The ability to recognise emotions predicts the time-course of sarcasm processing: Evidence from eye movements

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The ability to recognise emotions predicts the time-course of sarcasm processing: Evidence from eye movements

Henri Olkoniemi[*][a], Viivi Strömberg[a], & Johanna K. Kaakinen[a,b]

a. Department of Psychology, University of Turku, Finland.

b. Turku Institute for Advanced Studies, University of Turku, Finland

*Corresponding author: E-mail address: henri.olkoniemi@utu.fi; Address: Department of Psychology, 20014 University of Turku, Finland.

Abstract

A core feature of sarcasm is that there is a discrepancy between the literal meaning of the utterance and the context in which it is presented. This means that a sarcastic statement embedded in a story introduces a break in local coherence. Previous studies have shown that sarcastic statements in written stories often elicit longer processing times than their literal counterparts, possibly reflecting the difficulty of integrating the statement into the story's context. In the present study, we examined how sarcastic statements are processed when the location of the local coherence break is manipulated by presenting the sarcastic dialogues either before or after contextual information. Sixty participants read short text paragraphs containing sarcastic or literal target statements while their eye movements were recorded. Individual differences in ability to recognise emotions and working memory capacity were measured. The results suggest that longer reading times with sarcastic statements not only reflect local inconsistency, but also attempt to resolve the meaning of the sarcastic statement. The ability to recognise emotions was reflected in eye-movement patterns, suggesting that readers who are poor at recognising emotions are slower at categorising the statement as sarcastic. Thus, they need more processing effort to resolve the sarcastic meaning.

Introduction

Sarcasm can be defined as a form of verbal irony that is typically aggressive and negative in nature (Attardo, 2000). The use of sarcasm has been shown to serve a social role, and people often use it to soften criticism and remind each other that they belong in the same group (Colston, 1997; Dews, Kaplan, & Winner, 1995; Gibbs, 2000; Gibbs & Izett, 2005). It has been suggested that people use ironic language, such as sarcasm, more in computer-mediated communication in written form than in face-to-face conversations, although there is higher risk of miscommunication (Hancock, 2004), and the consequences of the misinterpretations can be vast (e.g. Ronson, 2015).

The purpose of the present study is to examine factors that influence the ease of comprehending sarcasm in written form. Recent eye-tracking studies have demonstrated that people take longer to read sarcastic statements than literal statements (Au-Yeung, Kaakinen, Liversedge, & Benson, 2015; Filik, Leuthold, Wallington, & Page, 2014; Filik & Moxey, 2010; Kaakinen, Olkoniemi, Kinnari, & Hyönä, 2014; Olkoniemi, Ranta, & Kaakinen, 2016; Turcan & Filik, 2016). In the present study, we examined whether this slowdown is an implication of a reader resolving the sarcastic meaning (e.g., Olkoniemi et al., 2016), or whether it merely reflects a coherence break caused by the sarcastic statement, which contradicts the context in which it is presented. Finally, recent eye-tracking studies have shown that there are individual differences in how readers process sarcastic statements, and that these differences are related to the time-course of resolving the sarcastic meaning (Kaakinen et al., 2014; Olkoniemi et al., 2016). In the present study, we examined how *the ability to recognise emotions* and *working memory capacity* are related to the processing of written sarcasm.

THE ABILITY TO RECOGNISE EMOTIONS PREDICTS THE TIME-COURSE OF SARCASM PROCESSING   4

Comprehension of Sarcasm

Most theoretical accounts of sarcasm comprehension assume that when the sarcastic

utterance is unfamiliar (i.e., not typically used in sarcastic contexts) and the context in which

it occurs does not immediately elicit sarcastic interpretation, the processing of the utterance

should take longer than when the same utterance is presented in literal meaning (Gibbs, 1994;

Grice, 1975; Giora, 2003; Pexman, 2008). This slowdown reflects problems in integrating

sarcastic statements into developing text representation, which results in a reanalysis of the

statement (e.g., Grice, 1975). Recent eye-tracking studies on sarcasm have demonstrated that

sarcastic statements attract longer total reading times (e.g., Filik & Moxey, 2010), and readers

are more likely to initiate regressions during first-pass reading of sarcastic target sentences

and to look back to them from subsequent parts of text (Kaakinen et al., 2014; Olkoniemi et

al., 2016).

However, it remains unclear why readers take longer to read sarcastic statements.

Most previous eye-tracking studies have used materials in which the sarcastic statement

appears after some contextual information is provided (e.g., Au-Yeung et al., 2015; Filik et

al., 2014; Filik & Moxey, 2010; Kaakinen et al., 2014; Olkoniemi et al., 2016; Turcan &

Filik, 2016). For example, consider the following passage (obtained from Olkoniemi et al.,

2016; translation from Finnish):

> *Max and Tony are roommates. One night Tony hears strange sounds from the shower*
>
> *room, as if someone would be crying from pain. Tony rushes into the bathroom and*
>
> *finds that Max is singing Elvis using a shampoo bottle as a microphone. 'You are a*
>
> *true singer!' Tony states. Max is confused by the comment and blushes.*

In this kind of setting, the sarcastic statement (*'You are a true singer!'*) is locally

inconsistent, which may cause additional cognitive load for readers (e.g., McKoon & Ratcliff,

1992), resulting in a slowdown in processing as readers try to solve the local inconsistency.

Thus, it is difficult to disentangle the effects related to the problems caused by the local

inconsistency from the effects related to resolving the sarcastic meaning of the statement.

In the present study, we used materials in which the sarcastic meaning of the

statement becomes clear only after the reader has moved on from it. We used short stories

that contained simple dialogues between two people, such as:

(1) *Paul: 'What a great concert!'*

(2a) *Sam: 'I'm sorry I asked you to come with me.'*

 (3a) *During a concert, Paul covers his ears with his hands.*

and

(1) *Paul: 'What a great concert!'*

(2b) *Sam: 'I'm happy I asked you to come with me.'*

(3b) *The boys bought tickets to a concert.*

In both examples, the first statement (1) is a *target statement,* which, in itself, is

neutral; it could be either sarcastic or literal, depending on the context in which it is

presented. The second statement is a *validation statement,* which confirms the meaning of the

first statement either as sarcastic (2a) or literal (2b). In the present study, these dialogues are

presented either before or after the context that matches either the sarcastic (3a) or literal (3b)

interpretation of the target sentence.

