



# The Open Psychology Journal

Content list available at: <https://openpsychologyjournal.com>



## RESEARCH ARTICLE

### Emojis in Deceptive Online Communication: The Frequency and Type of Emoji in Deceptive and NonDeceptive Online Messages

Michelle Walther<sup>1,\*</sup>, Petar Petrov<sup>2</sup>, Ryan Smith<sup>2</sup>, Danielle Young<sup>2</sup>, Peter W. de Vries<sup>1</sup> and Marielle Stel<sup>1</sup>

<sup>1</sup>Department of Psychology of Conflict, University of Twente, Risk, and Safety, Enschede, Netherlands

<sup>2</sup>Department of Psychology, University of Strathclyde, Stirling, Scotland

#### Abstract:

##### Background:

Little research has been done on nonverbal deception cues in computer-mediated communication (CMC). However, deception is a daily occurrence and since much communication is shifting towards CMC, it is important to understand the difference between truthful and deceptive messages.

##### Objective:

This research obtained more insight in the use of emoji in deceptive messages by answering the question: Are the frequency and type of emoji different in deceptive compared to truthful online messages?

##### Methods:

Participants sent three screenshots of deceptive, and truthful messages to WhatsApp. The used emoji were counted and sorted into levels of valence (positive, negative, and neutral) and intensity (strong versus weak).

##### Results:

The results indicated that participants used more negative, weak emoji in deceptive compared with truthful messages and more positive, weak, and strong emoji in truth compared with deceit. No difference was found for the emoji frequency.

##### Discussion:

The results are discussed in the light of earlier research. However, this is the first study investigating the use of emoji in the context of computer-mediated deception.

##### Conclusion:

The type of emoji can be indicative of the veracity as a nonverbal deception cue in online messages.

**Keywords::** Deception, Emoji, Nonverbal cues, Computer-mediated communication, Online messages, Online communication.

#### Article History

Received: February 25, 2022

Revised: March 8, 2022

Accepted: March 31, 2022

## 1. INTRODUCTION

People deceive on a daily basis [1, 2], mostly about feelings, actions, plans, and whereabouts [2]. In this paper, we define deception as the intentional act of spreading wrong information that is communicated to a receiver to achieve a false impression [2]. Deception research has mainly focused on how people can tell when a person is deceptive in face-to-face

(ftf) interaction [2, 3]. Nowadays, however, communication is not limited to ftf interactions but goes beyond it by means of computer-mediated communication (CMC), such as instant messaging [4]. Therefore, it is important to gain insight into the deception cues used in this medium. In this paper, we focus on the use of emojis when people deceive in online messages.

### 1.1. Cues to Deception

Nonverbal and verbal cues can indicate that a person is deceiving. Although the findings pertaining to verbal deception

\* Address correspondence to this author at the Department of Psychology of Conflict, Risk, and Safety, De Zul 10, 7522NJ Enschede, The Netherlands; Tel: +49(0) 151 20292959; E-mail: [michelle.walther@gmx.net](mailto:michelle.walther@gmx.net)

cues appear contradictory, meta-analyses have found that deceivers are more informal and expressive than truth tellers [5, 6]. Furthermore, deceivers convey their story with a higher pitch of voice, shorter time of talking, and fewer details than truth tellers [3]. Deceivers also differ in linguistic style compared to truth tellers, by showing lower cognitive complexity, using fewer self-references and other references, and using more negative emotional words [7]. Furthermore, research showed that truth tellers have lengthier text messages than deceivers [3].

Multiple indicators of deception can be found in the body language of a deceiver, such as using fewer illustrators, more dilated pupils, and adopting a more open posture [3, 4, 8]. There is, however, no set behaviour or group of behaviours only occurring while deceiving. Therefore, these behaviours are only more or less likely to occur when deceiving and as such do not flawlessly predict deception [3, 8, 9].

### 1.2. Computer-mediated Communication

The above-mentioned (non)verbal deception cues are predominantly researched in ftf interaction, whereas written communication *via* computers increases. Therefore, computer-mediated communication increases, by which we mean all communication that is done *via* some sort of device, including computers, mobile phones, laptops, and tablets and therefore including e-mails, text messages *via* online messengers or social media, but also video calls [10]. Most CMC interactions occur through text as opposed to for example video calls [11]. Communications mediated by computers differs from ftf communication in two important ways. First, the written CMC lacks the immediacy of distance and time of ftf communication. The asynchronous nature of most computer-mediated conversations means that more thoughts can be put into messages [1]. Second, the nature of online communication is more disembodied than ftf communication: written CMC lacks access to nonverbal visual and auditory cues which are usually present in ftf communication. The lack of nonverbal signals has been associated with difficulty interpreting the tone of the messages [12].

Deceptive cues in CMC found so far regard textual cues. As the act of deception places a demand on people's cognitive functions, this may affect people's language use in CMC. This is supported by the results of Toma and Hancock [13] showing that deceivers had shorter descriptive bodies of text in their online dating profiles compared to truth tellers. Research showed mixed evidence regarding the average sentence length and level of informality [5, 14, 15]. As most CMC research focused on verbal cues of deception [4, 16], it is important to gain insight into nonverbal cues of deception in CMC. Even though CMC lacks subtle nonverbal behaviours as displayed during ftf interactions, compensatory mechanisms have been developed for these behaviours: digital behaviours such as, but not exclusively, emojis [17 - 20].

