Title

Emotional Words — The Relationship of Self- and Other-Annotation of Affect in Written Text

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Braun, Nadine
Goudbeek, Martijn
Krahmer, Emiel

Publication Date

2020

Peer reviewed

Emotional Words – The Relationship of Self- and Other-Annotation of Affect in Written Text

Nadine Braun (N.Braun@uvt.nl)
Martijn Goudbeek (M.B.Goudbeek@uvt.nl)
Emiel Krahmer (E.J.Krahmer@uvt.nl)
All authors: Tilburg University, Department of Communication and Cognition, 5000 LE Tilburg

Abstract

For human and automatic text annotation of emotions, it is assumed that affect can be traced in language on (combinations of) individual words, text fragments, or other linguistic patterns, which can be identified and labelled correctly. For example, many sentiment analysis systems consider isolated words affectively meaningful units, whose proportions in a given text reveal its overall affective meaning. However, whether these words and their combinations as identified either by humans or algorithms also match the actual feelings of the authors remains unclear. Potential discrepancies between affect expression and perception in text have received surprisingly little scholarly attention, although a number of studies has already identified disparities between self- and other-annotation in affect detection for speech and audio-visual data. Therefore, we ask whether a similar difference shows in annotations of emotions in text.

Keywords: emotion expression; emotion perception; text annotation; language production; appraisals

Introduction

The correct understanding of language depends not only on an addressee's knowledge of the respective language itself but also on the correct decoding of a message with respect to both its semantic and its pragmatic meaning. For example, the connotation of a word such as "shoot" depends not just on the words succeeding it (e.g., "a ball", "a photo", or "a person"; see also "semantic prosody", Louw, 1993) but also on the pragmatic context. While scoring a goal is generally a positive thing in soccer, it is admittedly less so if the author is a supporter of the opposing team, in which case it would be the description of a rather negative event surrounded by disappointment or even anger. However, to correctly decode such affective meanings, additional background information about the author and their motivation are necessary. In spoken language, this information can be communicated indirectly or directly in the form of facial expressions or the prosody of an utterance. In written text, this can also entail additional text or one's own background knowledge about the respective subject. Whereas it is widely known that intended meanings are crucial to grasp phenomena like sarcasm, their importance for the expression and perception of affect in language has been less studied. Nevertheless, it seems fair to assume that author context and motivation are just as important when it comes to decoding emotions and that disregarding them might lead to discrepancies between affect expression and perception. This might be especially problematic for research fields dedicated to the detection of patterns that reveal affect and opinions, such as sentiment analysis.

Although sentiment analysis has become a fast-growing and popular research field over the past decades, these potential differential effects of context have received little scholarly attention. For both human and automatic approaches to annotation, the underlying idea is that affect can be traced in language (see, e.g., Bestgen, 1994; Hunston & Thompson, 2000) – on individual words, text fragments, or other linguistic patterns – and that these patterns, produced consciously or subconsciously by authors, can be identified and labelled correctly. Many sentiment analysis systems take a bag-of-words approach, which considers isolated words as affectively meaningful units, whose proportions in a given text reveal its overall affective meaning. However, whether these words and their combinations identified either by humans or algorithms also match the actual feelings of the authors is difficult to assess. Depending on its extent, a potential gap between affect expression and perception could cause problems for the field since sentiment analysis is not only used for commercial purposes, such as the identification of trends and opinions about products or people (Bae & Lee, 2012), but also for clinical applications, such as depression detection (Losada & Gamallo, 2018; Nguyen, Phung, Dao, Venkatesh, & Berk, 2014), the development of better suicide prevention strategies (Christensen, Batterham, & O'Dea, 2014), research into the perceived quality of healthcare in patient online communication (Denecke & Deng, 2015), or the automatic detection and, hence, better prevention of (cyber)bullying (Chatzakou et al., 2017). For these sensitive purposes, discrepancies could lead to incorrectly labeled data and inaccuracies in analyses even though accuracy and precision of analyses are of particular importance for clinical applications that aim at interventions.

