The Inner Theater: Decoding of Nonverbal Emotional Cues Described in Literary Texts

Carolin Brück1, Christina Gößling-Arnold1, Jürgen Wertheimer1, and Dirk Wildgruber1

Abstract
Nonverbal affective signals may not only serve as a valuable guide in any day-to-day social encounter, but their role may also translate to the world of literature and storytelling. This article presents a series of three experiments aimed at investigating the processing of emotional facial expressions and nonverbal vocal signals (e.g., prosody, laughter) described in literature. Experiments addressed research questions relating to processes of imagery and simulation that may guide the decoding of described vocal and facial cues with the experimental designs relying on both behavioral and psychophysiological measures (e.g., ratings of emotional valence, auditory and visual imagery, and skin conductance response). Results obtained in these studies indicated that not only the codes and symbols used to convey emotional information in literature may resemble real-life experiences, but similarities may also emerge with respect to the perceptual mechanisms used to derive meaning from described emotional cues. Observations of a vivid, modality-specific mental imagery evoked during reading suggest that reader may, in fact, emulate perceptual experiences associated with actually seeing or hearing similar affective cues during everyday social exchange. Latter assumptions are further corroborated by the observations of emotion-modulated changes in skin conductance responses elicited during reading suggesting that mental representations formed in the process may not only include vivid visual or auditory impressions of described facial and vocal cues but may also comprise bodily representations of the emotional states associated with the depicted affective cues.

Keywords
nonverbal communication, social cues, facial expressions, vocal expressions, literature

Human communication is shaped by the use of nonverbal cues, such as facial expressions, gestures, changes in body posture, or the tone of a person’s voice while speaking. Day after day in countless situations, we are confronted with nonverbal signals and we derive from these signals information regarding a sender’s emotional state, disposition, and intention—crucial knowledge that helps to understand our fellow human beings and aids “survival” in our social environment.

However, nonverbal affective cues may not only serve as a valuable guide in any “real-world” social encounter, but their role may also translate to the world of literature and storytelling: Similar to their real-world counterparts, descriptions of nonverbal emotional signals constitute essential elements of any dialogic structure in literature regardless of whether dialog is examined or viewed with respect to the internal communication system of literature (i.e., dialog between a set of characters within the plot), the external communication system (i.e., dialog between the author and his or her reader) or, most importantly, the crossing of boundaries between the two communication systems (i.e., direct or indirect appeal to the reader by fictional characters) (Pfister, 1993). In other words, just as facial expressions or vocal cues exchanged during close-up and personal contact may lead the observer to insights into the minds of his or her partners of interaction, literary descriptions of nonverbal emotional signals may provide the reader with important clues about a character’s state of mind, his or her intentions, or dispositions. Latter insight may guide the reader to a more detailed characterization of literary figures depicted in a story or book—just as real-life expressions of emotions allow us to construct an understanding of the people expressing them. Building on the assumption that similarities between real-world nonverbal cues and literary description of these signals may extend beyond their functional role in communication, the following series of experiments sought to explore perceptual processes associated with the decoding of nonverbal communication signals.

1Eberhard Karls University, Tübingen, Germany

Corresponding Author:
Dirk Wildgruber, Department of Psychiatry and Psychotherapy, Eberhard Karls University, Calwerstraße 14, 72076 Tübingen, Germany.
Email: Dirk.Wildgruber@med.uni-tuebingen.de
described in literary texts. Research questions related to the use of imagery and simulations of real-life perceptual experiences during the decoding process, as well as stylistic devices that may facilitate these re-enactment processes.

**Experiment 1: Mental Imagery**

Immersed in a text, readers might frequently find themselves so taken in by their reading that it is as if the places and characters described on the pages in front of them suddenly spring to life. A vivid picture of what the main characters are supposed to look may emerge before the reader’s eyes, a lifelike impression of the characters’ voices may echo in their minds. It appears that in the process of reading “[…] readers construct representations of the characters, events, states, goals, and actions that are described by the story. Readers create, as it were, a microworld of what is conveyed in the story” (Zwaan, Langston, & Graesser, 1995, p. 292).

