the software package SPM v12 (www.fil.ion.ucl.ac.uk/spm). Preprocessing covered slice-timing correction, realignment for motion correction, and sequential coregistration. Structural images were segmented into grey matter, white matter, cerebrospinal fluid, bone, soft tissue, and air/background using the "new segment" module (Ashburner and Friston 2005). A group anatomical template was created with the DARTEL (Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra; Ashburner 2007) toolbox from the segmented grey and white matter images. Transformation parameters for structural images were then applied to functional images to normalize them to the Montreal Neurological Institute brain template supplied with SPM software. Functional images were resampled to a resolution of 1.5 × 1.5 × 1.5 mm and spatially smoothed with a kernel of 6 mm full-width-at-half-maximum during normalization.

4.4.3 fMRI General Linear Model Analysis

We realized statistical parametric maps by multiple regressions of the data onto a model of the hemodynamic response. In the subject-level analysis, this model contained regressors for both experimental and control conditions. Regressors were convolved with the canonical hemodynamic response function in SPM software. Beta images of each of the four conditions in the factorial design for each participant were used at the group level to a random-effect paired t-test analysis (Holmes and Friston 1998) to explore potential differential correlations between the two experimental conditions (HighDeflection > LowDeflection) and between reading and nonreading conditions (AllSentences > AllThreads). Thus, we contrasted experimental conditions vs. control conditions ([HighDeflection + LowDeflection] > [HashSignWith@ + HashSignWithout], in short: AllSentences > AllThreads) at the whole-brain level to test whether the experiment successfully tapped into the language-processing network.

Whole-brain fMRI analyses included a cluster-forming threshold for uncorrected $p < 0.001$, and then a cluster-level threshold of family-wise error (FWE)–corrected $p < 0.05$ for the entire image volume. For our experimental manipulation (HighDeflection > LowDeflection), we also performed a small volume correction with a bilateral ACC mask (BA32 mask defined by the WFU PickAtlas tool; Maldjian et al. 2003). In the region of interest analyses, we applied an initial voxel-level threshold of uncorrected $p < 0.001$ for the entire image, followed by the threshold of voxel-level FWE–corrected $p < 0.05$ after applying a small volume correction with the bilateral ACC mask. The statistical thresholds for our fMRI analysis, including a cluster-forming threshold of uncorrected $p < 0.001$ and a subsequent cluster-level threshold corrected for FWE at $p < 0.05$, were carefully selected to balance sensitivity and specificity in detecting meaningful neural activations. These thresholds align with established practices in the field of neuroimaging and have been widely employed in similar studies, ensuring consistency with the existing literature (see Flandin and Friston 2019).

4.5 Postscan Sentence Analysis: Likelihood Ratings

For further analyses and as a manipulation check of affective deflection, we gathered likelihood ratings of the social interactions described in our stimulus sentences after scanning had been completed. Participants were asked to rate each sentence regarding the question “How likely do you rate the described interaction?” from “not at all” to “extremely” on a seven-point-Likert scale. Mean likelihoods were computed across all 106 sentences separated for each condition. Obtained likelihood means were compared using t-tests.

5 Results

5.1 Behavioral Performance

All 23 participants included in the final data set correctly responded to attention questions above chance (≥ 66%), with a mean accuracy of 86.77%.

5.2 fMRI Results

The comparison between experimental (AllSentences) and control (AllThreads) conditions showed activations in brain areas canonically associated with visual language and semantic processing. Small volume correction with the bilateral ACC mask showed increased neural activity in the left ACC (Brodmann area 32) for the high compared to the low affective deflection condition. Table 2 presents an overview of the results, Fig. 3 shows patterns of brain activation for contrasts between experimental and control conditions, and Fig. 4 shows brain activation in the ACC for the contrast between high and low affective deflection conditions.