The condition in which the context sentence (3a or 3b) is presented before the

dialogue resembles previous eye-tracking studies, and in this condition, sarcasm becomes

evident immediately when the reader encounters the target statement (1). Conversely, in the

context-last condition, sarcasm becomes evident after the target statement, in the validation

statements (2a and 2b). By using these types of materials, we aimed to tease apart the effects

related to local inconsistency from resolving the sarcastic meaning of the statement. If the

effects observed in previous studies are related to local inconsistency, we should observe

longer reading times for sarcastic than for literal validation statements in the context-last

condition. However, if resolving the sarcastic meaning requires reprocessing of the sarcastic

statement, we should observe that readers look back to the target statement in the context-last

condition.

Individual Differences in the Processing of Sarcasm

Recent eye-tracking studies have shown that individual differences in *working*

memory capacity (WMC) are related to how readers resolve sarcasm in written form

(Kaakinen et al, 2014; Olkoniemi et al., 2016). High WMC has been found to be related to

increased first-pass rereading of sarcastic sentences (Kaakinen et al., 2014; Olkoniemi et al.,

2016), whereas low WMC was related to an increased probability of making look-backs at

the sarcastic sentences from subsequent parts of text (Olkoniemi et al., 2016). In other words,

the time-course of resolving sarcasm seems to depend on WMC, such that high-WMC

readers detect sarcasm faster and/or resolve it earlier than low-WMC readers, who show

mainly delayed effects.

One possible explanation for these findings is that working memory is needed to keep

multiple potential interpretations (i.e., literal and sarcastic) in mind in the course of reading

(Just & Carpenter, 1992). Thus, high WMC should facilitate comprehending indirect

statements. In contrast, low-WMC readers may have trouble keeping multiple interpretations

in mind, making the interpretation process more effortful and resulting in more look-backs

(e.g., Walczyk & Taylor, 1996). Another possible explanation is that because efficient

inhibition of irrelevant material is a crucial characteristic of high WMC (e.g., Engle, 2010,

for review), high-WMC readers might be better able to suppress more salient literal

interpretations and start to process sarcastic meaning during the first-pass reading.

Conversely, low-WMC readers may have trouble suppressing or inhibiting the initial literal

interpretation of the statement, which is why they need to engage in later reprocessing to validate the sarcastic meaning (e.g., Giora, 1999; Miyake, Just, & Carpenter, 1994).

Based on previous studies, we expected that WMC should be related to the time-course of processing sarcastic statements. More specifically, we assumed that readers who have high WMC (as measured by the reading span task, Daneman & Carpenter, 1980) can process the meaning of a sarcastic target statement faster, or show more immediate reprocessing of the intended meaning than readers who have low WMC.

Sensitivity to the emotional state of the speaker plays a crucial role in sarcasm comprehension as well (Amenta, Noël, Verbanck, & Campanella, 2013; Nicholson, Whalen, & Pexman, 2013; Olkoniemi et al., 2016; Shamay-Tsoory, Tomer, & Aharon-Perez, 2005; Shany-Ur et al., 2012). For example, Nicholson et al. (2013) found that children (8- and 9-year-olds) with good empathy skills possessed better judgement of speakers' intent, as well as better comprehension of sarcasm, compared with those who have low empathy skills. In addition, Olkoniemi et al. (2016) showed that poor ability to make use of emotional information was reflected in eye-movement records. Readers who have poor ability to make use of emotional information were likely to look back from the sarcastic target sentences to earlier parts of text. Olkoniemi et al. suggested that if the reader does not have the emotional information readily available, he or she must rely on contextual information when resolving sarcastic meanings. Thus, we expected that readers' low ability to recognise emotions, as measured by the Toronto Alexithymia Scale (Bagby, Parker, & Taylor, 1994; Joukamaa et al., 2001), should be reflected in eye-movement patterns as an increased processing of sarcastic paragraphs.

Overview of Present Study

The present study explored the factors underlying comprehension of sarcasm. Of particular interest was the exact time-course of processing sarcastic statements. The location of the coherence break caused by the indirect meaning of the target statements was manipulated by presenting the contextual information either before or after each target statement. It was expected that the coherence break would cause problems in integrating the sentence with the developing memory representation, and that when the break coincides with the target statement (context before condition), it would immediately trigger longer reading times. However, when the coherence break comes after the target statement (context after condition), processing difficulty should be localised in the sentence where the coherence break becomes evident, i.e., at the validation statement. However, if resolving the meaning of the target statement requires reprocessing of the target statement itself, we should observe increased looking back to the target statement in the context after condition.

Individual differences were expected to influence the time-course of sarcasm processing. As for the WMC, we expect that high-WMC readers should show increased first-pass rereading of the statement when sarcasm becomes evident (Kaakinen et al., 2014; Olkoniemi et al., 2016). Low-WMC readers are expected to show increased look-backs (i.e., a relatively late reaction) to the sarcastic target statement in the context-first condition (as in Olkoniemi et al., 2016), as well as look-backs to either the sarcastic target statement or the validation statement in the context-last condition. As for the ability to recognise emotions, we expected that readers with poorer emotion-recognition abilities would show increased reading of the text parts that are crucial for the sarcastic interpretation. We expected that this would materialise as increased looking back to context, and/or increased first-pass rereading of the validation statement in the context-first condition, as well as increased first-pass rereading of the validation statement and/or context in the context-last condition.

Method

Participants

Sixty University of Turku (Finland) students (46 women, $M_{\mathrm{Age}} = 24.20$, $SD_{\mathrm{Age}} = 4.23$) participated in the study to fulfil a course requirement. All were native speakers of Finnish (the language studied here) and had normal or corrected-to-normal vision. All participants provided written, informed consent before the experiment.

Apparatus

Eye movements were recorded using a head-mounted EyeLink II (11 participants) or a desktop-mounted EyeLink 1000 eye-tracker system (49 participants) (SR Research Ltd., Ontario, Canada). The eye-movement registration was done monocularly, typically for the right eye. Sampling frequency was 500 Hz for EyeLink II and 1,000 Hz for EyeLink 1000. The stimuli were presented on a 21-inch CRT screen with a screen resolution of $1,024 \times 768$ pixels, with a 100 Hz refresh rate. Participants were seated 70 cm from the screen, and a chin rest was used with EyeLink 1000 to stabilise the head.

Materials

Text materials. Participants read a total of 60 short paragraphs. Forty of the paragraphs included sarcastic or literal statements (20 sarcastic and 20 literal). In addition, there were 20 filler items that included lies or literal statements (10 lies and 10 literal). The filler items purposefully were designed to include statements that required the reader to infer the intent of the speakers. Text paragraphs included one or two context sentences and a dialogue with two lines (see example in Table 1). In the dialogue, the first statement was a target statement, which was sarcastic or literal. The second statement was a validation statement that validated the meaning of the first statement. The context sentences (context)

were presented either before (context-first condition) or after the dialogue (context-last condition).