### 1.3. Emoji

Emojis denote graphical representations of facial and/or emotional expressions which mostly follow messages in written CMC [20 - 23]. Evolving since their implementation in

1982, emojis nowadays have various forms and meanings. Their use can have two main underlying motives: (1) aiding personal expressions, such as establishing emotional tone and lightening the mood and (2) reducing ambiguity of discourse [16, 19, 21, 22, 24 - 26]. Veytia-Bucheli *et al.* [27] found that people use emojis (1) in social interactions, (2) for expressing their emotions and feelings, (3) for understanding utensils, and (4) as part of their writing style. The five most commonly used emojis were the happy face (42%), the smile (22%), finger pointed up (21%) and angry and sadness emojis (each 18%) [27]. Arafah and Hasyim [28] found, furthermore, that emojis are part of the grammar of digital written communication, being used as punctuation at the end of a sentence.

The ability to recognize the emotion of the emoji depends on culture: The extent to which people are exposed to emojis in their culture influences their emotion recognition ability [29]. Generally, emoji help in understanding the emotions, attitudes, and attention that the sender intends to communicate to the reader [16, 21, 22]. Boutet *et al.* [30] showed that emojis help to intensify the perceived mood of the sender in a positive way with positive emojis and in a negative way with negative emojis. Additionally, they found that the processing speed and understanding of verbal messages increased when emojis were added in a congruent way [30]. Kimura-Thollander and Kumar [31] found that emojis help to establish a more close personal connection and overcome language differences and help to express peoples' cultural identities. They are, however, not the same as emotions or gestures [19, 21, 22]. Nonverbal behavior is mostly nonintentional and not always controllable, which makes it more trustworthy in showing a sender's intentions. Emojis, on the other hand, are intentional and added to messages in a controlled way and thus are less trustworthy [19, 22].

Research indicated that higher use of emoji results in lower detection of deceit [14]. Apparently, receivers are less skeptical when the deceiver tries to cover up their real emotions and thoughts by supporting their messages with more emojis. Briscoe and colleagues [14] argued that the use of emoji indicates that the sender wants to build a relationship and therefore the receiver is less likely to assume deception. However, the role of emojis as deception cues and therefore how they could be used in the detection of deception has not received much attention so far [32].

### 1.4. The Current Study

The current study investigates the use of emoji in deceptive messages compared to truthful messages. Leading to the research question: *How does emoji frequency and type differ between deceitful and truthful messages?*

Two contrasting hypotheses can be expected based on previous research. In the current study, we explore whether the *emoji frequency between deceitful and truthful messages varies*. On the one hand, one can expect *deceptive messages to contain more emoji than truthful messages*. As research indicates that deceivers tend to be intentionally more expressive to make their story believable [5, 6, 15] and want to appear honest and friendly through open and informal body language [16, 19, 21, 24, 26], deceivers can be expected to use

more emojis in their expressions. Assuming that emojis are used to support a message intentionally [4], one would expect deceivers to deliberately increase the believability of their story by being more expressive [5, 15]. On the other hand, one can expect *deceptive messages to contain fewer emojis than truthful messages* since DePaulo and colleagues [3] found that deceivers (unintentionally) use fewer illustrators and have less embraced statements.

Even though people are more likely to deceive about positive than negative matters [2], research indicated that deceivers use more negations [13, 33], more negative emotion words [7], and more negative statements and complaints in their messages [3] compared with truth tellers. Therefore, our third hypothesis is: *deceptive messages contain more negative emojis than truthful messages*.

To explore the use of emojis in deceptive and truthful messages, we chose an observational procedure. Participants were asked to fill in an online questionnaire and add screenshots of six conversations they had in the last via WhatsApp messenger. These six conversations should include three deceptive and three truthful answers to the questions participants were asked. The frequency and type (valence and intensity) of emoji were assessed for both deceptive and truthful messages. As our main aim of the study is to explore whether a single nonverbal emoji cue can predict deception, we choose to investigate the frequency and type of emoji in deceptive versus non-deceptive messages, regardless of the specific verbal content. In this study, we focused on the use of emojis in social networking rather than email as emojis are more often used in social networking [34]. We chose to focus on WhatsApp messages as WhatsApp is one of the most used CMC [35]. As such, WhatsApp may well provide a communication channel just as amenable to deception attempts as ftf communication. Additionally, WhatsApp has a gallery of emojis and GIF's that can be shared with others to indicate nonverbal cues and could therefore be used as cues of deception [4, 36].

## 2. MATERIALS AND METHODS

This study was approved by the Ethics Committee of the University of Twente in Enschede, the Netherlands (approval number 18001). Furthermore, the study was preregistered in AsPredicted (<https://aspredicted.org/7t8ub.pdf>).

### 2.1. Design

A within-participant design was used in which the frequency and type (positive vs negative) of emojis were compared between deceitful and non-deceitful messages. The dependent variable was the frequency of emojis per participant. The analysis of the third hypothesis was done per valence (positive versus negative).