5.3 Likelihood Ratings

We measured the likelihood of the described social interactions in the experimental sentences stimuli on the ratings given by the included 23 participants. A t-test comparison confirmed the well-known deflection-likelihood relation (Heise 2010); that is, interactions with higher affective deflection (arithmetic mean [AM] = 3.88, standard deviation [SD] = 0.91; $n = 106$) are perceived as less likely than those with lower affective deflection (AM = 4.53, SD = 0.91; $n = 106$; T(210) = 5.163; $p < 0.001$). Importantly, due to our tightly controlled sentence stimuli, this relation cannot be

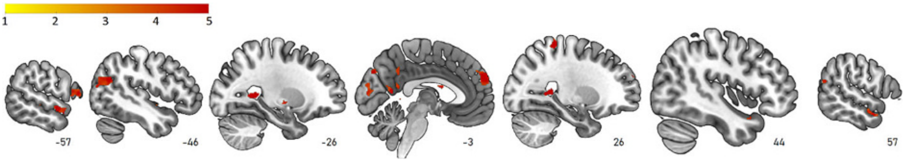

Fig. 3 Brain activation for the contrast of experimental vs. control conditions (AllSentences > AllThreads)

Fig. 4 Brain activation for region of interest analysis in the anterior cingulate cortex for contrast of high vs. low affective deflection conditions

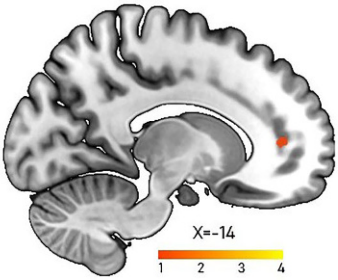

Table 2 Montreal Neurological Institute (MNI) coordinates and statistics for functional magnetic resonance imaging analyses. The reported x, y, z are related to peak coordinates. Depending on the analysis, they relate to either the whole brain or the region of interest analyses

Anatomical definition		MNI			Cl_size	Peak T
		x	y	z		
AllSentences > AllThreads (whole-brain corrected)						
R	Limbic lobe (hippocampus)	26	–44	6	521	7.22
L	Basal ganglia (caudate tail)	–15	–29	20	687	7.03
LR	Dorsolateral and medial prefrontal cortex (BA9)	–5	56	30	1554	7.00
L	Lateral temporal lobe (BA38, 22, 21)	–57	6	–15	728	6.80
R	Sensory association cortex (BA5)	21	–41	65	534	6.71
R	Lateral temporal lobe (BA38, 21)	51	8	–27	447	6.17
L	Angular area (BA 39, 19)	–47	–60	23	840	6.02
L	Basal ganglia (caudate body)	–9	18	9	491	5.97
L	Inferior frontal gyrus (BA45, 44)	–56	29	15	335	5.86
L	Ventral posterior cingulate area (BA23)	–3	–57	11	2362	5.75
R	Angular area (BA39)	56	–60	21	152	5.17
L	Basal ganglia (globus pallidus)	–27	–11	–5	145	5.08
HighDeflection > LowDeflection (small volume corrected)						
L	Anterior cingulate cortex (BA32)	–14	–44	8	14	4.72

Significant peak voxel for all clusters at $p < 0.05$ family-wise error–corrected; names of anatomical definition and x, y, z coordinates are given for first brain region of relevant cluster; in parentheses, involved brain regions are listed corresponding to Brodmann areas

Cl_size cluster size, *L/R* left/right, *BA* Brodmann area explained by other lexicosemantic variables known to influence expectancy-related processes.

6 Discussion

The aim of this article was to illustrate the viability of a neurosociological approach to better understand the neurophysiological foundations of processing conflicting affective meanings in linguistic descriptions of social interactions. Affect control theory proposes that individuals make sense of social situations and interactions based on the affective meanings of concepts used to mentally construe these situations. More specifically, ACT assumes that actors strive for consistency in the overall affective meaning of a given social situation and that affectively incoherent situations produce affective deflection, which lets actors mentally stumble across a situation and motivates them to act toward consistency in meanings.