[Table 1 near here]

There were four versions of each paragraph: a literal and a sarcastic version of each paragraph, plus a version in which the context was presented first and another in which the context was presented last. Also, the filler paragraphs were constructed similarly. Each participant saw only one version of a paragraph. The paragraph version and presentation order of the texts were pseudo-randomised across participants.

Participants read the 60 stories on a computer screen (font: Courier New; font size: 15; line height: 3), while their eye movements were recorded. Their understanding of the target statement and their memory for text were checked after predefined twenty paragraphs by presenting two questions: The first was an open question tapping into the meaning of the target statement (e.g., 'In your opinion, what did Paul mean?'). The other question required a yes-or-no response related to text memory (e.g., 'Did Paul cover his ears during the concert?'). Participants responded to the first question by typing their answers in a text box on the screen. As for the text-memory question, participants responded by pushing designated 'Yes' and 'No' buttons from the keyboard. As the presentation order of the texts was randomized, the questions appeared at random intervals.

For both types of questions, a correct answer was rewarded with one point, and the percentage of correct answers was computed. The reliability of the scoring of the inference questions was checked by selecting one-third of the answers that were scored by two independent scorers. The agreement between raters was good (96.90%; $\kappa = .84$, $p < .001$).

Separate rating studies were conducted to test: 1)how familiar the target statements were as sarcastic in comparison to literal meaning; and 2) how sarcastic statements were experienced compared with literal statements. The familiarity of target statements as sarcastic or as literal phrases was examined in a rating study. A survey tool was used to collect the data (Webropol, www.webropol.com) from 25 native Finnish speakers between the ages of 27-58 (13 women, M_{Age} = 34.12 years, SD_{Age} = 7.20). None of the participants took part in the actual experiment. Participants read all target statements without text context one at a time and evaluated, on a scale of 0 (never) to 10 (very often), how often they previously had seen or heard the statements in literal or sarcastic use. Before the evaluation, short definitions of the text types were given to the participants to read. The analysis showed that target statements were less familiar as sarcastic (M = 4.71, SD = 2.28, *range*: 0.33 –8.25) than literal (M = 5.40, SD = 2.17, *range*: 0.58–8.75), $t(24)$ = 2.51, p = .019, d' = -0.23.

In another rating study, **t**he text materials were tested for: 1) inferred meaning of the target statement and how 2) funny, 3) insulting and 4) natural the target statement was in the paragraph context. Fifty-two native Finnish speakers ages 19–52 (42 women, M_{Age} = 26.35, SD_{Age} = 7.66) participated in the study to fulfil a course requirement. None of the participants took part in the actual experiment. Participants were tested in groups of 3–10 participants in a computer classroom. Participants saw the paragraphs one at a time and were allowed to read the paragraphs and answer the questions at their own pace. The experimental session lasted for about 45 minutes.

When answering questions about the meaning of the target statements, participants chose from three options the one that he or she thought matched the target statement presented. The three options were literal, untruthful, or sarcastic interpretations (e.g., 'Paul likes the concert.'/'Paul tries to hide that he doesn't like the concert.'/'Paul doesn't like the concert and criticises Sam's choice.'). A correct answer was rewarded with one point, and the

percentage of correct answers was computed. In addition, participants evaluated how

insulting, funny and natural each target statement was on a scale of 0 to 10 (0 = not

funny/insulting/natural at all; 10 = very funny/insulting/natural). Descriptive statistics of the

ratings are presented in Table 2. Possible effects of context (before vs. after) and text type

(literal vs. sarcastic) were evaluated using 2×2 repeated-measure ANOVAs.

[Table 2 near here]

Analysis of the *funniness* of the target statements did not show an interaction between

text type and context, $F(1,204) = 0.20$, $p = .658$, $\eta_p^2<.01$, or a main effect of the context,

$F(1,204) = 0.07$, $p = .066$, $\eta_p^2<.01$. The funniness ratings differed between text types,

$F(1,204) = 69.08$, $p < .01$, $\eta_p^2 = .25$, indicating that sarcastic target statements were evaluated

as being funnier than their literal counterparts. The analysis on *level of insult* of the target

statement did not show an interaction between text type and context conditions, $F(1,204) =$

0.01, $p = .914$, $\eta_p^2<.01$, or a main effect of the context condition, $F(1,204) = 0.262$, $p = .609$,

$\eta_p^2 < .01$. The level of insult ratings of the target statements differed between text types,

$F(1,204) = 220.69$, $p < .001$, $\eta_p^2 = .52$, indicating that target statements were evaluated as

more insulting when presented in sarcastic than in literal meaning. The analysis on *how*

natural the target statement was did not show an interaction between text type and context

conditions, $F(1,204) = 0.21$, $p = .650$, $\eta_p^2<.01$, or a main effect of the context condition,

$F(1,204) = 0.50$, $p = .481$, $\eta_p^2<.01$. However, the naturality ratings of the target statements

differed between text types, $F(1,204) = 50.58$, $p < .001$, $\eta_p^2=.20$,indicating that sarcastic target

statements were evaluated as less natural in the story context than literal statements. As for

the *inferred meaning* of the target statement questions, the analysis did not show an

interaction between text type and context conditions, $F(1,204) < 0.01$, $p = .949$, $\eta_p^2<.01$, or a

main effect of the context condition, $F(1,204) = 1.16$, $p = .282$, $\eta_p^2 = .01$. However, there was

a difference between text types in inferring the correct meaning, $F(1,204) = 17.01$, $p < .001$,

$\eta_p^2 = .08$, indicating that readers were less adept at responding to the inference questions after

sarcastic statements compared with literal paragraphs.

In sum, target statements were rated as funnier and more insulting when presented in

sarcastic in comparison to literalmeaning, and were harder to comprehend. Target statements

were evaluated as less natural in sarcastic story context than in literal context. However, the

naturality scores overall were quite high, and the differences in perceived naturality simply

may reflect the nature of the sarcastic statements. Considering thatthe target statements were

overall more familiar as literal than sarcastic, they were incoherent within thesarcastic story

contexts. Finally, the context manipulation had no effect on how the paragraphs were

evaluated or comprehended.