### 2.2. Participants

#### 2.2.1. Power Analysis

A power analysis showed that the optimal sample size for a paired sample with a power of .80 ( $1-\beta$ ), a type 1 error rate of 5%, and an effect size of .1 is 620 pairs, of each one deceptive

and one nondeceptive message. Since each participant was asked to upload three deceptive and three truthful conversations and therefore would each provide three pairs of one deceptive and one nondeceptive message, the sample size was divided by three. This results in 207 participants ideally needed to obtain a power of .80 ( $1-\beta$ ).

#### 2.2.2. Requirements

Participants had to meet several requirements. Since a part of the participant pool would be retrieved from the University of Twente (with students mostly living in the Netherlands and Germany) *via* a platform where students get credit for participating, the requirements for all participants included that they had to be at least 18 years of age, live in the Netherlands or Germany. Another requirement was that, since our study focused on WhatsApp messenger as CMC, participants had to use WhatsApp at least five times per week. Because the questionnaire that participants had to fill in was done in English, participants who declared that their English skills are less than 'Average' (on a 5-point Likert-scale from 'Excellent' to 'Terrible'), were not included in the analysis. However, since English would not be the native language of most participants, participants were free to upload messages that were written in German, English, and Dutch in their text messages on WhatsApp, since these are the languages that could be translated by the researcher.

To make sure that a participant would have opportunities to deceive on WhatsApp, participants who used the texting application less than five times per week were not included in the analysis. Participants were asked to include six conversations which each include either a truthful or a deceptive message in response to a question. Therefore, deceptive/truthful messages that were not answers to the questions were excluded from the analysis. Furthermore, to exclude conversations with 'white lies' where the stakes are minimal and therefore leakage or deception cues would be minimal, screenshots that included answers to questions such as, 'Are you okay?', 'How old are you?' or 'How are you?' were not included in the analysis. These requirements were also included in the pre-registration of AsPredicted. As the last requirement, participants had to report at least one deceit and one truth so that there was enough material to compare deceptive with truthful messages.

#### 2.2.3. Demographics

In total, 281 people participated in this study, of which 89 were not included in the analysis because they did not meet the requirements (n=22 used WhatsApp less than five times per week, n=did not fill in their English skill level, there was no one excluded because of a too low English skill level, n=1 did not live in Germany or the Netherlands) or only partly filled in the study, for example, by leaving out screenshots (n=64). The mean age of the remaining 192 participants (40 male; 152 female) was 20.52 years ( $SD = 1.88$ ). Furthermore, 34 participants had Dutch nationality, 142 had German nationality, and 16 had a different nationality. All participants were students that participated in the study to earn credit points. A total of 1114 truthful and deceptive screenshots were uploaded by the 192 participants included in the analysis. Of

the deceptive and nondeceptive messages that were uploaded by the participants,  $n=66$  deceptive messages were in Dutch,  $n=342$  deceptive messages were in German,  $n=147$  deceptive messages were in English. Of the nondeceptive messages  $n=71$  were in Dutch,  $n=346$  nondeceptive messages were in German and  $n=142$  nondeceptive messages were in English. As the language was not equally divided among participants/messages and there is a lack of power, we did not include language in our analyses.

### 2.3. Procedure and Materials

The study was an online questionnaire that was published online *via* Qualtrics. Participants were asked to answer multiple questions about their demographics, the conversations they had with WhatsApp, and were asked to include screenshots from said conversations.

The study took participants five to ten minutes during five or six different moments in a week or about 50 to 60 minutes once. Participants read the informed consent and, after agreeing, they continued reading the instructions. In the instructions, the participants were asked to fill in the questionnaire and upload three deceptive and three nondeceptive messages of WhatsApp conversations. They were informed that they were able to access the questionnaire multiple times during a period of seven days, therefore, they could choose to collect the six messages during that week and fill in the questionnaire at the end of the week or to access the questionnaire multiple times during that week and upload the messages on different occasions. To prevent biased results, they were not fully informed about the true goals of the study; instead, they learned that it investigated deception cues in WhatsApp.

Then they filled in eight demographic questions about age, gender, highest educational status, nationality, country of residence, whether they were born in the Netherlands or Germany, and if not, for how long they had lived in either of the countries, WhatsApp usage per week, and the English skills of the participants.

Participants were then asked to disclose, during the participation, three WhatsApp messages in which they attempted to deceive the recipient, and three messages in which they told the truth. These could be either previously sent messages or messages sent during the week of participation and had to be messages which constituted answers to questions asked by the recipient. Disclosure of deceptive and truthful messages was done by taking screenshots, in which participants marked the part that constituted the lie or the truth and removed the name and other identifiable details of the person they were communicating with for confidentiality reasons.

As a backup, in case the screenshot was not uploaded or marked correctly, participants were also asked to fill in the exact text of the messages. Moreover, they were asked to indicate whether their message was truthful or deceptive. At the end of each form, the participants were able to upload their screenshot<sup>1</sup>, which was used for the analyses.

Whenever the participants came back to the Qualtrics-link,

they were able to continue where they stopped. After the last session, participants were thanked for their participation and were given the opportunity to be debriefed.