Although one might be inclined to think of such observations as phenomena restricted to a few readers endowed with a great imagination, the ability to build a mental model of what is described in a text may in fact constitute a key element in the process of comprehending written discourse (Bower & Morrow, 1990; Johnson-Laird, 1983; van Dijk & Kintsch, 1983). Some theories of discourse comprehension even argue that, “if we are unable to imagine a situation in which certain individuals have the properties or relations indicated by the text, we fail to understand the text itself” (van Dijk & Kintsch, 1983, p. 337). Literary descriptions or written language in general, consequently, may be considered a set of instructions on how to construct a mental representation of the facts and circumstances depicted by the author (Zwaan & Radvansky, 1998)—an impetus for the reader’s imagination so to say. As far as literary descriptions of nonverbal emotional signals are concerned, mental representations or rather simulations evoked by the text may, in fact, allow the reader to see with his or her mind’s eye the anger, fear, or joy a fictional character’s face is expressing, or hear with his or her mind’s ear the envy, scorn, or surprise carried by the protagonist’s voice.

In an effort to further explore processes of simulation associated with reading and decoding literary descriptions of facial and vocal affective signals, our first investigation aimed at evaluating the vividness of mental imagery evoked during reading. Particular emphasis was placed on exploring differential effects of facial and vocal affective signals on experiences of both visual imagery (VI) and auditory imagery (AI). Based on the assumption that mental imagery closely resembles perceptual processes that occur when in fact seeing or hearing similar information (Kosslyn, Ganis, & Thompson, 2001), we expected descriptions of facial expressions—that is, information closely tied to visual perception—to evoke more vivid VI, while descriptions of nonverbal vocal affective cues—in essence information associated with auditory perception—would more strongly appeal to the reader’s AI.

**Experiment 2: Effects of Direct Speech on Auditory Imagery**

Although each of the many stylistic variations created and developed by authors hold potential to facilitate mental imagery, research findings may allow singling out one potential candidate for further study. Particularly with respect to processes of AI evoked during reading, neuroimaging evidence points to a possible facilitating effect of the use of direct speech in written discourse: In line with the assumptions that AI, or rather the impressions of an inner voice while reading, should be reflected in a top-down activation of superior temporal brain areas known to be involved in voice perceptions (Belin, 2006; Belin, Fecteau, & Bedard, 2004; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000), recent findings evidencing a more pronounced recruitment of voice-sensitive temporal brain structures related to the processing of direct speech quotations (Brück, Kreiefelts, Gößling-Arnold, Wertheimer, & Wildgruber, 2014; Yao, Belin, & Scheepers, 2011) suggesting that “readers are more likely to engage in perceptual simulations (or spontaneous imagery) […] when reading direct speech as opposed to meaning-equivalent indirect speech statements” (Yao et al., 2011, p. 3146).

**Experiment 3: Psychophysiological Responses to Emotional Cues in Literature**

However, imagination or simulation process observed during reading may not only play a role in evoking vivid visual or auditory impressions of the depicted emotional signals, rather they may also relate to bodily states associated with the described emotional states. Over the years, a reasonable body of literature (reviewed, e.g., in Adolphs, 2002, 2003; Heberlein & Atkinson, 2009) has been gathered to suggest that, to understand what others might feel, we employ strategies that include simulating affective states of our fellow human beings. When aiming to decipher emotional expressions of our partners of communication, we may (at least in part) rely on representing in ourselves bodily sensations and feelings commonly associated with observed behavior (Adolphs, 2002; Heberlein & Atkinson, 2009).

However, not only the direct observation of emotional behavior may trigger such bodily changes, but thinking about different emotions alone may suffice to set of processes of simulation: Findings published on the role of simulation and embodiment of emotions during language processing, for instance, suggest an involvement of simulation in the representation of abstract emotion knowledge as well (for review, see Winkielman, Niedenthal, & Oberman, 2008). Electromyographical recordings obtained in different language processing studies, for
example, indicate that while judging whether or not a word was associated with a particular emotion, a systematic activation of facial muscles occurred that resembled facial expressions associated with the emotion the word related to (Niedenthal, Winkielman, Mondillon, & Vermeulen, 2009). Moreover, recent research findings published on the matter suggest that interfering with these bodily simulations may interfere with the judgment of emotions expressed in different words or sentences (Havas, Glencberg, & Rinck, 2007; Niedenthal et al., 2009). Latter findings once more underline the role of bodily simulation processes in the decoding of emotional information across a variety of different tasks and processing conditions.