Our study draws on neurocognitive evidence regarding the implications of words' affective meanings for the neural correlates of word processing and on our own previous work providing evidence for early and rapid processing of affective conflict measured by event-related brain potentials (Schauenburg et al. 2019). To gain a better understanding not only of the temporal processing of conflicting affective meanings in social interactions but also of the functional properties of conflict processing in specific brain areas, the present study employed an fMRI design and measured hemodynamic responses to visually presented sentences describing social interactions in two conditions of affective deflection. We hypothesized that the previously demonstrated early ERP effect should translate to the hemodynamic response and be associated with activation in the ACC, a brain area well known for conflict processing.

To ensure that our findings are actually related to semantic processing (instead of just visual processing), our design included attention questions and a control condition in which we visually presented controlled strings of hash signs. As a manipulation check for affective deflection, we acquired likelihood ratings for the social interactions described in our experimental sentences. Both behavioral and fMRI data support the assumption that participants were semantically engaged during sentence processing. Attention questions were answered at above-chance levels, and the results of whole-level brain contrasts were in line with the prevailing literature regarding language comprehension and semantic processing (e.g., Binder et al. 2009; Binder 2016; Federmeier et al. 2008; Ferstl et al. 2008).

With regard to our experimental manipulation, we expected reading sentences in the high compared to the low affective deflection conditions to correlate more strongly with hemodynamic responses in the ACC. As hypothesized, the ACC showed increased neural activity for sentences with high as compared to low affective deflection. Our study therefore confirms previous findings in supporting ACT's mathematical model of impression formation of linguistic representations of social interactions, and it suggests that affective deflection is related to activity in the ACC, even in the absence of explicit task demands. This supports the assumption of fast, implicit, and automatic processing of affective deflection.

Although our study was not explicitly designed to further probe the basic properties and (functional) implications of ACC activation, it motivates discussion of how ACC activity in response to affective deflection might be explained, alternatively, in terms of its involvement in prediction error signaling. Alexander and Brown (2019) presented a computational model of hierarchical error representation, rooted in the tradition of reinforcement learning models, suggesting that *surprise* as a (prediction) error is a key component in the formalization of association learning. The model assumes a hierarchical organization of frontal lobe processes in which error signaling is driven by bottom-up processes and prediction error modeling by top-down processes. By proposing that the ACC is not sensible, in the first place, to the affective components of a stimulus, but rather, and more generally, to surprise (i.e., the discrepancy of the occurrence or nonoccurrence of an expected or unexpected event), it aims at providing a comprehensive and unifying approach to ACC activation across various studies.

The pattern of ACC activation we found can thus also be explained as an error signal of predicted affective congruency induced by sentences' final words. Crucially, the concept of affective deflection shows considerable overlap with the concept of error prediction because it represents the degree to which an expected affective meaning is violated by an acutely perceived meaning. Similar to the concept of prediction error processing, the magnitude of affective deflection is an experienced-based correlate of the discrepancy between a predicted probability and an actually observed event. This interpretation also speaks to more recent predictive processing theories of affect and emotion, in which prediction error signaling and reduction play a decisive role such that interoceptive prediction signals (and errors) may initiate changes in affect (Barrett 2017, p. 9). According to this view, unanticipated information, such as affective deflection producing prediction errors, serves as feedback for "embodied simulations" (Barsalou 2008) that allow for predictions (e.g., regarding the course of a social interaction) to be made in the first place. Error signals, as suggested by the patterns of activation in the ACC, monitor differences between predicted and actual sensations (Barrett 2017, p. 7). How continuous changes in affect correspond to discrete emotions (e.g., fear, anger, guilt) has been outlined in detail by psychological constructionist accounts of emotion, which, in turn, have been shown to be compatible in many respects with ACT (Rogers et al. 2014; Lindquist and MacCormack 2014).

Furthermore, ACT postulates that impressions of affectively deflecting events would initiate cognitive and behavioral processes aimed at regaining an experience of affective congruency (Robinson et al. 2006). In the hierarchical error representation model, corresponding neural processes should be located at higher-order levels, top-down regulating areas in the prefrontal cortex. Rather than invalidating our findings and conclusions, we suggest that both models in fact complement one another with respect to the computational modeling of precise and testable patterns of brain activation and connectivity and to corresponding bottom-up and top-down processes in the frontal lobe.

