The role of defaultness and personality factors in sarcasm interpretation: Evidence from eye-tracking during reading

Ruth Filik[a], Hannah Howman[a], Christina Ralph-Nearman[a], & Rachel Giora[b]

[a]School of Psychology, University of Nottingham, UK

[b]Department of Linguistics, Tel Aviv University, Israel

Correspondence to:

Ruth Filik

School of Psychology

The University of Nottingham

University Park

Nottingham NG7 2RD

UK

Tel: +44 (0)115 951 5402

Fax: +44 (0)115 951 5324

Email: ruth.filik@nottingham.ac.uk

Abstract

Theorists have debated whether our ability to understand sarcasm is principally determined by the context (Gibbs, 1994; Utsumi, 2000) or by properties of the comment itself (Giora, 1997; 2003; Grice, 1975).[1] The current research investigated an alternative view which broadens the focus on the comment itself, suggesting that mitigating a highly positive concept by using negation generates sarcastic interpretations by default (Giora et al., 2015a, 2018). In the current study, pre-tests performed on the target utterances presented in isolation established their default interpretations; novel affirmative phrases (e.g., *He is the best lawyer*) were interpreted literally, whereas equally novel negative counterparts (e.g., *He isn't the best lawyer*) were interpreted sarcastically. In Experiment 1 (an eye-tracking study), prior context biased these utterances towards literal or sarcastic interpretations. Results showed that target utterances were easier to process in contexts supporting their default interpretations, regardless of affirmation/negation. Results from a second eye-tracking experiment suggested that readers' tendency to interpret negative phrases sarcastically is related to their own tendency to use malicious humor. Our findings suggest that negation leads to certain ambiguous utterances receiving sarcastic interpretations by default and that this process may be further intensified by personality factors.

[1] Sarcasm pertains here to verbal irony

Imagine that you had just given a lecture, and when reading the student evaluations of your class, you came across the comment, “Not the best teacher.” How would you feel? It is likely that you would initially adopt the sarcastic interpretation, according to which you are far from being the best teacher, and would thus be disappointed by such negative feedback. However, this comment is actually ambiguous between the sarcastic interpretation and a literal interpretation, in which you are, in fact, a good teacher, but some others may be better. The fact that readers or listeners seem to preferentially adopt the sarcastic interpretation of certain utterances causes problems for a number of contemporary accounts of sarcasm comprehension, and will be the focus of the current paper.

From the description above, it can be seen that sarcasm is a form of verbal irony that is used with the intent to criticize. A number of theories have been put forward in the literature to explain how such ironic or sarcastic language is understood, including: The Standard Pragmatic Model (Grice, 1975; Searle, 1979; 1993), the Direct Access View (Gibbs, 1994; 2002), the Graded Salience Hypothesis (Giora, 1997; 2003), and the Defaultness Hypothesis (Giora, Givoni, & Fein, 2015a). Following the Standard Pragmatic Model, the function of (affirmative) verbal irony is to communicate the opposite of what is said. According to this account, the literal interpretation of an utterance is accessed first, and if a mismatch with context is detected, the utterance will be reanalyzed as being nonliteral (here sarcastic), and the literal interpretation will be suppressed. The extra stages involved in the processing of irony would then result in a processing cost for ironic compared to literal language. This, for example, could be observed in longer reading times for an ironic comment compared to the same comment uttered literally (e.g., Giora et al., 2007), or in longer

response times to sarcastically than literally related probe words (e.g., Fein, Yeari, & Giora, 2015).

In contrast to this modular, stage-like account, the Direct Access View (Gibbs, 1994; 2002) would predict that literal and ironic targets are processed in essentially the same way; given a sufficiently supportive context, the ironic meaning or interpretation of a phrase can be accessed or constructed directly, bypassing the literal interpretation. Thus, following this account we should expect no differences in reading times for literal and sarcastic materials (given equally supportive contexts across conditions).

In contrast to the Standard Pragmatic Model, the Graded Salience Hypothesis (Giora, 1997; 2003) states that it is meaning salience (i.e., an entrenched, coded meaning affected by e.g., familiarity, frequency of usage, or prototypicality), rather than literalness, that is key. Hence, salient meanings are activated automatically on encountering the stimulus, regardless of other factors affecting processing (such as contextual support). For salient ironies (such as "That's just great!", which are regularly encountered ironically, and thus have an ironic entry in the mental lexicon), the salient ironic meaning can be accessed directly, and thus there should be no associated processing cost (e.g., Filik, Leuthold, Wallington, & Page, 2014; Giora, & Fein, 1999).