Questions, however, remain as to whether decoding mechanisms involving such bodily representations may, in fact, also apply to the processing of emotional expressions described in literary texts. In an effort to address this issue, we conducted an experiment during which we recorded fluctuations in skin conductance responses (SCR)—a measure of physiological arousal—during both the explicit and the implicit processing of facial and vocal affective signals described in literary texts. We hypothesized that, if readers engaged in a bodily representation of described mental states, texts depicting emotional (as compared with neutral) states of the mind should result in greater arousal (and thus greater SCR) observed in readers. Latter assumption builds on theoretical considerations associating emotional states with higher levels of arousal as relative to emotionally neutral states of the mind (Russell, 1980).

**Method**

**Experiment 1**

Twenty-four volunteers were asked to participate in a short reading experiment. All volunteers were students or staff of German universities participating in an interdisciplinary seminar on emotion research. All subjects included in this study were native speakers of German or spoke German fluently.

In the course of the experiment, participants were presented with short excerpts of different works of literature and asked to read and rate each passage with respect to the vividness of the AI and VI evoked by the text. Each passage described either facial or vocal characteristics assumed to communicate different emotions without explicitly stating the respective state of mind. A total number of 212 passages were selected from 23 novels and narratives published in German with the length of each excerpt ranging from four to 74 words ($M = 16.59; SD = 11.44$). The search for suitable stimulus material was limited to novels and narratives to ensure comparability among different text samples. Texts were presented using a paper-based questionnaire that asked participants to answer three questions for each text sample: Question 1 asked participants to rate the expressed emotional state on a scale ranging from 1 (very positive) to 9 (very negative). Question 2 required participants to indicate the vividness of their acoustic imagery (i.e., the extent to which one hears with his or her mind’s ear what is described in the text) in reference to a 9-point scale ranging from 1 (no acoustic image at all) to 9 (very clear and vivid acoustic image). Similarly, Question 3 asked the readers to rate the vividness of their visual images (i.e., the extent to which one sees with his or her mind’s eye what is described in the text) on a scale ranging from 1 (no visual image) to 9 (very clear and vivid visual image). Within each questionnaire, text order was randomized, and participants were free to read and rate as many excerpts as they could finish within a time frame of 90 min. On average, readers completed about 58% of the 212 items presented to them. Each item was rated at least by 42% of the participants.

For each participant, ratings of both VI and AI were averaged across text samples describing facial expressions as well as all text samples describing vocal cues resulting in two scores for each type of nonverbal cue: AI_face, AI_voice, as well as VI_face, VI_voice. To evaluate differences in imagery evoked by facial or vocal descriptions, VI as well as AI scores were subjected to paired-samples t tests.

To assure that the emotional valence of presented text samples did not influence the observed imagery effects, results were re-evaluated in a second group of participants (20 male, 20 female, $M_{age} = 25.20$, $SD = 5.03$, all native speakers of German) using a subset of the above-mentioned text samples balanced with respect to the expressed emotional valence. Each participant was asked to read 120 text samples: 60 literary descriptions of vocal expression as well as 60 literary descriptions of facial expressions selected from the original pool of 212 text samples. The selection contained 40 text samples expressing positive emotional states (20 through descriptions of facial expressions, 20 through descriptions of vocal expressions), 41 samples expressing neutral emotional states (21 through descriptions of facial expressions, 20 through descriptions of vocal expressions), and 39 expressing negative emotional states (19 through descriptions of facial expressions, 20 through descriptions of vocal expressions). Categorizations were based on valences ratings (9-point self-assessment manikin scale ranging from 1 (very positive) to 9 (very negative)) obtained from the 24 readers that constituted the first experimental group. A text sample was considered to express a neutral emotional state if average ratings of valence fell between the scores of 4.5 and 5.5. Positive emotional meaning was ascribed to a text if ratings averaged to scores lower than 4.5. Negative meaning was suggested if the average rating scores surpassed the mark of 5.5. Participants were instructed to read each text sample and rate the vividness of visual and auditory images evoked by each text. Ratings of AI were given in reference to a 9-point scale ranging from 1 (no auditory image) to 9 (very clear and vivid auditory image). Similarly, ratings of VI were given on a 9-point scale ranging from 1 (no visual image) to 9 (very clear and vivid visual image). Texts were
presented to readers using a computerized questionnaire. Text order was randomized among participants, and readers indicated their answers by pressing the number key (1-9) corresponding to their rating on the employed vividness scales.