Our study also showed increased activity only in the left ACC. Previous studies on conflict detection provide inconsistent findings concerning each hemisphere's differential contributions to overall ACC activity (Lütcke and Frahm 2008). Crucially, Stephan et al. (2003) found that cognitive control is localized in the same hemisphere as task execution: In their study, the left hemisphere was more involved in a letter decision, whereas the right hemisphere was more involved in a visuospatial decision task. With regard to the present study, this finding would explain activation of the left ACC based on the reading task being associated with activity in the left hemisphere.

Nonetheless, Lütcke and Frahm (2008) found that differential hemispheric involvement was related to differences between cognitive processes of conflict detection vs. error processing, rather than to task-induced lateralized processing. In this regard, Ochsner et al. (2009) reported increased activation in the left ACC in response to conflicting stimuli in a cognitive flanker task, whereas the right ACC was more involved in an affective flanker task. However, going into more detail concerning the applied tasks and the salience of emotionality (implicit processing of affective congruency in a reading task vs. an explicit categorization task for affective words), the comparison of reported findings of lateralized ACC involvement seems to be easier said than done. Future studies should therefore further investigate differential lateralization of ACC activity in different task and stimulus settings to overcome these inconsistencies.

Three limitations of our study need to be mentioned. First, hash signs might not be an ideal cognitive control condition in this design. They served as visually comparative stimuli, but the kinds of cognitive processes participants might recruit when perceiving hash signs was not tightly controlled (this also concerns the comparability of processing demands related to task difficulty; see Binder et al. 2009). Second, a combined EEG–fMRI study with the same group of participants is likely to provide more detailed information about the temporal evolvement of cognitive processes and their neural correlates. Third, individual differences in social cognition (e.g., trait empathy trait, theory of mind, or mentalizing abilities), personality traits, or language competencies should be accounted for in future studies.

The implications of our findings for ACT are threefold. First, the results help to better understand the concept of deflection, which until recently (e.g., Money 2023) had not received considerable attention. Our study provides evidence that different levels of affective deflection, mathematically derived from impression formation equations and existing out-of-context ratings of evaluation, potency, and activity, are correlated with increased activity in brain areas commonly associated with conflict processing. This lends support to the view that deflection is not simply an abstract, conceptual aide for ACT but is indeed a relevant psychological process with identifiable neurophysiological correlates.

Second, and going back to the roots of ACT in symbolic interactionism, our findings show how deviations from what individuals expect to typically occur in social interactions (as characterized by, for example, specific identities and role relations) are processed at the neurophysiological level. Importantly, this is contingent on the match or mismatch between transient (i.e., situational) and fundamental (learned, internalized) sentiments and affective meanings. Although there seems to be broad consensus within societies regarding the affective meanings of many socially relevant concepts, there is also systematic variation along indicators of culture and social stratification (Ambrasat et al. 2014; Ambrasat et al. 2016). Although there is no direct evidence, our research indicates that the social and cultural contexts and living conditions of, for example, social class, lifestyles, or social milieus bring about differences not only in cognitive and mental structures but also in neurophysiological processing.

Third, our study very generally speaks to existing research that has aimed to bring neuroscientific perspectives to ACT. This includes, for example, attempts at developing a neuroscientifically informed model of the relation between affect and cognition in ACT (MacKinnon and Hoey 2021), steps toward a neural computational model and implementation of ACT (Malhotra et al. 2020), and computational models of emotion that conceptualize the neural mechanisms involved in emotion processing as "semantic pointers," that is, patterns of neural firing binding neural representations of physiological input, evaluations of situations, and cultural context (Kajić et al. 2019; Thagard and Schröder 2015).