Ability to recognise emotions. The ability to recognise emotions was measured using

the Finnish version of the 20-item Toronto Alexithymia Scale (Bagby et al., 1994; Joukamaa

et al., 2001). TAS is a paper-and-pencil self-report scale that includes short claims, e.g., 'I am

often confused about what emotion I am feeling'. Participants answered the items on a five-

point Likert scale from 1 (strongly disagree) to 5 (strongly agree). The scale was scored by

summing up all responses. The scores vary between 20 and 100 points, with higher scores

indicating poorer ability to recognise emotions (i.e., higher alexithymia). The internal

reliability (Cronbach's alpha) of the TAS total score was α = .81. The average TAS score in

the experiment was 41.18 ($SD = 8.87$, range: 24–63).

Working memory capacity. The reading span test was used to measure verbal WMC

(Daneman & Carpenter, 1980; Kaakinen & Hyönä, 2007). Participants read aloud sets of

unrelated sentences presented on a computer screen. After every set, they were asked to recall

the last word of each sentence in the set. The test started with sets of two sentences. The set

size increased as long as the participant was able to recall the final words of the sentences.

Each set size was repeated three times. The test ended when the participant failed to recall the

final words of a sentence of a particular set size for its three repetitions. The test was

preceded by a practice session with three sets of two sentences. The test was scored for the

total number of correctly recalled final words, with test scores varying between 0 and 81

points. The average WMC score in the experiment was 25.90 ($SD = 12.18$, range: 9–68).

Procedure

Participants were tested individually. Upon arrival, participants were informed that

the experiment assessed reading. Participants also signed a consent form, and the specific

nature of the experiment was explained to participants when the experiment was over.

Before the reading task, the eye-tracking system was introduced to each participant,

and the experimental procedure was explained. The eye-tracker then was set up and

calibrated using a nine-point calibration screen. Participants were instructed to read each

paragraph at their own pace. Each paragraph was presented on one screen. Participants were

told to press the Enter key on the keyboard when they finished reading the paragraph. After

20 of the 60 paragraphs, two questions were presented one at a time. After the participant

answered the second question, the next paragraph was presented. The reading task was

followed by the reading span test, then participants filled out the TAS. Each experimental

session lasted for about 90 minutes.

Results

Data Analysis

Fixations shorter than 50 ms were either merged with a nearby fixation (if the distance

between the fixations was < 1°) or removed from the data. Sentence-level measures for target

statement, validation statement and context were computed from the eye-movement data

(Hyönä, Lorch, & Rinck, 2003). *First-pass reading time* is the summed duration of the fixations falling within the target region until the reader moved his or her eyes to fixate on another region. Note that we prefer to use the term *first-pass reading time* here rather than *gaze duration* because the regions of interest are not single words. *Regression path duration* is the summed duration of the fixations that occurred from the first fixation in a region until the participant moves his or her eyes beyond that region to the right. Therefore, regression path duration included all the fixations in a region and any regressive fixations on words in the previous parts of the text until a fixation is made to the right of the region. Moreover, we estimated *first-pass rereading time* by calculating regression path duration from the final word of the target statement. *Look-back fixation time* is the summed duration of fixations returning to the sentence from other parts of the text after the first-pass reading. From the look-back fixations, we computed the *probability to initiate a look-back* (binomial measure), and the summed fixation time on the condition that rereading was made.

For the target statements, all the reading time and probability measures described above were analysed. First-pass reading time, regression path duration and probability to initiate a look-back were analysed for validation statements and context. However, first-pass reading time and regression-path duration were analysed for target statements and context sentences only when they were presented last. When they were presented first, readers were unaware of the nature of the target statement and had no place to return. In addition, probability to initiate a look-back to validation statements and context was analysed only when they were presented as the last text region (context-first condition for context and context-last condition for validation statements). Validation statements ($M_{characters}$ = 30.78, *range:* 11–51) and contexts ($M_{characters}$ = 106.99, *range:* 40–199) varied in length, and consequently, length was controlled for in the analyses of fixation durations by using per-

character reading times (e.g., Ferreira & Clifton, 1986; Frazier & Rayner, 1982).[1] As for

look-backs to validation statements and context, only probabilities were analysed. The

reading-time measures were skewed; thus, they were logarithmically transformed before the

analyses.

Data were analysed with LMMs specifying participants and items as crossed random

effects (Baayen, Davidson, & Bates, 2008) using the lme4 package (Bates, Mächler, Bolker,

& Walker, 2015) in the R statistical software (Version 3.4.3; R Core Team, 2017). The

models were estimated using Maximum Likelihood estimation. Separate models were built

for each eye-movement measure for the different text regions (target statement, validation

statement and the context), and for the context-first and -last conditions. The influence of text

type (literal vs. sarcasm) was tested by fitting models with a sum-coded, fixed-effect variable.

The individual differences in variables were added to the models as centred, fixed-effect

variables; correlations between the measures were low, $r = -.19$, $p = .149$, 95% CI [-.42, .07].

The WMC score was distributed non-linearly and consequently divided into low and high

groups using a median split. To examine the potential effects of presentation order on the

observed effects (e.g., Olkoniemi et al., 2016), trial order was added as a fixed effect to each

model (first half of the experiment = -1, end half of the experiment = 1). Moreover, as two

different eye-tracking systems were used in the experiment, the eye tracker was added to the

models as a sum-coded, fixed-effect variable.

Model fitting was performed in a step-wise fashion, starting with the most complex

model, including all individual difference measures, text type, trial order, and their

---

[1] The use of millisecond-per-character reading times has been criticised because there is no linear

trend between fixation times and number of characters in text regions shorter than 10 characters

(Trueswell, Tanenhaus, & Garnsey, 1994). However, larger text regions (> 10 characters), as were

the case in the present experiment, are not likely to be distorted (Trueswell et al., 1994).

interactions as fixed effects. At this point, only participants and texts were fitted as random effects. The fixed effect associated with the smallest *t*-value was removed from the model, starting from the interaction terms, and the reduced model was compared with the former using *anova* function in the *lme4* package (Bates et al., 2015) to compare the log-likelihood of the two models. Fixed effects were removed one at a time (except for eye-tracker, which always was retained in the model as a control variable) until nothing else could be removed without significantly reducing the fit of the model. Finally, full random structure was fitted to the model (Barr, Levy, Scheepers, & Tily, 2013), and fixed and random effects were removed from the model if further changes did not significantly reduce the fit of the data. An exception was the eye-tracking system, which was a control variable and was not fitted to the random structure (e.g., Barr et al., 2013). If the model failed to converge after fitting the full random structure, the random structure of the model was trimmed top-down, starting with removing correlations between factors.