#### 2.3.1. Coding System of Emoji

To narrow the research down, only emojis on WhatsApp were included, reflecting facial expressions ( $N = 60$ ). These emojis were sorted into one of the four main categories of valence (neutral, positive, negative, or surprising) and six subcategories of intensity (weak and strong). This resulted in seven categories (neutral, positive-weak, positive-strong, negative-weak, negative-strong, surprising-weak, & surprising-strong) since no distinction was made between weak and strong intensity in the neutral valence. The valence surprising was added because a surprising state can either be positive or negative but is not a neutral expression [37 - 39].

The coding of the emojis was done because there was no suitable categorization found in previous studies. Rodrigues *et al.* [40] for example categorized emojis and emoticons on seven dimensions (aesthetic appeal, familiarity, visual complexity, clarity, valence, arousal, and meaningfulness). Their valence was measured by asking 'To what extent do you consider this stimulus refers to something positive/pleasant or negative/unpleasant'. In our study, however, we wanted to distinguish between strong and weak emojis as well as neutral and surprising emojis and only include facial expressions.

The emoji were independently sorted into different categories by four individual raters. These scorings were then compared and emoji that were sorted in the same category by all four individuals ( $N = 14$ ) and by three of the four individuals ( $N = 31$ ) were sorted into the final table. Finally, emojis that only two individuals rated similarly were checked for differences ( $N = 15$ ). For these 15 emojis, there were two individuals that sorted the emoji in the same valence and intensity, while a third individual chose the same valence but disagreed on the intensity. The fourth individual disagreed with both valence and intensity; therefore, it was chosen to sort the emoji in the valence of the three individuals and the intensity of the two raters agreeing with each other. No discussion occurred to confine disagreements. In addition to pre-classified emoji, participants also used 6 textual portrayals of facial expressions in the form of icons, which were additionally coded by three raters with complete agreement. Emoji that were used in the conversations that are not included in these categories were sorted as 'others'. See Appendix A for the division into the categories of emoji.

### 3. RESULTS

The data of the reported studies are available *via* the Open Science Framework (<https://osf.io/yh6zu/>). A test for normal distribution (Shapiro-Wilk test) on all relevant variables showed that there is a violation of the general assumption of normality,  $p < .01$  for the number of emoji used in truthful and deceptive messages, respectively  $W(192) = 0.89$ ,  $p < .001$  and  $W(192) = 0.92$ ,  $p < .001$ , and for the number of words in truthful ( $W(192) = 0.88$ ,  $p < .001$ ) and deceptive messages ( $W(192) = 0.90$ ,  $p < .001$ ). Log 10 transformations did not have sufficient effects. Therefore, Wilcoxon Signed Rank Tests for

non-parametrical data were used to compare means between two related variables in a sample, instead of the preregistered repeated measures ANOVA. Furthermore, the neutral and surprising emojis were excluded from the analysis since they were barely used (neutral emojis: Lie n=18, Truth n=8; surprising emojis: Lie n=18, Truth n=11; positive emojis: Lie n=143, Truth n=217; negative emojis: Lie n=95, Truth n=42) in the screenshots. Since there were a total of 1114 truthful and deceptive screenshots (557 pairs) which we included in the analysis, uploaded by 192 participants, the achieved power of this study was .74 with a type 1 error rate of 5% and an effect size of .1.

**3.1. The Number of Emoji**

To explore differences of the frequency of emoji used in deceptive versus truthful messages, a Wilcoxon Signed Rank Test was conducted with the content (deceit versus truth) as independent variable and the number of emoji as the dependent variable. The analysis showed that participants did not use different numbers of emoji in deceptive messages [*Median* = 0.67, *M* = 0.68] compared with truthful messages [*Median* = 0.67, *M* = 0.68], two-sided Monte-Carlo *p* = .79, *Z* = -0.27, 95% CI [0.782, 0.798], *d* = .039,  $\eta^2 < .001$ .

**3.2. The Type of Emoji**

To investigate the hypothesis that more negative than positive emojis are used in deceit compared to truth, two Wilcoxon Signed Rank Tests were conducted. The first test was conducted within the positive valence of emoji with the content (deceit versus truth) as the independent variable and frequency of emoji as the dependent variable. This test indicated that participants used significantly more positive emojis in truthful messages [*Median* = 0.17; *M* = 0.19] compared to deceptive messages [*Median* = 0.17; *M* = 0.12], two-sided Monte-Carlo *p* < .01, *Z* = - 3.48, 95% CI [0.000, 0.001], *d* = .518,  $\eta^2 = .063$ .

The second test was conducted within the negative valence of emoji with the content (deceit versus truth) as the independent variable and frequency of emoji as the dependent variable. It showed that participants used significantly more negative emojis while deceiving [*Median* < 0.01; *M* = 0.08] compared to telling the truth [*Median* < 0.01; *M* = 0.04], two-sided Monte-Carlo *p* < .01, *Z* = - 4.22, 95% CI [0.000, 0.000], *d* = .640,  $\eta^2 = .093$ .