Analogous to the data analysis conducted with the data from the first experimental group, ratings of VI and AI were averaged among face and voice descriptions, and resulting mean values (AI_face, AI_voice, VI_face, VI_voice) were subjected to pair-wise t tests.

**Experiment 2**

To test the effect of the stylistic device direct speech on perceptual simulation processes we re-analyzed data gathered in Experiment 1 (second evaluation) with respect to differences in ratings of AI depending on whether readers were presented a text encompassing direct speech or indirect speech.

Data of the group of 40 readers (20 male, 20 female, $M_{age} = 25.20, SD = 5.03$, all native speakers of German) were reanalyzed focusing on ratings provided in response to the set of 60 voice descriptions included in the study.

Of those 60 texts, 17 included direct speech, while the remaining 43 voice descriptions did not. As described above, participants were instructed to read each text sample and rate the vividness of AI experienced during reading on a scale from 1 (no auditory image) to 9 (very clear and vivid auditory image). Texts were presented to readers using a computerized questionnaire. Text order was randomized among participants. Readers indicated their answers by pressing the number key corresponding to their ratings.

To evaluate differences in mental imagery suggested for the different reporting styles, ratings of AI were averaged among texts containing direct speech as well as texts that did not include direct speech quotations, resulting in two imagery scores (AI_direct speech, AI_no direct speech) for each participant. A paired-samples t test computed on the two different AI scores was used to evaluate differences in vividness ratings associated with the different reporting styles.

**Experiment 3**

Data were collected from 10 volunteers (five male, five female, $M_{age} = 24.70$ years, $SD = 2.11$, all native speakers of German) asked to read a set of 78 text samples. Thirty-nine of the 78 text samples described facial cues and 39 described vocal cues. Within each subset of voice or face descriptions, one third of the text samples conveyed emotionally neutral states of the mind, while the remaining two thirds communicated either a positive ($n = 13$ text samples) or a negative ($n = 13$ text samples) emotional state. Texts were selected from an original pool of 212 excerpts (see Experiment 1), and the prior categorization of emotional meaning was based on valence ratings obtained for each text sample from a group of 24 volunteers (see Experiment 1).

To evaluate the SCR associated with explicit emotion processing, participants were asked to rate each of the 78 texts with respect to valence of emotions expressed in the text by choosing one of four different response alternatives: ++ (highly positive), + (positive), − (negative), −− (highly negative). To evaluate responses associated with implicit processing of described emotional signals, participants were instructed to give their opinion on how well they thought the presented text samples were written. Each volunteer was given a choice between four possible answers: ++ (very well written), + (well written), − (poorly written), −− (very poorly written). Task order was balanced among participants: Half of the participants started with judgments of emotional valence and half with judgments of literary quality. Texts were presented using the experimental control software Presentation (Neurobehavioral Systems, Inc., Albany, CA, USA) installed on a standard personal computer equipped with a 17-inch flat screen. Texts were displayed centered in the middle of the screen (left or right margin ~20% of the screen width), in 20 pt Arial font against a light gray background. Text lengths ranged from one to five lines or 33 to 260 characters in total. Participants indicated their answers by pressing one of four buttons on a Cedrus RB-730 response pad (Cedrus Corporation, San Pedro, CA, USA). Buttons were clearly labeled with the response alternatives (i.e., ++, +, −, −−). To avoid effects attributable to the positioning of response buttons, the order of response alternatives was reversed for half of the participants (i.e., −−, −, +, ++). Readers were allowed to advance through the stimulus material at their own pace; no time limitations were imposed.