Although obtaining affective "ground truth" is a challenging endeavor, attempts have been made by requesting authors and speakers to label their own data. In doing so, a number of studies has already identified disparities between self-reported and other-observed emotions. For speech and audio-visual data in the context of video games, Truong, Van Leeuwen, Neerincx, and de Jong (2009) showed that players, whose voices and faces were recorded while they played a computer game, categorized their own facial expressions into

different emotions and assigned different intensities to themselves than other coders who annotated the same videos on the same emotions. Barr and Kleck (1995) demonstrated a similar effect in two studies: participants evaluated the intensity of amusement expressed in their own faces, which were recorded while they watched video segments. Independent judges subsequently rated the intensity of the participants' faces in the same recordings, but generally judged the facial expressions as less intense than participants themselves. Other studies report similar findings (see, e.g., Afzal & Robinson, 2009). However, these studies only examined audio-visual data, so the question arises whether these effects can also be traced in written language.

Intuitively, the existence of differences between affect expression and perception in text makes sense, considering that people generally tend to infuse judgements about the opinions of others with their own knowledge and opinions. Research on "the curse of knowledge" , i.e. the inability to ignore one's preexisting knowledge, illustrates the difficulty of making objective judgements about others' mental states, both on a cognitive and an emotional level (see, e.g., Birch & Bloom, 2007; Damen, van der Wijst, van Amelsvoort, & Krahmer, 2018; Keysar, 1994). On a related note, Van Boven and Loewenstein (2005) discuss egocentricity in knowledge in the form of "empathy gaps" in the prediction of emotions in others: similar to biases about other's knowledge, people tend to judge others' affective states based on their own experiences and emotions. Moreover, not only do people tend to misjudge others' affective and mental states based on their own knowledge but they also misconstrue the intensity of their own emotional expression. This phenomenon is known as the "illusion of transparency" (see, e.g., Barr & Kleck, 1995; Gilovich, Savitsky, & Medvec, 1998).

While these effects are mostly investigated within the field of psychology, cognitive and affective expression-perception gaps have also been studied with regard to language, such as computer-mediated communication (CMC). For instance, Kruger, Epley, Parker, and Ng (2005) discuss people's overconfidence when communicating through emails: in five studies, they illustrate an expression-perception gap, among others, for sarcasm, sadness, or anger, and argue that this gap is a consequence of people's egocentricity, triggered by the focus on one's own intentions and the inability to adjust to the (emotional) perspective of another person. In similar experiments, Riordan and Trichtinger (2017) demonstrate the discrepancy between the perceived emotional intensity of emails between authors and readers. These studies suggest that, especially in email communication, more subtle information such as affective meanings can easily get lost, a finding that nicely illustrates potential gaps between own- and other-perceptions of texts.

Attempts have been made to assess and resolve, or at least, minimize these gaps in analyses. To address discrepancies, some studies for which large online text corpora were collected already include hashtags and emojis added by the authors of texts. In contrast to email communication, these texts are often not directed at specific addressee but are intended for an unspecified number of readers. The added tags are considered a form of emotional self-report or self-annotation and are used by authors to voluntarily indicate information about affective states (e.g., Liew, Turtle, & Liddy, 2016; Park, Xu, & Fung, 2018) or other meta-information about the text, such as intended sarcasm (see, e.g., González-Ibánez, Muresan, & Wacholder, 2011; Khodak, Saunshi, & Vodrahalli, 2017; Mihalcea & Liu, 2006). While for short texts, these tags that usually refer to the whole post and its context provide valuable information about the author and their intentions, the expression of affect can fluctuate within longer texts consisting of several sentences or even paragraphs. If no explicit cues like hashtags and emojis are provided – a practice that is mainly common on social media but less so in other registers – an author's true emotions are all but accessible. In particular, the intended affective meaning can diverge drastically from a reader's understanding of a text. Yet, it remains unclear how substantially author affect differs from the readers' perception. Therefore, in the current study, we attempt to investigate this affect expression-perception gap in longer written text further by examining the relationship between affect expression of the authors of affective texts and the perception of their readers (RQ).