Some broader implications of our findings, in particular from a sociological and neurosociological standpoint, concern theories that deal with implicit, automatic, embodied, prereflexive, and unconscious forms of information processing and behavior. This can be found, for instance, in dual-process theories of action (e.g., Vaisey 2009), in nondeclarative modes of culture (in action; e.g., Lizardo 2021), or, classically, in Bourdieu's ideas of the "incorporation" of practices and fields and its relevance for habitualized forms of action.

Taken together, our study showed, first, that the affective meanings of concepts—above and beyond their declarative meanings—are notably implicated in the processing of simple social situations (actor–behavior–object) and that affective deflection (i.e., affectively incoherent situations) recruits similar neural circuits as the processing of other types of conflicting information, such as semantic or perceptual conflict. The findings provide support for some of the key assumptions (and methodologies) of ACT and show that collaborations between sociology and the neurosciences can be fruitful for both disciplines.

Funding Open Access funding enabled and organized by Projekt DEAL.

Conflict of interest G. Schauenburg, A. Aryani, C.-T. Hsu, T. Schröder, M. Conrad, C. von Scheve, and A.M. Jacobs declare that they have no competing interests.

References

Alexander, William H., and Joshua W. Brown. 2019. The Role of the Anterior Cingulate Cortex in Prediction Error and Signaling Surprise. *Topics in Cognitive Science* 11:119–135.

Ambrasat, Jens, Christian von Scheve, Markus Conrad, Gesche Schauenburg and Tobias Schröder. 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.

Ambrasat, Jens, Christian von Scheve, Gesche Schauenburg, Markus Conrad und Tobias Schröder. 2016. Unpacking the habitus: Meaning making across lifestyles. *Sociological Forum* 31(4):994–1017.

Armony, Jorge, and Patrik Vuilleumier (eds.) 2013. *The Cambridge Handbook of Human Affective Neuroscience.* New York: Cambridge University Press.

Ashburner, John. 2007. A fast diffeomorphic image registration algorithm. *NeuroImage* 38:95–113.

Ashburner, J., and Karl J. Friston. 2005. Unified segmentation. *NeuroImage* 26:839–851.

Averett, Christine, and David R. Heise. 1987. Modified social identities: Amalgamations, attributions, and emotions. *Journal of Mathematical Sociology* 13:1–2.

Banich, Marie T., and Rebecca J. Compton. 2018. *Cognitive Neuroscience*. New York: Cambridge University Press.

Barrett, Lisa Feldman. 2017. The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience* 12(1):1–23.

Barsalou, Lawrence W. 2008. Grounded cognition. *Annual Review of Psychology* 59:617–645.

Berger, Peter, and Thomas Luckmann. 1966. *The social construction of reality*. Garden City, NY: Anchor.

Bernat, Edward, Scott Bunce and Howard Shevrin. 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.

Binder, Jeffrey R. 2016. fMRI of Language Systems. In *fMRI Techniques and Protocols,* ed. Massimo Filippi, 355–385. New York: Springer.

Binder, Jeffrey R., Rutvik H. Desai, Willam W. Graves and Lisa L. Conant. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex* 19:2767–2796.

Botvinick, Matthew M., Jonathan D. Cohen and Cameron S. Carter. 2004. Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences* 8:539–546.

Braem, Senne, Joseph A. King, Franziska M. Korb, Ruth M. Krebs, Wim Notebaert and Tobias Egner. 2017. The Role of Anterior Cingulate Cortex in the Affective Evaluation of Conflict. *Journal of Cognitive Neuroscience* 29:137–149.

Carter, Cameron S., Todd S. Braver, Deanna M. Barch, Matthew M. Botvinick, Douglas Noll and Jonathan D. Cohen. 1998. Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 280:747–749.

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

Clark, Herbert H. 1996. *Using language.* New York: Cambridge University Press.

Clay-Warner, Jody, Dawn T. Robinson, Lynn Smith-Lovin, Kimberly B. Rogers and Katie R. James. 2016. Justice Standard Determines Emotional Responses to Over-Reward. *Social Psychology Quarterly* 79:44–67.

Conley, Dalton, Jason Fletcher and Christopher Dawes. 2014. The emergence of socio-genomics. *Contemporary Sociology* 43:458–467.