The exact degrees of freedom are difficult to determine for the *t*-statistics estimated by LMMs, leading to the problem of determining exact *p*-values (Baayen et al., 2008). Consequently, degrees of freedom, or *p*-values, are not reported; statistical significance at the .05 level is indicated by values of the $|t$ and $z| > 1.96$. For the sake of brevity, only significant, or near-significant, effects involving text type are reported. Significant main effects are reported in text. Interactions were examined by computing the estimates of text type at low (1 *SD* below the mean) and high (1 *SD* above the mean) levels of the individual-differences variable, and the estimates and their 95% CIs are illustrated in figures. Final models, as well as the datasets and the R script, are reported in the Online Appendix available in the Open Science Framework (https://osf.io/4syah/).

Correct answers to questions presented after the paragraphs were analysed with a paired-samples t-test, as the number of observations was too low to fit LMMs.

THE ABILITY TO RECOGNISE EMOTIONS PREDICTS THE TIME-COURSE OF SARCASM PROCESSING   18

Reading Times in the Context-First Condition

Observed means and standard deviations of the different eye-movement measures for

the context-first and -last conditions are presented in Table 3. The results are presented

separately for each text part (context, target statement and validation statement) in the order

they appeared in the context-first condition.

[Table 3 near here]

Reading of the Context

The analysis of the *probability to initiate a look-back to context* did not show effects of

text type.

Reading of Target Statement

The analysis of the *first-pass reading times on the target statement* did not show effects

of text type. However, the analysis of the *first-pass rereading time* on the target statement

showed a main effect of text type, indicating that sarcastic target statements attracted longer

first-pass rereading times than their literal counterparts, $b = 56$ms, 95% CI [13.80, 117.68].

The analysis of the *regression path duration on the target statement* showed an interaction

between text type and trial order. Readers showed longer regression path duration for

sarcastic compared with literal target statements at the beginning of the experiment, but the

difference between text types wore off toward the end of the experimental session (see Figure

1). The analysis of the *probability to initiate a look-back to target statement* did not show

effects of text type. Finally, the analysis of *the look-back time to target statement* revealed a

main effect of the text type, indicating that if a look-back to the target statement was made,

sarcastic target statements attracted longer look-backs than literal statements, $b = 89$ms, 95%

CI [12.70, 204.55].

[Figure 1 near here]

Reading of Validation Statement

The analysis of the *first-pass reading time on the validation statement* showed an

interaction between text type and TAS. The result indicates that only readers with a relatively

high TAS score (i.e., higher alexithymia traits and poorer ability to recognise emotions)

showed a sarcasm effect (i.e., longer reading times on sarcastic paragraphs, compared with

literal ones; see Figure 2). The analysis of the *regression path duration on the validation*

statement did not show effects of text type.

[Figure 2 near here]

Reading Times in the Context-Last Condition

The results are presented separately for each text part (target statement, validation

statement and context) in the order that they appeared in context-last condition.

Reading of Target Statement

The analysis of the *probability to initiate a look-back to the target statement* did not

show effects of text type. In addition, the analysis of *the look-back time-to-target statement*

showed a weak effect of text type, indicating that sarcastic target statements attracted longer

look-backs than literal statements, $b = 95$ms, 95% CI [-1.65, 240.32].

Reading of Validation Statement

The analysis of *first-pass reading time on the validation statement* did not show

effects of text type. The analysis of the *regression-path duration on the validation statement*

revealed a main effect of the text type, indicating that validation statements from the sarcastic

paragraphs attracted longer regression-path duration than literal statements, b =

4ms/character, 95% CI [0.06, 5.54]. Finally, the analysis of the *probability to initiate a look-*

back to the validation statement did not show effects of text type.

Reading of Context

The analysis of the *first-pass reading time on the context* did not show effects of text

type. The analysis of the *regression-path duration on the context* revealed a main effect of the

text type, indicating that context of the sarcastic paragraphs attracted longer regression-path

duration times than their literal counterparts, b = 3ms/character, 95% CI [0.27, 6.89].

Text Memory and Inference Questions

Readers were better at responding to the *text memory questions* after sarcastic (M =

95.67%, SD = 9.09) more than literal (M = 89.67%, SD = 13.01) paragraphs, $t(59)$ = 3.34, p =

.001, d = .43. As for *inference question*s, readers were poorer at responding to inference

questions after sarcastic (M = 80.67%, SD = 22.39) more than literal (M = 94.00%, SD =

12.38) paragraphs, $t(59)$ = -4.45, p < .001, d = -.57.

In sum, although the overall accuracy in responses to the questions was relatively

high, readers were better at responding to the text-memory questions for the sarcastic than for

literal paragraphs, and had more problems responding to inference questions for the sarcastic

compared with literal target sentences.

Discussion

In the present study, we examined the moment-to-moment processing of sarcastic and

literal statements embedded in story contexts. The location of the coherence break introduced

by the sarcastic statement was manipulated by presenting the sarcastic dialogue either before

or after the contextual information. Moreover, we were interested in individual differences in

the processing of sarcasm.

Role of context in processing sarcasm

When story context preceded the target statement, readers tended to do more

immediate rereading and longer returns to the sarcastic target statement. Regression path

duration also was longer for sarcastic than literal statements, indicating that readers reread the

context, as well as the target statement, before moving on, especially at the beginning of the

experimental session. These results are in line with previous results showing that processing

of sarcasm takes more time than processing literal statements, and that the effects are mostly

located on the target statement itself (Au-Yeung et al., 2015; Filik et al., 2014; Filik &

Moxey, 2010; Kaakinen et al., 2014; Olkoniemi et al., 2016; Turcan & Filik, 2016). In other

words, when there is a story context after which a sarcastic statement is presented, the

sarcasm is hard to integrate into the developing text representation. Thus, there is a need to

reassign a new, sarcastic meaning to the statement (e.g., Gibbs, 1994; Giora, 2003; Grice,

1975).

In the condition in which the contextual information was presented after the target

statement, the effects on the target statement itself were rather weak. Sarcastic target

statements attracted only slightly longer look-backs than literal statements. Instead, for the

validation statement, regression path duration was longer after a sarcastic than a literal target

statement. The result indicates that readers did rereading of validation and target statements

immediately when the sarcasm became evident. Moreover, regression path duration from

context was longer for sarcastic than for literal texts, suggesting that context also needed to

be integrated with the dialogue containing the sarcasm to form a coherent text representation.