The Current Study

Based on the research question and existing literature containing evidence for an affective expression-perception gap, such as studies on audio-visual data (e.g., Truong, Van Leeuwen, & De Jong, 2012), the "curse of knowledge" (e.g., Keysar, 1994), the "illusion of transparency" (e.g., Gilovich et al., 1998), and discrepancies in email communication in CMC, we intend to measure the discrepancy between author and reader perception of affective texts using a two-part study (writing and annotation). Since many sentiment analysis approaches assume a bag-of-words approach to affective texts by using individual words and their proportions in a text as markers of overall text affect, we will investigate text affect in a similar way, using annotations of individual, emotionally meaningful words and compare their proportions to the overall valence of the texts. Further, we are not only interested in the valence (positivity, negativity) of words and texts, which most studies focus on, but also in the discrete emotions conveyed by individual words (see, e.g., Ekman, 1992 or, more recently, Cordaro et al., 2018). Although, for example, anger and sadness are both generally categorized as negative emotions, they tend to arise in different contexts (Smith & Ellsworth, 1985) and they usually serve different purposes (e.g., Tiedens, 2001): anger tends to convey dominance and competence, while sadness often triggers compassion in addressees. However, since the wording of a writing task could prime authors' language if emotion terms were used, we will use an appraisal-based framework for the instructions. Appraisals are dimensions related to a situation or trigger that can be used to explain and describe emotional experiences, such as arousal, pleasantness, or certainty, the combination of which, in turn, is believed to prompt specific emotions (Ellsworth & Scherer, 2003; Smith & Ellsworth, 1985). In our

case, we expect the combination of different appraisals to cue authors to write about different emotional experiences. Therefore, we hypothesize that texts produced according to different combinations of appraisals will differ in valence (H1a) and emotion categories annotated by authors (H1b). Further, we assume that self-annotations will contain more emotion tags than other-annotations (H2) and that self-annotators will rate their texts to be more strongly valenced overall compared to other-annotators (H3). Based on the difference in audio-visual data, we predict that inter-annotator agreement will be higher between other-annotators than agreement between self- and other-annotators (H4).

Method

Text Production In order not to prime authors with specific emotion terms, appraisals were used to formulate writing instructions. Based on the appraisal theory of emotion (Smith & Ellsworth, 1985), two appraisals, which are likely to elicit a range of different emotions when combined, were chosen: pleasantness, which is related to valence, and control, which is related to the attribution of responsibility for an event or situation. Combined, four combinations and, hence, conditions emerge: 1) high pleasantness, low control (HPLC), involves emotions such as amusement, happiness, awe, relief, surprise; 2) high pleasantness, high control (HPHC), e.g., happiness, relief, coyness, contentment, desire, interest, triumph; 3) low pleasantness, low control (LPLC), e.g., sadness, fear, pain, confusion, boredom), and 4) low pleasantness, high control (LPHC), e.g., embarrassment, anger, contempt, disgust, shame.

Annotation For the annotation task, participants were asked to identify individual words with emotional meaning in the texts and assign emotion categories to these words by creating "tags" (max. one per word). The categories to be annotated were inspired by Cordaro et al. (2018), who identified 22 emotions, categorized according to valence (i.e. positive or negative) and arousal: amusement, food desire, sexual desire, interest, surprise, triumph, contentment, coyness, relief, anger, confusion, embarrassment, fear, pain, bored, contempt, disgust, sadness, shame, sympathy. The terms were translated to Dutch by three native Dutch speakers, who, after intensive discussion, agreed on the final translations. The annotation of emotions in the texts was done on the word level using the freely available offline annotation tool MMAX2 (Müller & Strube, 2006). Emotion categories were provided as 22 buttons that participants could assign by highlighting a word and then clicking on a button.

Procedure

Part 1: Text production and self-annotation Upon arrival, participants were informed about the study and gave consent, after which they were asked to start a Qualtrics survey. For the text collection part of the study, participants completed a writing task. They entered their age and gender in the beginning and were subsequently presented with four different writing instructions based on the conditions, whose order was automatically randomized in the survey**, e.g.:** "Think of a situation in which you felt like you had no control over what was happening but which was ultimately very pleasant for you. Take your time and try to remember details about the experience. After you have visualized the situation in your mind, describe it below. You can continue with the next part of the survey after five minutes. If you are not done by then, don't worry and take your time to finish your **st**ory." (translated from Dutch). Upon finishing each text, participants were asked to rate the valence of the described experience on a 9-point Likert scale. After participants were finished with the writing task, the survey concluded with a neutral nature video without voiceover or subtitles.