Moreover, four Wilcoxon Signed Rank Tests were conducted to investigate whether these effects hold for different intensities (strong versus weak) of the emoji (Table 1) for the mean number of emoji per message). The tests were done within the valence (positive versus negative) and intensity (strong versus weak) of emojis, with content (lie versus truth) as the independent variable and frequency of emojis as the dependent variable.

First, participants used significantly more positive strong emoji in truth [*Median* < 0.01, *M* = 0.13] compared to lies [*Median* < 0.01, *M* = 0.07], two-sided Monte-Carlo *p* = .01, *Z* = -2.52, 95% CI [0.009, 0.013], *d* = .37,  $\eta^2 = .033$ . Second, participants also used significantly more positive weak emoji while telling the truth [*Median* < 0.01, *M* = 0.24] compared with telling a lie [*Median* < 0.01, *M* = 0.17], two-sided Monte-

Carlo *p* = .01, *Z* = -2.45, 95% CI [0.010, 0.014], *d* = .36,  $\eta^2 = .031$ . Third, participants did not differ in the number of negative strong emoji in deceptive [*Median* < 0.01, *M* = 0.04] than truthful messages [*Median* < 0.01, *M* = 0.02], two-sided Monte-Carlo *p* = .50, *Z* = 0.68, 95% CI [0.528, 0.547], *d* = .10,  $\eta^2 = .002$ . Finally, participants used significantly more negative weak emoji while telling a lie [*Median* < 0.01, *M* = 0.13] compared to telling the truth [*Median* < 0.01, *M* = 0.05], two-sided Monte-Carlo *p* < .01, *Z* = 4.61, 95% CI [0.000, 0.000], *d* = .71,  $\eta^2 = .111$ . These four tests indicated that participants used significantly more positive strong and positive weak emoji in truths compared with lies and used more negative weak emojis in lies compared with truths.

**Table 1. 'Indicated number of emojis per message for deceitful and truthful messages per valence and intensity.**

	Positive ( <i>M</i> )*		Negative ( <i>M</i> )*	
Lie	0.12		0.08	
Truth	0.19		0.04	
	Strong ( <i>M</i> )*	Weak ( <i>M</i> )*	Strong ( <i>M</i> )	Weak ( <i>M</i> )*
Lie	0.07	0.17	0.04	0.13
Truth	0.13	0.24	0.02	0.05

\* significant difference (*p*<.05).

**3.3. Potential Differences in Message Length**

As stated earlier research showed that truth tellers have lengthier text messages than deceivers [3], it is conceivable that the number of emoji is affected by the length of text. We did not include this in the hypotheses but after exploring the data we wanted to check for the potential differences in message length. Therefore, an additional explorative analysis was conducted. A Wilcoxon Signed Rank test was conducted to investigate whether deceivers and truth tellers differed in the number of words used in their messages. The test indicated that participants did not use more words in deceptive [*Median* = 9.00, *M* = 10.24] compared with truthful messages [*Median* = 8.17, *M* = 9.64], two-sided Monte-Carlo *p* = .10, *Z* = -1.67, 95% CI [0.094, 0.106], *d* = .243,  $\eta^2 = .015^2$ .

**4. DISCUSSION**

This study investigated whether the frequency and type of emoji differ between deceptive and truthful messages. The results revealed no difference in the frequency of emojis between deceptive and truthful messages. However, the type of emoji did differ: in deceptive messages, less positive, strong and weak emoji and more negative weak emoji were used than in truthful messages.

The results indicate that it is not the frequency of emoji, but the type of emoji indicating deception. The finding that less positive and more negative emojis were used in deceptive than truthful messages can be explained by deceivers expressing more negativity in their messages than truth tellers. This is in line with previous research showing that more negations and more negative emotion words were used by deceivers [7, 13, 33]. Additionally, DePaulo and colleagues [3] concluded that deceptive messages contain more negative statements and complaints than truthful messages. As in negative messages, more negative emoji are used [41], it is conceivable that the more negative messages of deceivers elicited more use of negative emoji, but only the negative emoji of weaker

intensity.

The current study is the first to shed light on the use of emojis under truthful and deceptive conditions. However, no actual deception manipulation occurred as participants were asked to select deceptive and truthful messages from their logged WhatsApp conversations rather than being instructed to create a deceptive or truthful message. Hence, no causal relationship could be established, leaving open the possibility that other factors may have influenced emoji use. Although this could be seen as a limitation, one could also argue that deception manipulation in the current context inevitably led to a contrived study with possibly limited results, and, at best, limited ecological validity. The current study, however, focused on past instances of communication, and, as such, on actual communication behavior, displayed in real life with peers or relations. In turn, this allows for greater validity and generalizability of the results. Since communication *via* WhatsApp is mostly informal between family, friends, and classmates [42], the results cannot be generalized to more formal communication, such as work-related e-mails.

Our explorative analysis of the potential difference in message length was not in line with previous research stating that truth-tellers have lengthier text messages than deceivers [3]. Briscoe and colleagues [14] and Zhou and Zhang [15] found that there are more words and more sentences used in deceptive messages. It is hard to speculate why we obtained different results. A possibility could be that most of the studies explored in the meta-analysis [3] were face-to-face studies and did not analyze written accounts. Briscoe and colleagues [14] and Zhou and Zhang [15] concentrated on social media communication. However, both studies were experimental lab designs with scripts while our study was an observational design in an uncontrolled environment. Future research should explore the differences between written and spoken accounts in lies with both field data and experimental lab studies.