With respect to the processing of the statement in which sarcasm became evident (target statement in the context-first condition and validation statement in the context-last condition), the results showed that readers did increased reprocessing of the statement that represented the inconsistency, and also returned to the previous text part from it. The result suggests that readers reacted to the local inconsistency and possibly tried to integrate the source of inconsistency with the existing contextual information. However, the target statements and the validation statements differed in respect to look-backs, in that only sarcastic target statements were looked back more, compared with their literal counterparts. This result suggests that the processing of sarcastic meaning is at least, to some extent, localised to the sarcastic statement, supporting the idea that resolving the meaning of sarcasm requires re-evaluating the meaning of the statement (Grice, 1975). Furthermore, the results suggest that look-backs probably reflect integrating the meaning of the sarcastic statement with the story context. The result is in line with eye-tracking studies suggesting that look-backs reflect a conscious, strategic effort to build a comprehensive mental representation of the text content (Hyönä, Lorch, & Kaakinen, 2002; Hyönä & Nurminen, 2006). However, as the validation statements were not exactly the same in the sarcastic and literal stories ('I'm sorry I asked you to come with me' vs. 'I'm happy I asked you to come with me'), this interpretation should be considered with some caution.

Some of the effects related to the processing of sarcasm changed during the experimental session, replicating previous findings (Olkoniemi et al., 2016). It is possible that sarcastic statements encountered during the experiment created a global context, in which sarcastic statements increasingly were more likely to appear, affecting processing of the paragraphs. It is noteworthy that this happened even though our experimental materials included non-sarcastic stories containing statements that did not directly fit into the context (i.e., lies), requiring the reader to infer their meaning. The result is in line with theoretical

views that assume a role of context in biasing the interpretation toward the non-salient, non-literal meaning (Gibbs, 1994; Giora, 2003; Pexman, 2008).

In addition, the results showed that readers were poorer in responding to the inference questions after sarcastic than after literal paragraphs, replicating previous findings (Au-Yeung et al., 2015; Kaakinen et al., 2014; Olkoniemi et al., 2016). Despite the extra processing effort that readers invest in reading the sarcastic statements, they do not always understand the sarcastic meaning in them. Maybe because of this extra processing effort, readers also were more accurate in answering text-memory questions related to sarcastic than literal texts. In other words, extra processing effort related to the processing of sarcasm helps readers better recall the text content. The results of the norming study showed that the comprehension of the sarcastic statements was unaffected by the contextual manipulation, suggesting that the location of the contextual information affects processing, but not the comprehension of sarcastic statements. However, it should be noted that in the norming study, comprehension of statements was measured with multiple-choice questions, whereas in the eye-movement experiment, participants were free to provide their own answers. Thus, one should be cautious when comparing results.

Individual Differences in Processing Sarcasm

Individual differences in the ability to recognise emotions, as measured by TAS, were related to the processing of sarcasm. In the context-first condition, readers scoring relatively high in TAS (i.e., poor ability to recognise emotions) showed increased processing of validation statements of the sarcastic paragraphs during first-pass reading. These findings support the hypothesis that poor ability to recognise emotions is related to greater confusion when encountering sarcastic statements and that this effect spills over into validation statements. As suggested by Olkoniemi et al. (2016), the emotional component in sarcasm serves as a marker pointing toward the sarcastic interpretation, helping the reader infer the

sarcastic meaning. Those having difficulties noticing or interpreting the emotional marker in sarcasm need more contextual information to form the correct inference.

However, the effect was not seen in the context-last condition. The results of the norming study showed that the target statements used in the present experiment were perceived to be emotionally laden (i.e., more insulting and funnier) when presented in sarcastic than in literal meaning, and that the emotional component did not differ between context-first and -last conditions. It might be that when the context is presented after the sarcastic statement, it is easier to form an interpretation of the statement and easier to process for those who would otherwise need extra processing of the contextual information. This interpretation is in line with the results reported by Ackerman (1982), who showed that correct sarcastic interpretation was more difficult to make when the context was presented first than when it was presented last. Ackerman suggested that when the context is presented first, integrating the statement meaning with the context is more difficult than when the sarcastic statement precedes the context (Ackerman, 1982; cf. Grice, 1975). When the context precedes the sarcastic statement, readers already have started to build a literal text representation in their minds; thus, the reader expects a literal statement (e.g., Gibbs, 1994; Giora, 2003). This causes the extra processing in the context-first condition for sarcastic statements. However, when the context comes after the statement, there is no text representation that the statement should be integrated with, and the reader is more open to different interpretations: The statement might be a literal comment or sarcasm, which would become evident only later, and there is less need for extra processing of the target statement. This notion is supported by our data because the effects related to the sarcasm in the context-last condition were relatively small.

Finally, we failed to replicate previous findings (Kaakinen et al., 2014; Olkoniemi et al., 2016) showing that high WMC is related to the increased rereading of the sarcastic target

statement. The result might be related to the text materials that were very short in the present

experiment (3–4 sentences; in previous studies 5-14 sentences were used). Shorter paragraphs

do not strain working memory, which is likely to diminish the effects related to working

memory capacity.

Conclusions

The results suggest that even though the comprehension slowdown typically observed

with sarcastic statements in text is partly related to resolving a coherence break, there also is

a component related to resolving the sarcastic meaning. This is reflected as increased look-

backs to the sarcastic target statement regardless of when sarcasm becomes evident. The

results also suggest that forming a sarcastic interpretation is somewhat easier when the

context is presented after the sarcastic statement, at least for those who are poorer at

recognising emotions. In the context-last condition, readers have not started to build a literal

text representation before the statement, but rather start to build a sarcastic interpretation as

early as possible. This especially aids readers who have poorer abilities to recognise emotions

and, thus, may not be able to recognise the emotional cues in sarcasm and may need to form

an inference based on other cues provided in the text.

The present results are in line with theoretical views that assume that the text context

may provide support for either literal or sarcastic interpretation of a statement (Gibbs, 1994;

Giora, 2003; Pexman, 2008), and that reader characteristics moderate how much reader

makes use of contextual information. Furthermore, the results show that readers who are

better able to recognise emotions can use the emotional marker (i.e., the emotional

discrepancy between what the protagonist says and the context) as a cue in interpreting the

statement, lending support to the parallel constraint-satisfaction framework (Pexman, 2008).

The framework states that sarcasm comprehension depends on complex social, emotional and

cognitive inferences, as well as on an individual's ability to rapidly coordinate the

information needed for interpretation formation (Pexman, 2008).