The writing task was followed by a 5-minute break, in which participants watched a nature video with a calming melody but without spoken or written descriptions. During this break, the experimenter prepared the annotation task by loading the produced texts into the offline annotation tool on a second computer. When the video finished, the participant informed the experimenter, who then explained the annotation tool and task to them. A brief instruction manual was kept on the table and the experimenter stayed in the room in case problems arose. Participants proceeded to annotate all four texts in random order and again notified the experimenter after they were finished. Finally, they received the debriefing form, were thanked for their participation, and dismissed from the session.

Part 2: Other-annotation Again, participants were first informed about the study and gave consent, after which the experimenter explained the annotation tool and task. Participants received a sheet of paper on which they rated each annotated text on valence on a 9-point Likert scale and on which they noted their age and gender. Similar to the first part of the study, participants could also consult an instruction manual or ask the experimenter questions about the annotation tool. After they completed the ten annotations, they informed the experimenter, who handed them the debriefing form. Finally, they were thanked for their participation and dismissed from the session.

Participants

In Part 1 of the study, 30 participants (25 female, M_{Age} = 22.03, SD = 1.59) completed the writing task and self-annotation. For Part 2, another 60 participants completed the annotation-only task (47 female, M_{Age} = 21.85, SD = 2.15). All participants in both parts of the study were Bachelor and Premaster students of Tilburg University (Department of Communication and Cognition) and participated for course credit.

Analysis plan

To investigate the proposed hypotheses, different types of analyses were used. For H1, H2, H3, and part of H4, linear mixed effects models were run in R using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) to account for repeated measures due to multiple texts and annotations by the

Table 1: Comparison of valence ratings by conditions (HP/LC, HP/HC, LP/LC, LP/HC) as rated by authors (H1).

	Mean	SD	Condition	*b*	99% CI	SE
HP/LC	7.90	1.29				
			HP/HC	0.37	-0.40, 1.17	0.28
			LP/LC	-6.17	**-6.94, -5.33**	0.32
			LP/HC	-5.17	**-5.94, -4.44**	0.30
HP/HC	8.26	1.20				
			LP/LC	-6.53	**-7.31, -5.82**	0.29
			LP/HC	-5.53	**-6.33, -4.83**	0.29
LP/LC	1.73	0.94				
			LP/HC	1.00	**0.19, 1.83**	0.31
LP/HC	2.73	1.33				

Note: Significant comparisons in bold.

same authors across conditions. For H1, the dependent variables were overall text valence as indicated on a 9-point scale and the number of positive and negative emotion tags per text, and the independent variable is Appraisal Combination (HP/LC, HP/HC, LP/LC, LP/HC). Authors and individual texts will be added as random factors. For H2 and H3, again, the dependent variables were the overall number of emotion tags and valence, while the independent variable is Annotation Perspective (self, other). The confidence intervals were set to 99% to correct for multiple comparisons. For H4, we calculated the inter-annotator agreement between other-annotators, and other-annotators and authors using Krippendorff's alpha with the "irr" package (and the implemented kripp.alpha function) in R. For the comparison, we considered the distribution of Kalphas and the changes for the agreement of other-annotators and other-annotators/authors across all texts. Since we did not have an indication of data spread for Kalphas, we compared the Kalphas of the two groups (other-annotators with/without the authors) in a linear mixed model with Annotator Group as the predictor and Texts as random factors.

Table 2: Comparison of valence ratings (H3) by conditions (HP/LC, HP/HC, LP/LC, LP/HC) as rated by authors and other-annotators (intercept).

	Mean	SD	*b*	99% CI	SE
HP/LC	7.05	1.13	0.85	**0.26, 1.39**	0.22
HP/HC	7.77	1.06	0.50	-0.20, 1.18	0.27
LP/LC	2.43	1.53	-0.69	**-1.41, -0.03**	0.27
LP/HC	3.18	1.40	-0.44	-1.13, 0.26	0.27

Note: Significant comparisons in bold.