For nonsalient noncoded ironies (i.e., novel phrases that the reader or listener is not used to encountering ironically), the predictions are the same as for the Standard Pragmatic Model, that is, the salience-based *interpretation* (based on the salient meanings of the utterance components) will be activated initially, and then reanalyzed as ironic, if contextual cues indicate that this is appropriate. In contrast to the Standard Pragmatic Model, the salience-based (often literal) interpretation of

sarcastic stimuli would be retained rather than suppressed, since the contrast between the ironic and salience-based literal interpretations underlies the function of irony (Giora, 1995), which is to draw attention to a failed expectation.

Enlarging the scope of the Graded Salience Hypothesis, the Defaultness Hypothesis (Giora et al., 2015a) focuses on *interpretations* rather than *meanings* (but will treat salient meanings as default responses). Default interpretations need to be constructed, rather than accessed directly from the mental lexicon. Still, they are constructed automatically. Thus, a phrase such as "He isn't the best teacher" will not have its sarcastic interpretation encoded in the mental lexicon. Nevertheless, the initial interpretation that is constructed unconditionally, regardless of strongly supportive context or explicit cues to the contrary, is the sarcastic one.

To be considered a default interpretation, (i) stimuli must be novel (noncoded/unfamiliar), in order to distinguish them from salient meanings. In addition, (ii) they must be potentially ambiguous between a literal and nonliteral interpretation, so that a preference is allowed. They should, therefore, be free of any kind of internal cues such as semantic anomaly (Beardsley, 1958) or internal incongruity (Partington, 2011), which prompt nonliteralness, and (iii) free of any biasing external cues, such as contextual information or tone of voice, which might rule out either interpretation (for a comprehensive consideration of the conditions necessary for default interpretations, see Giora et al., 2015a).

A number of recent studies have examined the hypothesis that certain negative constructions (in particular, phrases containing explicit negation modifying highly positive concepts, such as *Punctuality is not his forte, Ambitious she is not, He is not the most organized student*, or implicit negation modifying such concepts, as in *Do you really believe she is honest?*) can elicit sarcastic interpretations by default (e.g.,

Giora et al., 2015a; Giora, Drucker, Fein, & Mendelson, 2015b; Giora, Givoni, Heruti, & Fein, 2017; Giora et al., 2013; 2018). For example, Giora et al. (2015a, Experiment 1.1) demonstrated that when presented out of context, negative phrases (in Hebrew) of the form "X is not the most Y" (e.g., *He is not the most organized student*), received a sarcastic interpretation (i.e., participants rated the utterance interpretation as closer to "He is quite messy" than "He is quite orderly, but less so than others"), suggesting that their default interpretation is sarcastic. In contrast, corresponding affirmative items (e.g., *He is the most organized student*) were interpreted literally (Experiment 1.2). In addition, when participants were asked to explicitly rate how sarcastic the items were (Experiment 1.3), negative items were rated as being significantly more sarcastic than affirmative counterparts.

Giora et al. then conducted a self-paced reading study (Experiment 2) to examine how these negative and affirmative utterances were processed when presented in contexts biased towards either a literal or sarcastic interpretation of the target (strength of contextual bias was equally strong across all conditions, as determined by a pre-test). Results showed that reading times were shorter for default affirmative literal targets than for nondefault affirmative sarcastic counterparts and nondefault negative literal counterparts. Moreover, reading times were shorter for default negative sarcastic targets than for nondefault negative literal counterparts and nondefault affirmative sarcastic counterparts. The same pattern of effects was observed on a 'spillover' region (the two words following the critical target utterance). This was taken as evidence in support of the Defaultness Hypothesis in that default interpretations (i.e., sarcastic interpretations of negative items and literal interpretations of affirmative items) were easier to process than their nondefault

counterparts (i.e., literal versions of negative items and sarcastic versions of affirmative items).