Falk, Emily B., Luke W. Hyde, Colter Mitchell, Jessica Faul, Richard Gonzalez, Mary M. Heitzeg, Daniel P. Keating, Kenneth M. Langa, Meghan E. Martz, Julie Maslowsky, Frederick J. Morrison, Douglas C. Noll, Megan E. Patrick, Fabian T. Pfeffer, Patricia A. Reuter-Lorenz, Moriah E. Thomason, Pamela Davis-Kean, Christopher S. Monk and John Schulenberg. 2013. What is a representative brain? Neuroscience meets population science. *Proceedings of the National Academy of Sciences of the USA* 110(44):17615–17622.

Federmeier, Kara D., Edward W. Wlotko and Aaron M. Meyer. 2008. What's "right" in language comprehension: Event-Related Potentials reveal right hemisphere language capabilities. *Language and Linguistics Compass* 2:1–17.

Ferstl, Evelyn C., Jane Neumann, Carsten Bogler and D. Yves von Cramon. 2008. The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping* 29:581–593.

Flandin, Guillaume, and Karl J. Friston. 2019. Analysis of family-wise error rates in statistical parametric mapping using random field theory. *Human Brain Mapping* 40:2052–2054.

Franks, David D. 2010. *Neurosociology*. New York: Springer.

Gore, John C. 2003. Principles and practice of functional MRI of the human brain. *Journal of Clinical Investigation* 112(1):4–9.

Greene, Joshua D., Leigh E. Nystrom, Andrew D. Engell, John M. Darley and Jonathan D. Cohen. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44:389–400.

Hatzenbuehler, Mark L., Katie A. McLaughlin, David Weissman and Mina Cikara. 2024. A research agenda for understanding how social inequality is linked to brain structure and function. *Nature Human Behavior* 8:20–31.

Heider, Fritz. 1946. Attitudes and cognitive organization. *The Journal of Psychology* 21:107–112.

Heise, David R. 2007. *Expressive Order: Confirming Sentiments in Social Actions*. Boston, MA: Springer.

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

Holmes, A.P., and Karl J. Friston. 1998. Generalisability, Random Effects & Population Inference. *NeuroImage* 7.

Kajić, Ivana, Tobias Schröder, Terrence C. Stewart and Paul Thagard. 2019. The semantic pointer theory of emotion: Integrating physiology, appraisal, and construction. *Cognitive Systems Research* 58:35–53.

Kalkhoff, Will, Shane R. Thye and Joshua Pollock. 2016. Developments in Neurosociology. *Sociology Compass* 10:242–258.

Kutas, Marta, and Steven A. Hillyard. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307(5947):161–163.

Lindquist, Kristen A., and Jennifer K. MacCormack. 2014. Comment: Constructionism is a multilevel framework for affective science. *Emotion Review* 6:134–135.

Lizardo, Omar. 2021. Culture, Cognition, and Internalization. *Sociological Forum* 36:1177–1206.

Lütcke, Henry, and Jens Frahm. 2008. Lateralized anterior cingulate function during error processing and conflict monitoring as revealed by high-resolution fMRI. *Cerebral Cortex* 18:508–515.

MacKinnon, Neil J., and David R. Heise. 2010. *Self, identity, and social institutions*. New York: Palgrave.

MacKinnon, Neil J., and Jesse Hoey. 2021. Operationalizing the Relation Between Affect and Cognition With the Somatic Transform. *Emotion Review* 13(3):245–256.

Maldjian, Joseph A., Paul J. Laurienti, Robert A. Kraft and Jonathan H. Burdette. 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage* 19:1233–1239.

Malhotra, Aarti, Terrence Stewart and Jesse Hoey. 2020. A Biologically-Inspired Neural Implementation of Affect Control Theory. In *Proceedings of the 18th International Conference on Cognitive Modeling*. https://iccm-conference.neocities.org/2020/papers/Contribution_259_final.pdf

Mansouri, Farshad A., Keiji Tanaka and Mark J. Buckley. 2009. Conflict-induced behavioural adjustment: a clue to the executive functions of the prefrontal cortex. *Nature Reviews Neuroscience* 10:141–152.