Results

On average, the texts consist of 149.42 words (*SD* = 54.25). Valence assigned by authors to their own texts differs significantly between almost all combinations of the four conditions (H1a), with the exception of the comparison of HP/LC (*M* = 7.90, *SD* = 1.29) to HP/HC (*M* = 8.26, *SD* = 1.20; see Table 1). The types of emotions annotated vary by condition (H1b; see Figure 1): e.g., contentment is most prominent in HP/HC, while surprise is mostly annotated in HP/LC. In LP conditions, sadness and pain are indicative of LC, while shame and embarrassment occur mostly in HC situations. However, the difference in number of annotated positive and negative tags is only significant between HP (positive: *M* = 5.07, *SD* = 2.43; negative: *M* = 1.6, *SD* = 2.03) and LP (positive: *M* = 1.67, *SD* = 1.60, *b* = -3.40, SE = 0.36, BC 95% CI [-4.08, -2.65]; negative: *M* = 5.1, *SD* = 3.14, *b* = 3.50, SE = 0.42, BC 95% CI [2.65, 4.33]), not between LC and HC conditions. In addition, there is no significant difference between the number of tags used by authors and other-annotators in either of the conditions (H2). A comparison between own- and other-ratings of valence (H3) shows that authors rated LC conditions to be significantly more positive (HP) and more negative (LP) than other-annotators (Tab. 2; Fig. 2), while there is no significant difference between HC conditions. While inter-annotator agreement for annotated emotions (H4) is low overall, agreement between other-annotators only (*M* = 0.31, *SD* = 0.09) is higher than between other-annotators and the author (*M* = 0.30, *SD* = 0.08, *b* = -0.009, SE = 0.004, BC 95% CI [-0.02, -0.002]).

Discussion

In a two-part study, we showed that authors and other readers differ in their understanding of affective texts. For authors, four conditions based on two appraisal categories, high/low pleasantness and control (HP/LC, HP/HC, LP/LC, LP/ HC), differed significantly in valence ratings except for the two HP conditions, while the annotated emotion tags mainly reflected the difference between high and low pleasantness. In con-