The aim of the current Experiment 1 is to further test the predictions of the Defaultness Hypothesis by replicating and extending the study of Giora et al. (2015a). Firstly, we aim to investigate whether the findings reported by Giora et al. generalize from Hebrew to English. Secondly, since eye-tracking has successfully been used in a number of recent studies to investigate factors that can influence the on-line processing of written irony (e.g., Au-Yeung, Kaakinen, Liversedge, & Benson, 2015; Filik et al., 2014; Filik & Moxey, 2010; Kaakinen, Olkoniemi, Kinnari, & Hyönä, 2014; Olkoniemi, Ranta, & Kaakinen, 2016; Ţurcan & Filik, 2016; in press), we aim to use this methodology to examine the timecourse of processing in more detail. Specifically, in Experiment 1, we will monitor participants' eye movements as they are reading negative and affirmative phrases (such as *He isn't the most popular hairdresser*/*He is the most popular hairdresser*), which will be placed in contexts that bias either the sarcastic or literal interpretation of the target (see Table 1 for an example material in all four conditions).

Most of the theories and experiments discussed thus far consider how aspects of the text influence the ease in which a comment is interpreted as literal or sarcastic. However, it is also important to consider how aspects of the perceiver may influence this process. Thus, in addition, we will explore whether aspects of the reader's personality may influence the on-line processing and interpretation of the same negative utterances (Experiment 2). Specifically, we will investigate whether the participants' own tendency to use malicious humor influences on-line processing of ambiguous utterances such as the negative constructions tested here.

In terms of predictions for Experiment 1[2], given that the items contain novel ironies, following the Graded Salience Hypothesis, we would expect to observe processing difficulty for nonsalient ironic targets compared to salience-based literal counterparts. Specifically, we would expect longer reading times on critical portions of the text in nonsalient sarcastic than salience-based literal conditions. Following the Direct Access View, we would expect no differences across conditions, given equally strong supportive contexts in both cases. Finally, following the Defaultness Hypothesis, we would predict shorter reading times for default readings compared to nondefault readings, regardless of degree of non/literalness. Specifically, for affirmative materials, we would predict shorter reading times for default literal materials than for nondefault sarcastic versions. We would also predict shorter reading times for default literal affirmative materials than nondefault negative literal materials. For negative materials, we would predict shorter reading times for default sarcastic than nondefault literal conditions. We would also predict shorter reading times for default negative sarcasm than nondefault affirmative sarcasm.

Experiment 1

Method

Participants: Forty native English-speakers from the University of Nottingham community (28 females) participated in the experiment. They received a small monetary award as an inconvenience allowance for participating.

[2] We do not include predictions for the Standard Pragmatic Model for the following reason. According to this model, the ironic interpretation of an utterance is the opposite of the literal interpretation. Following this, for the negative materials used in the current study, such as "He's not the most popular hairdresser", the Standard Pragmatic Model would make erroneous predictions regarding the intended interpretation of these utterances when uttered sarcastically. Specifically, this model would predict that the sarcastic interpretation of "He's not the most popular hairdresser" would be "He is the most popular hairdresser", since this would be the opposite of not being the most popular hairdresser (see e.g., Coulson, 2005, and Tobin and Israel, 2012, for discussion).

Materials and Design: Twenty-four experimental materials were constructed (see Table 1 for an example; the full set is available from the first author).

Table 1. Example Material (Forward Slashes Denote Analysis Regions)

Affirmative	
Literal	Jenny and Emily walked past the local beauty salon and noticed that it was full of customers. "I always see people inside the salon," said Jenny. Emily replied: "Me too. The number of customers has definitely increased since the new owner took over and I believe his salon is busier than the others in town." "I heard the owner plans on expanding the salon to meet demand," said Jenny. Emily said: "Yes, I heard that too. It's not surprising though, as/ **he is**[pre-critical]/ **the most**[first-critical]/ **popular hairdresser."**[second-critical]/ "**You're right.**[spillover]/ It's because everybody in this town appreciates his skills," replied Jenny.
Sarcastic	Jenny and Emily walked past the local beauty salon and noticed that there were no customers. "I never see people inside the salon," said Jenny. Emily replied: "I know. The number of customers has definitely plummeted since the new owner took over." "The funny thing is," said Emily, "he goes around town bragging about how he has to turn customers away as he is so busy." "Indeed!" laughed Jenny. "Look at his empty salon. Clearly **he is**[pre-critical]/ **the most**[first-critical]/ **popular hairdresser."**[second-critical]/ **"You're right,"**[spillover]/ replied Emily. "He must be totally stupid as well."
Negative	
Literal	Jenny and Emily walked past the local beauty salon and noticed that it was full of customers. "I always see people inside the salon," said Jenny. Emily replied: "Me too. The number of customers has definitely increased since the new owner took over." "Still, it isn't as busy as the one in town. I heard the one in town is going to expand to meet demand," said Jenny. Emily said: "Yes I know. So even though our local salon owner is quite busy, **he isn't**[pre-critical]/ **the most**[first-critical]/ **popular hairdresser."**[second-critical]/ "**You're right.**[spillover]/Give him time and things might be different," replied Jenny.
Sarcastic	Jenny and Emily walked past the local beauty salon and noticed that there were no customers. "I never see people inside the salon," said Jenny. Emily replied: "I know. The number of customers has definitely plummeted since the new owner took over." "The funny thing is," said Emily, "he goes around town bragging about how he has to turn customers away because he is so busy." "Look at his empty salon," laughed Jenny, "clearly, **he isn't**[pre-critical]/ **the most**[first-critical]/ **popular hairdresser."**[second-critical]/ "**You're right."**[spillover]/ replied Emily. "He must be totally stupid as well".