SCR data were acquired using 8-mm standard silver/silver chloride electrodes (TSD203; BIOPAC Systems, Goleta, CA, USA) filled with Gel101 electrode paste (BIOPAC Systems) and fastened to the index and middle finger (palmar surface, medial phalanx) of the non-dominant hand using Velcro® straps. Participants rested for 10 min following electrode attachment to ensure acclimation. To avoid movement-related artifacts, participants were instructed to place their non-dominant hand on a cushion next to the response pad and keep it as still as possible throughout the experiment. Signals were acquired using a BIOPAC MP-100 data acquisition system (BIOPAC Systems) equipped with an SDR 100C electrodermal activity amplifier module and an isolated digital interface (STP100C) allowing synchronizing physiological measurements with stimulus presentation. Data acquisition and recording were controlled using the software package AcqKnowledge version 3.9.1 (BIOPAC Systems). Signals were processed with an amplification of 2 $\mu$mho/V, low-pass filtered with a cutoff frequency of 1 Hz, high-pass filter with a cutoff frequency of 0.5 Hz, and sampled with a rate of 10 data points per second. Employed settings provided relative SCR recordings as compared with
absolute levels of conductivity allowing to interpret SCR changes associated with reading the texts. To facilitate an event-related analysis of gathered physiological signals, markers of stimulus onset were recorded along with the SCR data. Moreover, to allow SCR responses to return to baseline between trials, inter-stimulus intervals ranging from 5 to 6.7 s as well as six null events (i.e., 10-s periods of no stimulation) were employed.

Data analysis of SCR measurements included a peak detection protocol (i.e., MATLAB processing script, developed in-house) that allowed to determine the maximum amplitude of skin conductance changes within a predefined time window beginning 1 s following stimulus onset and ending 5 s after participants had indicated their answers and thus had finished reading. Peaks were determined for each presented text sample and subsequently averaged among valence categories (i.e., positive, neutral, negative) and cue types (i.e., facial or vocal) resulting in six single measures obtained for each participant within each task conditions (i.e., explicit emotion processing [judgments of emotional valence] or implicit emotion processing [judgments of literary quality]): $M_{\text{face}}_{\text{post}}$, $M_{\text{face}}_{\text{neg}}$, $M_{\text{face}}_{\text{neg}}$, $M_{\text{voice}}_{\text{post}}$, $M_{\text{voice}}_{\text{neu}}$, $M_{\text{voice}}_{\text{neg}}$. To evaluate differences in the magnitude of change among the different tasks, valence categories, and cue modalities, obtained measures were subjected to a repeated measures analysis of variance (ANOVA) with task, valence, and cue type defined as within-subject factors.

Results and Discussion

Experiment 1

Comparisons computed on data gathered in the first group of participants revealed a significant difference between AI ratings obtained for descriptions of facial and vocal cues, $t(23) = -6.76, p = .000$. Mean scores indicated higher AI ratings following literary descriptions of the vocal signals ($M_{\text{AI}}_{\text{voice}} = 4.65, \pm 0.30$ standard error of the mean [SEM]) as compared with descriptions of facial expressions ($M_{\text{AI}}_{\text{face}} = 1.88 \pm 0.35$ SEM). Comparisons computed on VI scores similarly indicated a significant difference between descriptions of facial and vocal cues, $t(23) = 8.19, p = .000$. Compared with the analysis of AI scores, however, the reverse pattern of results emerged: Mean VI scores revealed higher ratings for text samples presenting facial ($M_{\text{VI}}_{\text{face}} = 5.17 \pm 0.32$ SEM) as compared with vocal cues ($M_{\text{VI}}_{\text{voice}} = 2.96 \pm 0.29$ SEM). As far as data from the second experimental group are concerned, comparisons computed on AI scores once again revealed a significant difference between ratings obtained for descriptions of facial and vocal cues, $t(39) = -12.45, p = .000$, with higher mean AI ratings obtained for vocal ($M_{\text{AI}}_{\text{voice}} = 5.54, \pm 0.25$ SEM) as compared to facial cues ($M_{\text{AI}}_{\text{face}} = 2.18, \pm 0.19$ SEM; Figure 1a). Comparisons computed on VI scores again indicated a significant difference between descriptions of facial and vocal expressions, $t(39) = 11.34, p = .000$, with
higher mean VI ratings obtained for facial \( (M_{\text{VI, face}} = 6.43, \pm 0.21 \text{ SEM}) \) as compared to vocal cues \( (M_{\text{VI, voice}} = 4.19, \pm 0.21 \text{ SEM}; \text{Figure 1a}) \).