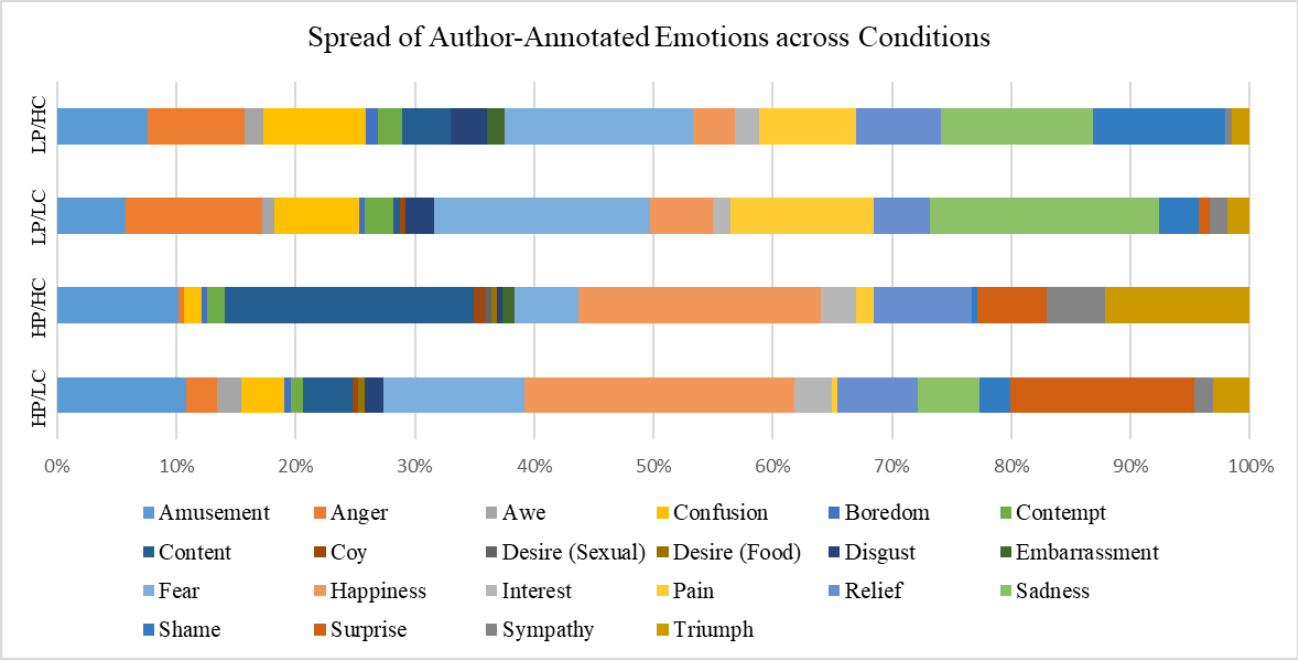

Figure 1: Percentages of all emotion tags per condition as annotated by authors (HP/LC, *N* = 194; HP/HC, *N* =206; LP/LC, *N* = 209; LP/HC, *N* = 197).

trast, other-annotators considered LC conditions to be respectively more negative (HP) and more positive (LP) than authors themselves. Additionally, agreement between other-annotators for annotated words and assigned emotion categories was higher than between other-annotators and authors. Although inter-annotator agreement appears to be low and the observed difference between own- and other-annotations seems small, it is important to consider the low chance level in the current study. Since 22 emotion categories were used, each text consisted of about 150 words, and 6 people annotated each text chances of participants "accidentally" annotating the same words the same way in the texts were negligible, making such a seemingly small difference meaningful in the context of the study and topic. This finding is in line with earlier research on audio-visual perception differences between the self and other observers by Truong et al. (2012) and with well-known phenomena observed in psychology, such as the "illusion of transparency" and the "curse of knowledge".

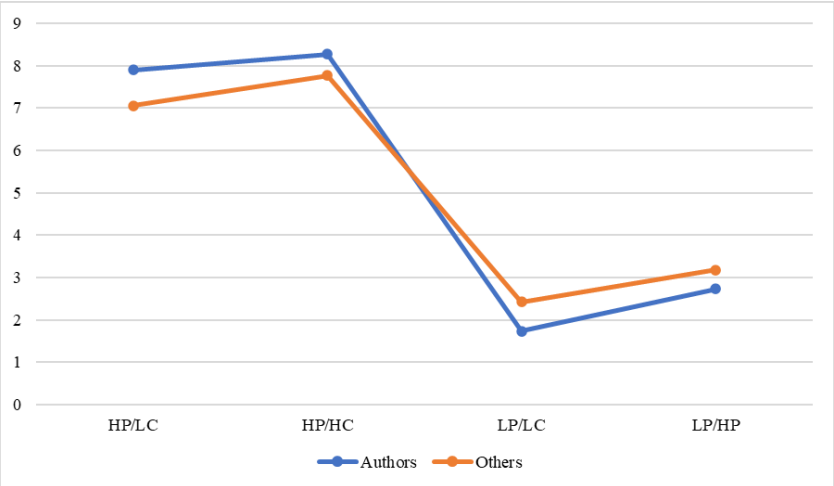

Figure 2: Valence ratings by conditions (HP/LC, HP/HC, LP/LC, LP/HC) as rated by authors and other-annotators.

While the results suggest a difference between emotion expression and perception, some characteristics of the data and the experimental setup that could have potentially influenced the results should be considered. Firstly, gender distribution across the two parts of the experiments was not ideal. More women than men participated in both the production and annotation study, which is likely due to a higher proportion of women in our student population and, hence, the participant pool used for the study. In previous studies, gender has been shown to have an effect on emotion recognition abilities and emotional intelligence, with women being supposedly more susceptible to others' emotional states (Hall, Carter, & Horgan, 2000; Hoffmann, Kessler, Eppel, Rukavina, &Traue, 2010), although recent findings suggest more subtle differences (Fischer, Kret, & Broekens, 2018). While these differences might also exist in our dataset, we assume that the higher proportion of female participants would, if anything, have reduced the gap between emotion expression and perception