A number of pre-tests were performed on the materials to ensure that they met the criteria for defaultness. As mentioned earlier, these criteria require that items be potentially ambiguous between literal and nonliteral interpretations. To avoid any bias, items should (i) share a similar degree of novelty, (ii) be free of semantic anomaly or internal incongruity, which invite nonliteralness, and (iii) be rated for degree of defaultness, when presented in isolation.

Novelty pre-test: Twenty participants rated affirmative and negative items presented in isolation on a scale from 1 (unfamiliar) to 7 (familiar). Both sets of items were perceived as being unfamiliar and there was no significant difference in novelty ratings between affirmative (M = 1.79, SD = .59) and negative (M = 1.67, SD = .59) items, $t < 1$.

Interpretation pre-test: Items presented in isolation were rated by 20 participants on a 7-point scale, with one end of the scale indicating a literal interpretation (e.g., 'He was a very popular hairdresser, but others were more popular'), and the other end indicating a sarcastic interpretation (e.g., 'He was far from being the most popular hairdresser'). A score of 1 would indicate a literal interpretation, and a score of 7 would indicate a sarcastic interpretation. Results showed that affirmative items (M = 1.80, SD = .56) were interpreted significantly less sarcastically than negative items (M = 4.79, SD = .75), $t(23) = 27.12$, $p < .001$. In support of the suggestion that negative items received a sarcastic interpretation by default, and affirmative items received a literal interpretation by default, the results of one-samples t-tests indicate that negative items were consistently rated above the midpoint of the scale, $t(23) = 5.14$, $p < .001$, and affirmative items were consistently rated below the midpoint, $t(23) = 19.33$, $p < .001$.

Criticism pre-test: In order to assess whether items presented out of context were perceived as being critical and derisive (i.e., sarcastic), 24 participants rated each item on a 7-point scale ranging between 1(not at all critical or derisive) to 7(extremely critical or derisive). Results showed that negative items (M = 5.30, SD = 1.09) were rated as being significantly more critical and derisive than affirmative items (M = 2.89, SD = 1.68), $t(23) = 4.9$, $p < .01$. Having ensured that items presented out of context met conditions (i-iii) for defaultness, we wished to assess that the contexts that we created for the eye-tracking experiment were equally strongly biasing towards the intended interpretation across all conditions.

Strength of contextual bias pre-test: Items were presented in context, and rated by 24 participants on a 7-point scale, with one end of the scale indicating a literal interpretation and the other end indicating a sarcastic interpretation (this was balanced across materials). For our purposes, a score of 1 would indicate weak contextual support, and a score of 7 would indicate strong contextual support for the intended interpretation. Results showed that for literal materials, there was no difference in contextual support between affirmative (M = 6.38, SD = .63) and negative (M = 6.17, SD = .69) materials, $t < 1$. For sarcastic items, affirmative materials (M = 6.55, SD = .61) were judged as receiving slightly more contextual support than negative materials (M = 6.20, SD = .66), $t(23) = 2.71$, $p < .05$. However, this would not explain any benefits in reading time predicted for negative compared to affirmative items, since the effect is in the opposite direction to our predictions. For affirmative items, there was no difference between literal (M = 6.38, SD = .63) and sarcastic (M = 6.55, SD = .61) conditions, $t < 1$. For negative items, there was no difference between literal (M = 6.17, SD = .69) and sarcastic (M = 6.20, SD = .66) conditions, $t < 1$.