In sum, while descriptions of visual affective signals such as facial expression elicited more vivid visual images, literary depictions of auditory signals were accompanied by more vivid experiences of AI. Findings of both vivid VI and AI in the process of reading descriptions of facial and vocal emotional signals appear to be in line with a variety of research reports (e.g., Alexander & Nygaard, 2008; Kurby, Magliano, & Rapp, 2009; Long, Winograd, & Bridge, 1989; Sadoski, 1985; Sadoski, Goetz, Olivarez, Lee, & Roberts, 1990) as well as anecdotal evidence suggesting experiences of mental imagery to be a common phenomenon associated with reading. Similarities in perceptual details or cerebral mechanisms engaged in the processes (Kosslyn et al., 2001), underline the notion that mental images created by the reader emulate direct perception in many aspects (Brück et al., 2014; Yao et al., 2011), and visual and acoustic images evoked during reading may, in fact, rely on recalling or re-combing stored knowledge gathered from previously perceived events that fit the descriptions provided by the text (Kosslyn et al., 2001; van Dijk & Kintsch, 1983). Empirical evidence suggests that in the course of reading, readers might not only access knowledge structures pertaining to temporal or spatial information encoded in the text (Bower & Morrow, 1990; Zwaan & Radinsky, 1998), but they may also activate knowledge about human emotions allowing them to infer emotional states of a fictional character (Gernsbacher, Goldsmith, & Robertson, 1992; Gernsbacher, Hallada, & Robertson, 1998) even if emotions are not explicitly stated in the text. Using this knowledge to understand emotional information represented by text may to a certain degree resemble recapturing past sensory-motor states activated while directly perceiving or encountering similar situations or events (see, for instance, Niedenthal, 2007), including perhaps visual and auditory impressions formed during these experiences that may reflect themselves in mental images evoked during reading.

**Experiment 2**

As far as the effect of text style on imagery is concerned, results revealed a significant difference between mean AI scores obtained for the reporting styles of direct and indirect speech, \( t(39) = -5.37, \ p = .000 \). On average, texts encompassing portions of direct speech received higher ratings of AI \( (M_{\text{AI, direct speech}} = 6.06 \pm 0.28 \text{ SEM}) \) as compared to texts that did not include direct speech \( (M_{\text{AI, non-direct speech}} = 5.34 \pm 0.24 \text{ SEM}; \text{Figure 1b}) \).

The respective findings, in general, support ideas derived from neuroimaging evidence suggesting direct speech statements to elicit a more vivid representation or perceptual simulation of a reported voice (Brück et al., 2014; Yao et al., 2011), resulting in more intense experiences of AI while reading. From a linguistic point of view, direct speech has been considered a means of communication through which the recipient is provided with a demonstration rather than a description of a speech act (Clark & Gerrig, 1990). Such propositions may be corroborated by our personal experiences with direct speech in day-to-day communication: Considering spoken discourse, for instance, direct speech in the form of quotations may not only provide a verbatim account of what was said, but may perhaps also entail a simulation of the vocal behavior including idiosyncrasies in speech rate or voice pitch. A female speaker may, for instance, lower her voice while quoting her husband, or adopt an accent that fits his.

Empirical data provided on the silent reading of direct speech statements suggest similar simulation processes associated with the processing of direct speech in written discourse: Reading rates, for instance, have been shown to differ depending on whether the fictional speaker of direct speech statement had been introduced as a slow or fast talker (Yao & Scheepers, 2011). Context information, however, did not affect the reading of similar text samples employing other reporting styles, such as indirect speech statements and rendering observed effects specific to direct speech. Latter findings, in turn, may be interpreted to reflect the unique potential of direct speech quotations to evoke vivid simulation or imagination processes in readers (Yao & Scheepers, 2011) that may aid the decoding of information conveyed in the text.