A further point with regard to the messages is that the deceptive and nondeceptive messages were offered by what users perceive as respectively deceptive or nondeceptive. It is conceivable that users selected the messages in which the deception/truth was very clear. The messages we found were mostly both, truthful and deceptive messages, about future plans, if they wanted to go to a party or what their plans for the weekend were, and about their studies, if they already finished them. We furthermore found that participants send screenshots from conversations about meetings with friends if they wanted to meet. In deceptive messages, we found that having to work or having to study was used frequently used if they cancelled a meeting or party.

We would like to outline that as the languages of the screenshots differed within participants and were not evenly distributed among the messages and since we had more German participants than Dutch participants or participants of other nationalities, we did not include language and cultural differences in our analysis. However, it is important to mention that cues to deception can vary across cultures and languages and whether the person speaks in their first or second language. Research showed that there are cultural and language differences in verbal (spoken and written) communication [31, 33, 43 - 45]. Liu [43], for example, found that there is a difference between low-context communication

(predominantly used in individualistic cultures) and high-context communication (predominantly used in collectivistic cultures). In low-context communication, the meaning of a message is mostly made explicit in the verbal message, while in high-context communication the meaning of the message is conveyed through physical context or internalized in the person [43]. This means that in low-context communication lie detection cues can be found more likely in the actual message while in high-context communication non-verbal lie detection cues are more likely. Since Dutch and German cultures are both predominantly low-context communication, individualistic cultures' cultural differences can be expected to be low. However, future research should examine this.

Berebey-Meyer *et al.* [46] found that people are less inclined to lie in spoken language when using a different language than their native tongue. This might be because a second language is used with more deliberation and therefore reduce the temptation to lie [46]. For our study, this would implicate that people lied more in their native tongue. Cheng *et al.* [47] found that when lying in their second language, participants displayed more non-verbal cues of deception in face-to-face communication, due to more cognitive load. Future research should investigate this for written and non-verbal communication.

We would also like to address that the categorization of the emojis was done by only four judges, while the categorization done by Rodriguez *et al.* [40] was done by 505 participants. However, in qualitative research, it is a common practice to only have two or three raters, such as in the coding of interviews [48] or the selection of research papers in a systematic literature review [49]. Furthermore, in Ott *et al.* [50] three human judges were used to categorize deceptive and nondeceptive reviews. We, therefore, argue that four judges at this point are sufficient for our research, but it should still be handled with caution in further research.

With the digitalization of society, new ways of communication will keep emerging. Consequently, new cues of deception will be displayed. The communication possibilities offered by the specific media influence the way in which people can differentiate when deceiving and telling the truth, and thus which deception cues may be used in that specific media. The deception cues found in one media may thus differ from deception cues as displayed in other media. Importantly, however, as emojis are incorporated into our smartphone keyboards, people can use the emoji on WhatsApp, Facebook, and other messaging systems. Therefore, we expect that our results will generalize to all online media in which there is a possibility of communicating with the use of emoji.

Thus far, research has either concentrated on verbal and nonverbal deception cues in ftf interaction or on verbal/linguistic deception cues in CMC. To the researchers' knowledge, no research has been done on the use of emoji use in deceptive and truthful online messages, and if emojis could be used as nonverbal deception cues. This is the first study investigating the use of emojis in the context of computer-mediated deception.

**4.1. Implications for Practice**

The practical implication of the knowledge about how to identify lies on social media can on the one hand help when communicating with friends and family *via* WhatsApp, for identifying when somebody lies. Especially now in pandemic times, when most communication is not face-to-face. On the other hand, the analysis of deception detection cues in social media can also help law enforcement identify fraudsters online.

**CONCLUSION**

The results showed that the frequency of emoji cannot indicate if a message is deceptive or truthful, however, the type of emoji can. Therefore, we provided a new cue to deception detection. As with any other deception cue so far, the type of emoji is not 100% reliable to predict deception [3]. However, the results show that the use of weak negative emoji can indicate deception, while positive, weak, and strong emoji can indicate truth, allowing the first step towards deception detection in online messages.

**LIST OF ABBREVIATIONS**

- FTF** = Face to Face
- CMC** = Computer Mediated Communication

**ETHICAL STATEMENT**

This study was approved by the Ethics Committee of the University of Twente in Enschede, the Netherlands (approval number 18001).

**CONSENT FOR PUBLICATION**

All authors and the University of Twente consent to the publication of this article. The participants declared that their data may be used anonymously in a publication.

**AVAILABILITY DATA AND MATERIALS**

The data of the reported studies are available *via* the Open Science Framework (<https://osf.io/yh6zu/>). The data was acquired by the authors during the project, there was no third source involved.

**FUNDING**

None.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest, financial or otherwise.

**ACKNOWLEDGEMENTS**

The author contribution are as follows: The first author wrote most of the article as part of her Master's thesis at the University of Twente. The idea of the design and research direction came from Petar Petrov, Ryan Smith, Danielle Young. The supervisors, Dr. Marielle Stel and Dr. Peter W. de Vries, were involved in the detailed design idea and throughout the whole project, the editing process and approval decision to publish the article.