Money, Victoria. 2023. Demonstrating Anticipatory Deflection and a Preemptive Measure to Manage It: An Extension of Affect Control Theory. *Social Psychology Quarterly* 86(2):151–169.

Ochsner, Kevin N., Brent Hughes, Elaine R. Robertson, Jeffrey C. Cooper and John D. E. Gabrieli. 2009. Neural systems supporting the control of affective and cognitive conflicts. *Journal of Cognitive Neuroscience* 21:1842–1855.

Oldfield, Richard C. 1971. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9:97–113.

Osgood, Charles E. 1962. Studies of the generality of affective meaning systems. *American Psychologist* 17:10–28.

Pornpattananangkul, Narun, and Joan Y. Chiao. 2014. Comment: Affect control theory and cultural priming: A perspective from cultural neuroscience. *Emotion Review* 6:136–137.

Robinson, Dawn T., Lynn Smith-Lovin and Alllison K. Wisecup. 2006. Affect Control Theory. In *Handbook of the Sociology of Emotions*, eds. Jan E. Stets and Jonathan H. Turner, 179–202. Boston, MA: Springer.

Rogers, Kimberly B. 2019. Sources of Consensus and Variegation in Cultural Affective Meanings. *Social Currents* 6(3):219–238.

Rogers, Kimberly B., Tobias Schröder and Christian von Scheve. 2014. Dissecting the sociality of emotion: A multi-level approach. *Emotion Review* 6(2):124–133.

Schauenburg, Gesche, Jens Ambrasat, Tobias Schröder, Christian von Scheve and Markus Conrad. 2015. Emotional connotations of words related to authority and community. *Behavior Research Methods* 47:720–735.

Schauenburg, Gesche, Markus Conrad, Christian von Scheve, Horacio A. Barber, Jens Ambrasat, Arash Aryani and Tobias Schröder. 2019. Making sense of social interaction: Emotional coherence drives semantic integration as assessed by event-related potentials. *Neuropsychologia* 125:1–13.

von Scheve, Chistian. 2018. Affective neuroscience as sociological inquiry? In *The Palgrave Handbook of Biology and Society*, eds. Maurizio Meloni, John Cromby, Des Fitzgerald and Stephanie Lloyd, 391-415. Basingstoke: Palgrave Macmillan.

Schmidtke, David S., Tobias Schröder, Arthur M. Jacobs and Markus Conrad. 2014. ANGST: affective norms for German sentiment terms, derived from the affective norms for English words. *Behavior Research Methods* 46:1108–1118.

Schröder, Tobias. 2011. A Model of Language-Based Impression Formation and Attribution Among Germans. *Journal of Language and Social Psychology* 30:82–102.

Schröder, Tobias, and Wolfgang Scholl. 2009. Affective Dynamics of Leadership: An Experimental Test of Affect Control Theory. *Social Psychology Quarterly* 72(2):180–197.

Schutt, Russell K., Larry J. Seidman and Matcheri S. Keshavan (eds.). 2015. *Social Neuroscience: Brain, Mind, and Society.* Cambridge, MA: Harvard University Press.

Sheridan, Margaret A. 2023. Measuring the impact of structural inequality on the structure of the brain. *Proceedings of the National Academy of Sciences of the United States of America* 120(25):e2306076120.

Stephan, Klaas E., John C. Marshall, Karl J. Friston, James B. Rowe, Afra Ritzl, Karl Zilles and Gereon R. Fink. 2003. Lateralized cognitive processes and lateralized task control in the human brain. *Science* 301:384–386.

Stevens, Francis L., Robin A. Hurley and Katherine H. Taber. 2011. Anterior cingulate cortex: unique role in cognition and emotion. *The Journal of Neuropsychiatry and Clinical Neurosciences* 23:121–125.