**Experiment 3**

As targeted by Experiment 3, imagination or simulation process observed during reading may not only play a role in evoking vivid visual or auditory impressions, but they may also relate to bodily states associated with the described emotional states. Results of Experiment 3 indicated that while no significant effects were found with respect to the influence of task, \( F(1, 9) = 0.46, \ p = .514, \text{ cue modality}, F(1, 9) = 0.03, \ p = .868, \text{ or the interactions of the respective factors (all interactions} \ p > .416), \text{ data analysis revealed a significant main effect of valence, } F(1.61, 14.52) = 4.13, \ p = .045, \text{ (Greenhouse-Geisser corrected)} \text{ on SCR changes observed during reading. Post hoc paired-samples } t \text{ tests computed among the three different valence categories allowed to further delineate the finding: Relative to responses obtained for descriptions of neutral facial and vocal signals, reading descriptions of positive or negative emotional states was associated with significantly higher SCR changes (positive vs. neutral: } t(9) = 2.26, \ p = .050, \text{ negative vs. neutral: } t(9) = 3.46, \ p = .007; \text{ positive vs. negative: } t(9) = -0.43, \ p = .677). \text{ In other words, irrespective of whether task instructions required the explicit or implicit decoding of emotional signals, text samples conveying positive as well as negative states of the mind induced in the reader higher changes of physiological activity (mean peak}
amplitudes of change: $M_{\text{pos}} = 0.035$ µmho ± 0.005 SEM, $M_{\text{neu}} = 0.032$ µmho ± 0.004 SEM, $M_{\text{neg}} = 0.036$ µmho ± 0.004 SEM; Figure 2a). To evaluate whether observed emotional-related increases in SCR were further modulated by the perceived emotional intensity, a regression analysis was conducted with mean SCR changes obtained for each text sample defined as dependent variable and mean emotional intensity of each text modeled as regressor. Emotional intensity scores were derived by averaging individual valence ratings (++ = 1, + = 2, − = 3, −− = 4) for each stimulus among participants, then subtracting each mean from 2.5 (i.e., the mean value to be expected if participants regarded any given text sample as neutral), and finally converting calculated differences to absolute values. Regression analysis revealed a significant association between the emotional intensity of text samples and SCR changes obtained.
during reading \((R = .265, R^2 = .07, p = .001)\) indicating that the more emotionally intense a given text sample was perceived, the higher the SCR changes measured while reading the respective text (Figure 2b).

In sum, observed emotion-modulated increases in bodily responses may be explained by applying a simulation model of emotion recognition (Adolphs, 2002; Heberlein & Atkinson, 2009) not only to direct perception of emotional cues but also to the decoding of written descriptions of affective signals: Assuming that emotion recognition may at least in part rely on simulating bodily sensations associated with the perceived (or in this case described) mental states, one would expect the decoding of emotional signals to be accompanied by increases in physiological activity that resemble changes in arousal induced by emotional experiences. As both positive and negative emotional states have been assumed to induce higher levels of bodily activity compared with neutral states of the mind (Russell, 1980), one would expect simulation of the respective emotional states to result in more pronounced increases of physiological activity. The pattern of findings obtained in this experiment proves to be in accordance with the outlined predictions and thus might be considered evidence of simulation. Mental representation formed by the reader might consequently involve not only vivid pictures or auditory images but also bodily sensations associated with described mental states. As observed emotion-modulated SCR responses extended to both the processing of described facial and vocal cues, and emerged regardless of whether or not attention was focused directly on emotional information, suggested simulation processes might be considered to reflect a more general decoding mechanism associated with perceiving descriptions of emotions across a variety of different processing conditions and cues.

**Concluding Remarks**

With only a facial expression or a slight change in the tone of their voices, human beings communicate a wealth of information, that allows access into the sender’s current state of mind. In the field of literature and storytelling, readers might frequently encounter descriptions of facial or vocal cues, and—as with their real-world counterparts—readers might utilize these signals to derive valuable information regarding a character’s attitudes, intentions, or feelings. Concluding from the research finding reported above, our ability to decipher verbal descriptions of nonverbal signals might be at least in part driven by similar decoding mechanisms used to understand our fellow human beings during everyday social interaction. Reports of a vivid, modality-specific mental imagery, moreover, suggest that, in the process of reading, readers might be able to emulate the direct perception of nonverbal affective signals allowing them to “see” described visual cues before their mind’s eye or “hear” described vocal cues with their mind’s ear.

A novel, or perhaps even literature in general, “is in its broadest definition a personal, a direct impression of life” (James, 1888, p. 384) and in a sense it is our life experience that allows us to comprehend or appreciate literature—and vice versa. Not only the symbols and codes used to convey emotional information in literary texts might mimic real-life experiences, but also the skills and mechanisms used to decipher meaning from presented cues might closely resemble the “tactics” employed to navigate the social terrain on a day-to-day basis.