and that, in a more balanced setup, the gap might be even greater; based on the research on gender differences, this difference should be especially pronounced for male annotators. Aside from gender differences, another point of concern about the current study might be generally different abilities in terms of emotion expression and recognition, in particular alexithymia. Alexithymia refers to a subclinical condition which describes impaired introspective processes and a lack of mental representations of emotions, which can lead to problems with recognizing and verbalizing one's own emotions and which can also affect recognition of others' emotional states (see, e.g., Lane et al., 1996). Consequently, alexithymic participants would have experienced more difficulties performing the tasks required for our study than non-alexithymic participants. Parallel to gender differences in emotion recognition abilities, the condition is also considered more prevalent in men than in women (Levant, Hall, Williams, & Hasan, 2009). Unfortunately, as we did not control for alexithymia in the current study, we cannot make claims about its influence on the current results. However, a recent study suggests that the condition might be moderated by verbal abilities, meaning that highly alexithymic individuals with a high verbal IQ might be able to recognize emotions similarly well as low-alexithymic individuals (Montebarroci, Surcinelli, Rossi, & Baldaro, 2011). Since a high verbal IQ might be expected for a communication student population and since the majority of our participants were females, we assume that the condition might have been less problematic in this study. Nevertheless, to ensure that the results were not affected by alexithymia, this could be controlled for in a follow-up study. Further, the differences between authors and annotators might also reduce over time, in that an extended period of time between recall and annotation of emotional events might increase the author's emotional distance to the words used to tell the story. Again, this might open up interesting possibilities for future research.

In future work, we aim to do more extensive analyses. For example, the overlap between authors and other-annotators will be analyzed further; in particular, whether the valence of the annotated emotion categories coincides and to which degree the annotated words match for both groups. Additionally, we will compare the emotion words annotated by our participants with existing affective word lists, such as the Dutch Linguistic Inquiry and Word Count (LIWC; Zijlstra, Van Meerveld, Van Middendorp, Pennebaker, & Geenen, 2004). In theory, the words identified in the current study and should match entries in existing affective lexicons and these lexicons should classify the valence of texts produced for the current study in a similar way as our human annotators did.

The results of the current study hint at a misalignment of author and reader perception, which might be caused either by misunderstanding or miscommunication of the emotions conveyed through the texts. If the emotions supposedly communicated through text do not match the ones detected, this might have serious implications for automatic emotion detection as already done by sentiment analysis systems, especially for diagnostic purposes, e.g., aimed at online suicide prevention (e.g., Christensen et al., 2014). In this case, the approach to analyzing affect in texts should likely be reconsidered.

Additionally, disparities do not only lead to potential issues for sentiment analysis but also raise questions about the role of self-report and observation in affective science. A discrepancy between authors' intentions and readers' perceptions might either indicate a mismatch between the intentional affect communication by authors and the affect experience of others, or an unawareness of one's own affective states – and hence, their forms of expression.

References

Afzal, S., & Robinson, P. (2009). *Natural affect data—Collection & annotation in a learning context.* Paper presented at the 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops.

Bae, Y., & Lee, H. (2012). Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers. *Journal of the Association for Information Science and Technology, 63*(12), 2521-2535.

Barr, C. L., & Kleck, R. E. (1995). Self-other perception of the intensity of facial expressions of emotion: Do we know what we show? *Journal of personality and social psychology, 68*(4), 608.

Bestgen, Y. (1994). Can emotional valence in stories be determined from words? *Cognition & Emotion, 8*(1), 21-36.

Birch, S. A., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological science, 18*(5), 382-386.

Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., & Vakali, A. (2017). *Mean birds: Detecting aggression and bullying on twitter.* Paper presented at the Proceedings of the 2017 ACM on web science conference.

Christensen, H., Batterham, P., & O'Dea, B. (2014). E-health interventions for suicide prevention. *International journal of environmental research and public health, 11*(8), 8193-8212.

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion, 18*(1), 75.

Dadvar, M., Trieschnigg, D., Ordelman, R., & de Jong, F. (2013). *Improving cyberbullying detection with user context.* Paper presented at the European Conference on Information Retrieval.

Denecke, K., & Deng, Y. (2015). Sentiment analysis in medical settings: New opportunities and challenges. *Artificial intelligence in medicine, 64*(1), 17-27.

Ekman, P. (1992). Are there basic emotions?

Fischer, A. H., Kret, M. E., & Broekens, J. (2018). Gender differences in emotion perception and self-reported

emotional intelligence: A test of the emotion sensitivity hypothesis. *PloS one*, *13*(1).