**APPENDICES**

**Appendix A. Emojis sorted by the raters in valence (positive, negative, neutral and surprising) and intensity (strong and weak).**

Positive Emoji		Negative Emoji		Neutral Emoji	Surprising Emoji	
Strong	Weak	Strong	Weak		Strong	Weak
	:)/:-) :D/:-D :p/:P/:-P/:-p		:(/:-( 			:o/:-o/:- O:/O

- study during middle adolescence. *Int J Adolesc Youth* 2018; 24: 1-13.
- [12] Kiesler S, Siegel J, McGuire TW. Social psychological aspects of computer-mediated communication. *Am Psychol* 1984; 39(10): 1123-34.  
[http://dx.doi.org/10.1037/0003-066X.39.10.1123]
- [13] Toma CL, Hancock JT. What lies beneath: The linguistic traces of deception in online dating profiles. *J Commun* 2012; 62(1): 78-97.  
[http://dx.doi.org/10.1111/j.1460-2466.2011.01619.x]
- [14] Briscoe EJ, Appling DS, Hayes H. Cues to deception in social media communications. 2014 47th Hawaii International Conference on System Sciences. 1435-43.  
[http://dx.doi.org/10.1109/HICSS.2014.186]
- [15] Zhou L, Zhang D. Following linguistic footprints. *Commun ACM* 2008; 51(9): 119-22.  
[http://dx.doi.org/10.1145/1378727.1389972]
- [16] Lo SK. The nonverbal communication functions of emoticons in computer-mediated communication. *Cyberpsychol Behav* 2008; 11(5): 595-7.  
[http://dx.doi.org/10.1089/cpb.2007.0132] [PMID: 18817486]
- [17] Nelson RA, Tossell CC, Kortum P. Emoticon Use in Mobile Communications. *Encyclopedia of Mobile Phone Behavior*. Hershey: IGI Global 2015; pp. 1-11. <https://www.igi-global.com/chapter/emoticon-use-in-mobile-communications-/www.igi-global.com/chapter/emoticon-use-in-mobile-communication-s-/130123> Internet  
[http://dx.doi.org/10.4018/978-1-4666-8239-9.ch001]
- [18] Thompson D, Filik R. Sarcasm in Written Communication: Emoticons are Efficient Markers of Intention. *J Comput Mediat Commun* 2016; 21(2): 105-20.  
[http://dx.doi.org/10.1111/jcc4.12156]
- [19] Walther JB, D'Addario KP. The impacts of emoticons on message interpretation in computer-mediated communication. *Soc Sci Comput Rev* 2001; 19(3): 324-47.  
[http://dx.doi.org/10.1177/089443930101900307]
- [20] Gawne L, McCulloch G. Emoji as digital gestures 2019. <https://www.languageatinternet.org/articles/2019/gawne>
- [21] Dresner E, Herring SC. Functions of the Nonverbal in CMC: Emoticons and Illocutionary Force. *Commun Theory* 2010; 20(3): 249-68.  
[http://dx.doi.org/10.1111/j.1468-2885.2010.01362.x]
- [22] Kralj Novak P, Smailović J, Sluban B, Mozetič I. Sentiment of Emojis. *PLoS One* 2015; 10(12):e0144296  
[http://dx.doi.org/10.1371/journal.pone.0144296] [PMID: 26641093]
- [23] Tang Y, Hew KF. Emoticon, emoji, and sticker use in computer-mediated communication: A review of theories and research findings. *Int J Commun* 2019; 13(0): 27.
- [24] Derks D, Bos AER, von Grumbkow J. Emoticons in computer-mediated communication: social motives and social context. *Cyberpsychol Behav* 2008; 11(1): 99-101.  
[http://dx.doi.org/10.1089/cpb.2007.9926] [PMID: 18275321]
- [25] Kaye LK, Malone SA, Wall HJ. Emojis: Insights, affordances, and possibilities for psychological science. *Trends Cogn Sci* 2017; 21(2): 66-8.  
[http://dx.doi.org/10.1016/j.tics.2016.10.007] [PMID: 28107838]
- [26] Skovholt K, Grønning A, Kankaanranta A. The communicative functions of emoticons in workplace e-mails: -. *J Comput Mediat Commun* 2014; 19(4): 780-97.  
[http://dx.doi.org/10.1111/jcc4.12063]
- [27] Veytia-Bucheli MG, Gómez-Galán J, Vergara D. Presence of new forms of intercultural communication in higher education: Emojis and social interactions through whatsapp among graduate students. *Educ Sci (Basel)* 2020; 10(11): 295.  
[http://dx.doi.org/10.3390/educsci10110295]
- [28] Arafah B, Hasyim M. The Language of Emoji in Social Media. *KnE Soc Sci* 2019; 1: 494-504.  
[http://dx.doi.org/10.18502/kss.v3i19.4880]
- [29] Takahashi K, Oishi T, Shimada M. Is ☺ Smiling? Cross-Cultural Study on Recognition of Emoticon's Emotion. *J Cross Cult Psychol* 2017; 48(10): 1578-86.  
[http://dx.doi.org/10.1177/0022022117734372]