Thagard, Paul, and Tobias Schröder. 2015. Emotions as semantic pointers: Constructive neural mechanisms. In *The psychological construction of emotion*, eds. Lisa Feldman Barrett and James A. Russell, 144–167. New York: Guilford Press.

Vaisey, Stephen. 2009. Motivation and Justification: A Dual-Process Model of Culture in Action. *American Journal of Sociology* 114(6):1675–1715

Võ, Melissa L. H., Arthur M. Jacobs and Markus Conrad. 2006. Cross-validating the Berlin Affective Word List. *Behavior Research Methods* 38(4):606-609.

Võ, Melissa L. H., Markus Conrad, Lars Kuchinke, Karolina Urton, Markus J. Hofmann and Arthur M. Jacobs. 2009. The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods* 41:534–538.

Zhang, Dandan, Weiqi He, Ting Wang, Wenbo Luo, Xiangru Zhu, Ruolei Gu, Hong Li and Yue-jia Luo. 2014. Three stages of emotional word processing: an ERP study with rapid serial visual presentation. *Social Cognitive and Affective Neuroscience* 9(12):1897–1903.

Gesche Schauenburg 1982, Dr. rer. nat., licensed psychotherapist and research psychologist at University Medical Center Hamburg-Eppendorf, Germany. Fields of research: clinical psychology, affective psychology, cognitive psychology. Publication: Using affect control theory as an approach to investigate neural correlates of affective meanings in sentence context in visual language processing, Dissertation, Freie Universität Berlin, 2022.

Arash Aryani 1976, PhD, Postdoctoral Researcher, Freie Universität Berlin. Fields of research: cognitive neuroscience. Publication: Affective arousal links sound to meaning. Psychological Science 31, 2020 (with E. S. Isbilen and M. H. Christiansen).

Chun-Ting Hsu 1981, Dr. phil., research scientist at the Guardian Robot Project of the Institute of Physical and Chemical Research (Riken). Fields of research: social neuroscience, neurolinguistics, affective neuroscience. Publications: Enhanced mirror neuron network activity and effective connectivity during live interaction among female subjects. NeuroImage 263, 2022 (with W. Sato, T. Kochiyama, R. Nakai, K. Asano, N. Abe and S. Yoshikawa).

Tobias Schröder 1979, professor and vice president at the Potsdam University of Applied Sciences. Fields of research: social psychology, computational social science, change management, sustainable cities. Publications: Disruption of social orders in societal transitions as affective control of uncertainty. American Behavioral Scientist 67, 2023 (with J. Hoey); Individual attitude change and societal dynamics. Experimenting with psychological theories. Psychological Review 28, 2021 (with J. Lorenz and M. Neumann); Modeling dynamic identities and uncertainty in social interaction: Bayesian affect control theory. American Sociological Review 81, 2016 (with J. Hoey and K. B. Rogers).

Markus Conrad 1971, Dr. phil. in Psychology, Associate Professor at the Universidad de La Laguna, San Cristóbal de La Laguna, Spain. Fields of research: reading processes, visual word recognition, bilingualism, emotion, sound-symbolism. Publications: ERPs reveal an iconic relation between sublexical phonology and affective meaning. Cognition 226, 2022 (with S. Ullrich, D. Schmidtke and S. A. Kotz).

Christian von Scheve 1973, Dr. rer. pol., Professor of Sociology at Freie Universität Berlin. Fields of research: sociology of emotion, microsociology, cultural sociology, social psychology. Publications: Affective Societies—Key Concepts. London: Routledge 2019 (ed. with J. Slaby); Angry populists or concerned citizens? How linguistic emotion ascriptions shape affective, cognitive, and behavioural responses to political outgroups. Cognition and Emotion 37, 2022 (with P. Wunderlich and C. Nguyen).

Arthur M. Jacobs 1958, Professor of Psychology at Freie Universität Berlin. Fields of research: affective neuroscience, neurocomputational poetics. Publications: Handbook of Empirical Literary Studies. Berlin 2021 (ed. with Don Kuiken); Neurocomputational Poetics: How the Brain Processes Verbal Art. London 2023.