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications*, *40*(16), 6266-6282.

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of personality and social psychology*, *75*(2), 332.

González-Ibánez, R., Muresan, S., & Wacholder, N. (2011). *Identifying sarcasm in Twitter: a closer look.* Paper presented at the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2.

Greaves, F., Ramirez-Cano, D., Millett, C., Darzi, A., & Donaldson, L. (2013). Use of sentiment analysis for capturing patient experience from free-text comments posted online. *Journal of medical Internet research*, *15*(11), e239.

Hall, J. A., Carter, J. D., & Horgan, T. G. (2000). Gender differences in nonverbal communication of emotion. *Gender and emotion: Social psychological perspectives*, 97-117.

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta psychologica*, *135*(3), 278-283.

Hunston, S., & Thompson, G. (2000). *Evaluation in text: Authorial stance and the construction of discourse: Authorial stance and the construction of discourse*: Oxford University Press, UK.

Keysar, B. (1994). The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive psychology*, *26*(2), 165-208.

Khodak, M., Saunshi, N., & Vodrahalli, K. (2017). A large self-annotated corpus for sarcasm. *arXiv preprint arXiv:1704.05579*.

Kruger, J., Epley, N., Parker, J., & Ng, Z.-W. (2005). Egocentrism over e-mail: Can we communicate as well as we think? *Journal of personality and social psychology*, *89*(6), 925.

Lane, R. D., Lee, S., Reidel, R., Weldon, V., Kaszniak, A., & Schwartz, G. E. (1996). Impaired verbal and nonverbal emotion recognition in alexithymia. *Psychosomatic medicine*, *58*(3), 203-210.

Levant, R. F., Hall, R. J., Williams, C. M., & Hasan, N. T. (2009). Gender differences in alexithymia. *Psychology of Men & Masculinity*, *10*(3), 190.

Liew, J. S. Y., Turtle, H. R., & Liddy, E. D. (2016). *EmoTweet-28: a fine-grained emotion corpus for sentiment analysis.* Paper presented at the Proceedings of the tenth international conference on language resources and evaluation (lrec 2016).

Losada, D. E., & Gamallo, P. (2018). Evaluating and improving lexical resources for detecting signs of depression in text. *Language Resources and Evaluation*, 1-24.

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. *Text and technology: In honour of John Sinclair*, *240*, 251.

Montebarocci, O., Surcinelli, P., Rossi, N., & Baldaro, B. (2011). Alexithymia, verbal ability and emotion recognition. *Psychiatric Quarterly*, *82*(3), 245-252.

Mihalcea, R., & Liu, H. (2006). *A Corpus-based Approach to Finding Happiness.* Paper presented at the AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs.

Müller, C., & Strube, M. (2006). Multi-level annotation of linguistic data with MMAX2. *Corpus technology and language pedagogy: New resources, new tools, new methods*, *3*, 197-214.

Nahar, V., Unankard, S., Li, X., & Pang, C. (2012). *Sentiment analysis for effective detection of cyber bullying.* Paper presented at the Asia-Pacific Web Conference.

Nguyen, T., Phung, D., Dao, B., Venkatesh, S., & Berk, M. (2014). Affective and content analysis of online depression communities. *IEEE Transactions on Affective Computing*, *5*(3), 217-226.

Park, J. H., Xu, P., & Fung, P. (2018). PlusEmo2Vec at SemEval-2018 Task 1: Exploiting emotion knowledge from emoji and# hashtags. *arXiv preprint arXiv:1804.08280*.

Pestian, J. P., Matykiewicz, P., Linn-Gust, M., South, B., Uzuner, O., Wiebe, J., . . . Brew, C. (2012). Sentiment analysis of suicide notes: A shared task. *Biomedical informatics insights*, *5*, BII. S9042.

Riordan, M. A., & Trichtinger, L. A. (2017). Overconfidence at the keyboard: Confidence and accuracy in interpreting affect in e-mail exchanges. *Human Communication Research*, *43*(1), 1-24.

Rude, S., Gortner, E.-M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, *18*(8), 1121-1133.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of personality and social psychology*, *48*(4), 813.

Tiedens, L. Z. (2001). Anger and advancement versus sadness and subjugation: the effect of negative