Acknowledgements

References

Ackerman, B. P. (1982). Contextual integration and statement interpretation: The ability of children and adults to interpret sarcastic statements. *Child Development, 53*, 1075–1083. doi:10.2307/1129149

Amenta, S., Noël, X., Verbanck, P., & Campanella, S. (2013). Decoding emotional components in complex communicative situations (irony) and its relation to empathic abilities in male chronic alcoholics: An issue for treatment. *Alcoholism: Clinical & Experimental Research, 37*, 339–347. doi:10.1111/j.1530-0277.2012.01909.x

Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of Pragmatics, 32*, 793–826. doi:10.1016/S0378-2166(99)00070-3

Au-Yeung, S. K., Kaakinen, J. K., Liversedge, S. P., & Benson, V. (2015). Processing of written irony in autism spectrum disorder: An eye-movement study. *Autism Research, 8*, 749–760. doi:10.1002/aur.1490

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. doi:10.1016/j.jml.2007.12.005

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale-I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*, 23–32. doi:10.1016/0022-3999(94)90005-1

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. doi:10.18637/jss.v067.i01

Colson, H. L. (1997). Salting a wound or sugaring a pill: The pragmatic functions of ironic criticism. *Discourse Processes, 23*, 24–53. doi:10.1080/01638539709544980

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450–466. doi:10.1016/S0022-5371(80)90312-6

Dews, S., Kaplan, J., & Winner, E. (1995). Why not say it directly? The social function of irony. *Discourse Processes, 19*, 347–367. doi:10.1080/01638539509544922

Engle, R. W. (2010). Role of working-memory capacity in cognitive control. *Current Anthropology, 51*, 17–26. doi:10.1086/650572

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368. doi:10.1016/0749-596X(86)90006-9

Filik, R., Leuthold, H., Wallington, K., & Page, J. (2014). Testing theories of irony processing using eye-tracking and ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 811–828. doi:10.1037/a0035658

Filik, R., & Moxey, L. M. (2010). The on-line processing of written irony. *Cognition, 116,* 421–436. doi:10.1016/j.cognition.2010.06.005

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210. doi:10.1016/0010-0285(82)90008-1

Gibbs, R. W. (1986). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General, 115*, 3–15. doi:10.1037/0096-3445.115.1.3Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding.* Cambridge: Cambridge University Press.

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol, 15*, 5–27. doi:10.1080/10926488.2000.9678862

Gibbs, R. W., & Izett, C. D. (2005). Irony as persuasive communication. In H. L. Colston & A. N. Katz (Eds.), *Figurative language comprehension: Social and cultural influence* (pp. 131–151). Mahwah, New Jersey: Lawrence Erlbaum Associates Inc.

Giora, R. (1995). On irony and negation. *Discourse Processes, 19*, 239–264. doi:10.1080/01638539509544916

Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics, 31*, 919–929. doi:10.1016/S0378-2166(98)00100-3

Giora, R. (2003). *On our mind: Salience, context, and figurative language.* New York: Oxford University Press.

Grice, H. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.

Hancock, J. T. (2004). Verbal irony use in face-to-face and computer-mediated conversations. *Journal of Language and Social Psychology, 23*, 447–463. doi:10.1177/0261927X04269587

Hyönä, J., Lorch, R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology, 94*, 44–55. doi:10.1037//0022-0663.94.1.44

Hyönä, J., Lorch, R. F., & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam: Elsevier Science. doi:10.1016/B978-044451020-4/50018-9

Hyönä, J., & Nurminen, A. M. (2006). Do adult readers know how they read? Evidence from eye movement patterns and verbal reports. *British Journal of Psychology, 97*, 31–50. doi:10.1348/000712605X53678

Joukamaa, M., Miettunen, J., Kokkonen, P., Koskinen, M., Julkunen, J., Kauhanen, J., . . . Järvelin, M.-R. (2001). Psychometric properties of the Finnish 20-item Toronto Alexithymia Scale. *Nordic Journal of Psychiatry, 55*, 123–127. doi:1080/08039480116694

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122–149. doi:10.1037/0033-295X.99.1.122

Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory, 15*, 634–645. doi:10.1080/09658210701457096

Kaakinen, J. K., Olkoniemi, H., Kinnari, T., & Hyönä, J. (2014). Processing of written irony: An eye movement study. *Discourse Processes, 51*, 287–311. doi:10.1080/0163853X.2013.870024

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*, 440–466. doi:10.1037/0033-295X.99.3.440

Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language, 33*, 175–202. doi:10.1006/jmla.1994.1009

Nicholson, A., Whalen, J. M., & Pexman, P. M. (2013). Children's processing of emotion in ironic language. *Frontiers in Psychology, 4*, 1–10. doi:10.3389/fpsyg.2013.00691

Olkoniemi, H., Ranta, H., & Kaakinen, J. K. (2016). Individual differences in the processing of written sarcasm and metaphor: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 433–450. doi:10.1037/xlm0000176

Pexman, P. M. (2008). It's fascinating research: The cognition of verbal irony. *Current Directions in Psychological Science, 17*, 286–290. doi:10.1111/j.1467-8721.2008.00591.x

R Core Team (2017). R: A language and environment for statistical computing. Retrieved from http://www.R-project.org.

Ronson, J. (2015) How one stupid tweet ruined Justine Saccos life. Retrevied from: https://www.nytimes.com/2015/02/15/magazine/how-one-stupid-tweet-ruined-justine-saccos-life.html?_r=0

Shamay-Tsoory, S. G., Tomer, R., & Aharon-Peretz, J. (2005). The neuroanatomical basis of understanding sarcasm and its relationship to social cognition. *Neuropsychology, 19*, 288–300. doi:10.1037/0894-4105.19.3.288

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285–318. doi:10.1006/jmla.1994.1014

Turcan, A., & Filik, R. (2016). An eye-tracking investigation of written sarcasm comprehension: The role of familiarity and context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 1867–1893. doi:10.1037/xlm0000285

Walczyk, J. J., & Taylor, R. W. (1996). How do the efficiencies of reading subcomponents relate to looking back in text? *Journal of Educational Psychology, 88*, 537–545. doi:10.1037/0022-0663.88.3.537

Figure Captions

Figure 1. Model estimates for the regression-path duration on the target statement. The y-axis represents the sarcasm effect, which is the difference in reading times between sarcastic and literal texts. The model means and confidence intervals are back-transformed from log values. Error bars represent 95% CI.