- [30] Boutet I, LeBlanc M, Chamberland JA, Collin CA. Emojis influence emotional communication, social attributions, and information processing. *Comput Human Behav* 2021; 119106722  
[http://dx.doi.org/10.1016/j.chb.2021.106722]
- [31] Kimura-Thollander P, Kumar N. Examining the "Global" Language of Emojis: Designing for Cultural Representation In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* [Internet] Glasgow Scotland UK: ACM. 2019. <https://dl.acm.org/doi/10.1145/3290605.3300725>
- [32] Bai Q, Dan Q, Mu Z, Yang M. A Systematic Review of Emojis: Current Research and Future Perspectives. *Front Psychol* 2019; 10: 2221.  
[http://dx.doi.org/10.3389/fpsyg.2019.02221] [PMID: 31681068]
- [33] Taylor PJ, Lerner S, Conchie SM, Menacere T. Culture moderates changes in linguistic self-presentation and detail provision when deceiving others. *R Soc Open Sci* 2017; 4(6):170128  
[http://dx.doi.org/10.1098/rsos.170128] [PMID: 28680668]
- [34] Kaye LK, Wall HJ, Malone SA. "Turn that frown upside-down": A contextual account of emoticon usage on different virtual platforms. *Comput Human Behav* 2016; 60: 463-7.  
[http://dx.doi.org/10.1016/j.chb.2016.02.088]
- [35] O'Hara KP, Massimi M, Harper R, Rubens S, Morris J. Everyday dwelling with WhatsApp. *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* [Internet].  
[http://dx.doi.org/10.1145/2531602.2531679]
- [36] Utz S. Social information processing in MUDs: The development of friendships in virtual worlds. *J Online Behav* 2000; 1: 1.
- [37] Noordewier MK, Breugelmann SM. On the valence of surprise. *Cogn Emotion* 2013; 27(7): 1326-34.  
[http://dx.doi.org/10.1080/02699931.2013.777660] [PMID: 23560688]
- [38] Noordewier MK, Topolinski S, Van Dijk E. The temporal dynamics of surprise. *Soc Personal Psychol Compass* 2016; 10(3): 136-49.  
[http://dx.doi.org/10.1111/spc3.12242]
- [39] Noordewier MK, van Dijk E. Surprise: Unfolding of facial expressions. *Cogn Emotion* 2019; 33(5): 915-30.  
[http://dx.doi.org/10.1080/02699931.2018.1517730] [PMID: 30185108]
- [40] Rodrigues D, Prada M, Gaspar R, Garrido MV, Lopes D. Lisbon Emoji and Emoticon Database (LEED): Norms for emoji and emoticons in seven evaluative dimensions. *Behav Res Methods* 2018; 50(1): 392-405.  
[http://dx.doi.org/10.3758/s13428-017-0878-6] [PMID: 28364283]
- [41] Derks D, Bos AER, Grumbkow J. Emoticons and social interaction on the Internet: the importance of social context. *Comput Human Behav* 2007; 23(1): 842-9.  
[http://dx.doi.org/10.1016/j.chb.2004.11.013]
- [42] Gasaymeh A-MM. University students use of whatsapp and their perceptions regarding its possible integration into their education 2017. <https://computerresearch.org/index.php/computer/article/view/1641>
- [43] Liu M. Verbal Communication Styles and Culture 2016. <https://oxfordre.com/communication/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-162>  
[http://dx.doi.org/10.1093/acrefore/9780190228613.013.162]
- [44] Leal S, Vrij A, Vernham Z, et al. Cross-cultural verbal deception. *Leg Criminol Psychol* 2018; 23(2): 192-213.  
[http://dx.doi.org/10.1111/lcrp.12131]
- [45] Taylor PJ, Tomblin S, Conchie SM, van der Zee S. Cross-cultural deception detection. *Detect Decept Curr Chall Cogn Approaches* 2014; p. 175.
- [46] Bereby-Meyer Y, Hayakawa S, Shalvi S, Corey JD, Costa A, Keysar B. Honesty speaks a second language. *Top Cogn Sci* 2020; 12(2): 632-43.  
[http://dx.doi.org/10.1111/tops.12360] [PMID: 29961266]
- [47] Cheng KHW, Broadhurst R. The detection of deception: The effects of first and second language on lie detection ability. *Psychiatry Psychol Law* 2005; 12(1): 107-18.  
[http://dx.doi.org/10.1375/ppl.2005.12.1.107]
- [48] Ansari S, Gupta S. Fake reviews and manipulation: Do customer reviews matter? *Res Pap* 2005.2019. [https://aisel.aisnet.org/ecis2019\\_rp/143](https://aisel.aisnet.org/ecis2019_rp/143)
- [49] Kitchenham B. Procedures for performing systematic reviews. In: *Keele UK Keele Univ.* 2004; 33: pp. (2004) 1-26.
- [50] Ott M, Choi Y, Cardie C, Hancock JT. Finding Deceptive Opinion Spam by Any Stretch of the Imagination. *ArXiv11074557 Cs ArXiv11074557 Cs* 2011. <http://arxiv.org/abs/1107.4557>