**Authors’ Note**

Carolin Brück and Christina Gößling-Arnold contributed equally to the work and share first authorship.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The author(s) received financial support by the German Research Foundation (Deutsche Forschungsgemeinschaft) and Open Access Publishing Fund of the University of Tübingen to help publish this research.

**References**

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews, 1*, 21-62.

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Review Neuroscience, 4*, 165-178.

Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: Talker-specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 446-459.

Belin, P. (2006). Voice processing in human and non-human primates. *Philosophical Transactions of the Royal Society of London B: Biological Science, 361*, 2091-2107.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Science, 8*, 129-135.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*, 309-312.

Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science, 247*, 44-48.

Brück, C., Kreifelts, B., Gößling-Arnold, C., Wertheimer, J., & Wildgruber, D. (2014). “Inner voices”: The cerebral representation of emotional voice cues described in literary texts. *Social Cognitive and Affective Neuroscience, 9*, 1819-1827.

Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language, 66*, 764-805.

Gernsbacher, M. A., Goldsmith, H. H., & Robertson, R. R. W. (1992). Do readers mentally represent characters’ emotional states? *Cognition and Emotion, 6*, 89-111.
Gernsbacher, M. A., Hallada, B. M., & Robertson, R. R. W. (1998). How automatically do readers infer fictional characters’ emotional states? *Scientific Studies of Reading, 2*, 271-300.

Havas, D. A., Glenberg, A. M., & Rinck, M. (2007). Emotion simulation during language comprehension. *Psychonomic Bulletin & Review, 14*, 436-441.

Heberlein, A. S., & Atkinson, A. P. (2009). Neuroscientific evidence for simulation and shared substrates in emotion recognition: Beyond faces. *Emotion Review, 1*, 162-177.

James, H. (1888). *Partial portraits*. London, England: Macmillan.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience, 2*, 635-642.

Kurby, C. A., Magliano, J. P., & Rapp, D. N. (2009). Those voices in your head: Activation of auditory images during reading. *Cognition, 112*, 457-461.

Long, S. A., Winograd, P. N., & Bridge, C. A. (1989). The effects of reader and text characteristics on imagery reported during and after reading. *Reading Research Quarterly, 24*, 353-372.

Niedenthal, P. M. (2007). Embodiment of emotion. *Science, 316*, 1002-1005.

Niedenthal, P. M., Winkielman, P., Mondillon, L., & Vermeulen, N. (2009). Embodiment of emotion concepts. *Journal of Personality and Social Psychology, 96*, 1120-1136.

Pfister, M. (1993). *The theory and analysis of drama* (Reprint ed.). Cambridge, UK: The Press Syndicate of the University of Cambridge.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161-1178.

Sadoski, M. (1985). The natural use of imagery in story comprehension and recall—Replication and extension. *Reading Research Quarterly, 20*, 658-667.

Sadoski, M., Goetz, E. T., Olivarez, A., Lee, S., & Roberts, N. M. (1990). Imagination in story reading—The role of imagery, verbal recall, story analysis, and processing levels. *Journal of Reading Behavior, 22*, 55-70.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.

Winkielman, P., Niedenthal, P. M., & Oberman, L. (2008). The embodied emotional mind. In G. R. Semin & E. R. Smith (Eds.), *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches* (pp. 263-288). New York, NY: Cambridge University Press.

Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience, 23*, 3146-3152.

Yao, B., & Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition, 121*, 447-453.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*, 292-297.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162-185.

**Author Biographies**

**Carolin Brück** is a researcher associated with the affective neuroscience research group in the Department of Psychiatry and Psychotherapy at the University of Tübingen (lead by Dirk Wildgruber).

**Christina Gößling-Arnold** is a researcher working in the field of international literature at the Department of Modern German Literature at the University of Tübingen (lead by Jürgen Wertheimer).

**Jürgen Wertheimer** is a professor of Comparative Literature and International Literature at the University of Tübingen. One of his fields of research contains the representation of affective signals in literary texts including pity, horror, confidence and laughter.

**Dirk Wildgruber** is an associate professor in the Department of Psychiatry and Psychotherapy at the University of Tübingen. His area of research is the neurobiology of social interaction and emotion processing in healthy subjects and patients with psychiatric disorders.