Figure 2. Model estimates for the first-pass reading time on the validation statement. The y-axis represents the sarcasm effect, which is the difference in the per-character reading times between sarcastic and literal paragraphs. For illustration purposes, the TAS score is divided into high and low (± 1 SD), with a high TAS score indicating higher alexithymia traits and poorer ability to recognise emotions. The model means and confidence intervals are back-transformed from log values. Error bars represent 95% CI.
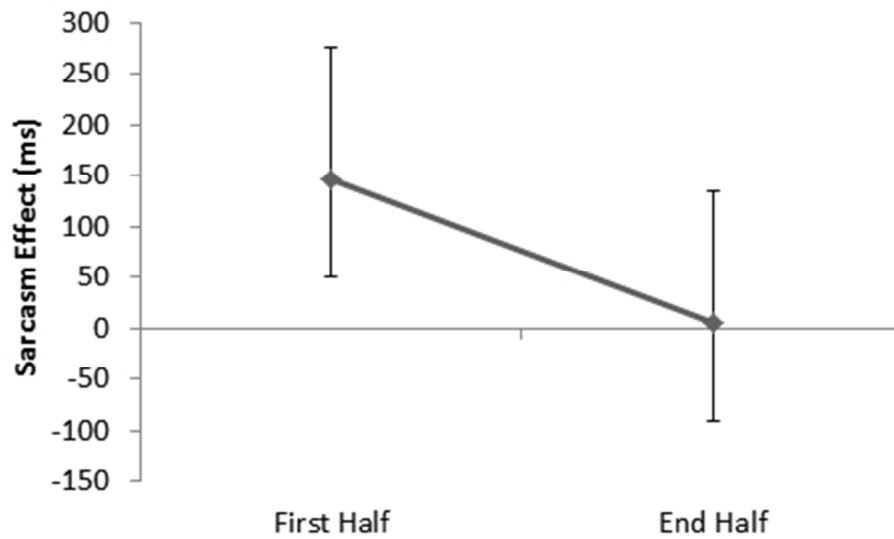
Figure 1. Model estimates for the regression-path duration on the target statement. The y-axis represents the sarcasm effect, which is the difference in reading times between sarcastic and literal texts. The model means and confidence intervals are back-transformed from log values. Error bars represent 95% CI.

40x24mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 2. Model estimates for the first-pass reading time on the validation statement. The y-axis represents the sarcasm effect, which is the difference in the per-character reading times between sarcastic and literal paragraphs. For illustration purposes, the TAS score is divided into high and low (± 1 SD), with a high TAS score indicating higher alexithymia traits and poorer ability to recognise emotions. The model means and confidence intervals are back-transformed from log values. Error bars represent 95% CI.
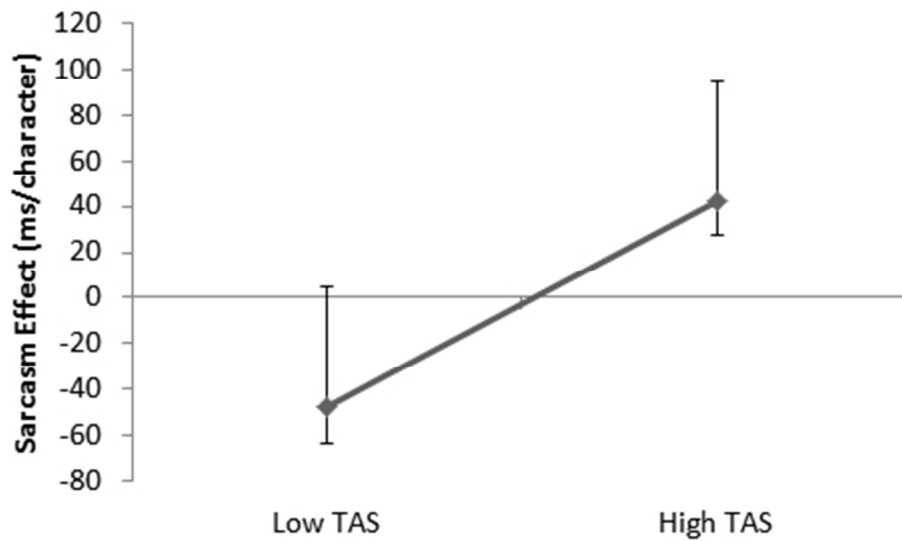
40x24mm (300 x 300 DPI)

Table 1

Examples of Experimental Paragraphs (translation from Finnish).

		Context-First Condition
Region	*Text type*	*Text*
Context	*Literal*	The boys bought tickets to a concert.
	Sarcastic	During a concert, Peter covers his ears with his hands.
Target statement		Paul: 'What a great concert!'
Validation statement	*Literal*	Sam: 'I'm happy I asked you to come with me.'
	Sarcastic	Sam: 'I'm sorry I asked you to come with me.'

		Context-Last Condition
Region	*Text type*	*Text*
Target statement		Paul: 'What a great concert!'
Validation statement	*Literal*	Sam: 'I'm happy I asked you to come with me.'
	Sarcastic	Sam: 'I'm sorry I asked you to come with me.'
Context	*Literal*	The boys bought tickets to a concert.
	Sarcastic	During a concert, Peter covers his ears with his hands.

Note. There were two versions of each paragraph (literal and sarcastic); each participant read only one of the versions, which were counterbalanced across participants. English translations of the stimuli are available upon request from the first author.

Quarterly Journal of Experimental Psychology

Table 2
Descriptive Statistics of the Paragraph Ratings

		Context First		Context Last	
Measure	*Text Type*	*M*	*SD*	*M*	*SD*
Level of funniness	Literal	2.12	1.70	2.16	1.57
	Sarcasm	4.11	1.64	3.95	1.66
Level of insult	Literal	0.50	0.68	0.58	0.57
	Sarcasm	3.52	1.99	3.64	1.99
Naturality	Literal	8.31	1.10	8.10	1.20
	Sarcasm	6.92	1.52	6.87	1.44
Correct inference (%)	Literal	93.27	9.01	94.81	7.79
	Sarcasm	86.92	11.97	88.65	13.87

Table 3

Descriptive Statistics of the Reading Time Measures in Context-First and -Last Conditions

			Text type			
			Literal		Sarcasm	
Context	Text region	Measure	*M*	*SD*	*M*	*SD*
First	Context	L-B prob	0.73	0.44	0.68	0.47
	Target statement	F-PR time	722	454	737	505
		F-PRR time	559	624	636	651
		RPD	1,140	779	1,212	790
		L-B time	961	883	1,115	1,024
		L-B prob	0.71	0.45	0.75	0.44
	Validation statement	F-PR time	29	17	29	17
		RPD	116	113	107	96
Last	Target statement	L-B time	1,463	1,209	1,612	1,346
		L-B prob	0.86	0.35	0.89	0.32
	Validation statement	F-PR time	32	21	33	20
		RPD	50	37	54	33
		L-B prob	0.77	0.42	0.76	0.42
	Context	F-PRR	29	13	29	15
		RPD	53	31	57	34

Note. Reading-time measures for target statement are in ms. Reading time measures for validation statement and context are ms/character reading times. F-PR time = First-Pass Reading time, F-PRR time = First-Pass Rereading time, RPD = Regression Path Duration, L-B time = Look-Back time, L-B prob = Probability to Look-Back