Eye-tracking procedure: Eye movements were recorded via an SR Research EyeLink 1000 eye-tracker, which sampled eye position every millisecond. Viewing was binocular, but only the right eye was recorded. Materials were displayed on a 17-inch monitor, 56 cm from participants' eyes. Before the experiment, the procedure was explained and participants were instructed to read normally. Participants were seated at the eye-tracker and their head was placed on a chin and forehead rest to minimize movements. Participants then completed a calibration procedure. Before the start of each trial, a fixation box appeared in the upper left quadrant of the screen. Once the participant fixated this box, the stimulus computer displayed the text. If the participant's apparent point of fixation did not match with the fixation box then the experimenter re-calibrated the eye-tracker. Once a participant completed reading each item, they fixated a post-it note that was attached to the lower right-hand edge of the monitor, and pressed a key. A question (e.g., *Were Jenny and Emily talking about the hairdresser?*) was displayed following 25% of trials in order to ensure that participants were reading for comprehension. An average correct response rate of 89.4 % indicated that this was the case.

Data analysis. Materials were divided into analysis regions as shown in Table 1. The *pre-critical region* comprised the beginning of the target phrase, and typically contained the words "(s)he is(n't)". The *first-critical region* was the first point in the sentence at which the participant might be able to detect whether the comment is intended literally, or sarcastically, and typically comprised words such as "the best/the most". The phrase was fully disambiguated at the *second-critical region*, which contained the remainder of the utterance (e.g., "popular hairdresser"). The *spillover region* was the following two words.

An automatic procedure pooled short contiguous fixations. Fixations under 80 ms were incorporated into larger adjacent fixations within one character and fixations under 40 ms that were not within three characters of another fixation were deleted, as were fixations over 1200 ms. Prior to analysis, trials where participants failed to read the sentence or there had been track loss, were eliminated. Specifically, we removed trials where two or more adjacent regions had zero first-pass reading times, accounting for 8.02 % of the data.

Four measures of reading behavior are reported. Measures of early processing are *first-pass reading time*, which sums all fixations in a region until the point of fixation leaves the region either to the left or the right, and *regression path* (or *go-past*) *reading time*, which is the sum of fixations from the time that a region is first entered until a saccade transgresses the right region boundary. Measures indicative of later processing are *second-pass reading time*, which sums the duration of fixations in a region after having left it either to the left or the right, and *total reading time*, which sums the duration of all fixations made within a region.

When reading times were zero for a particular region, the relevant point was excluded from the analysis, and means were calculated from the remaining data points in the design cell. In the pre-critical region, this procedure accounted for 21.97% of first-pass and regression path data, and 10.42% of total time data; in the first-critical region, 18.91% of first-pass and regression path and 10.42% of total time data; in the second-critical region, 2.38% of first-pass and regression path and 2.15% of total time data, and in the spillover region, 9.40% of first-pass and regression path and 4.19% of total time data.

Results and Discussion

Data for each region were subjected to two 2 *polarity* (affirmative vs. negative) x 2 *context* (literal vs. sarcastic) within-subjects ANOVAs, one treating participants (*F*1) and one treating items (*F*2) as random variables (see Table 2 for descriptive statistics, and Figures 1-3 for illustrations of the key findings).

Table 2. Descriptive statistics for Experiment 1.

Region	Measure (msec)	Affirmative Literal		Affirmative Sarcastic		Negative Literal		Negative Sarcastic	
		M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Pre-critical	First-pass	267	15.6	265	14.0	286	16.7	275	19.2
	Regression path	329	19.8	372	22.0	376	26.2	380	29.1
	Second-pass	106	13.4	199	29.4	142	17.6	133	15.0
	Total times	354	18.1	440	31.5	403	23.9	390	19.8
1st Critical	First-pass	236	10.5	225	10.1	239	9.6	230	10.1
	Regression path	311	16.2	342	27.9	307	18.6	308	16.6
	Second-pass	108	13.4	175	21.9	136	21.1	98	11.5
	Total times	340	17.5	390	22.4	362	22.9	313	13.4
2nd Critical	First-pass	466	20.7	471	30.0	515	35.0	476	22.2
	Regression path	646	27.0	775	48.4	785	59.9	710	32.8
	Second-pass	141	18.9	192	30.2	162	28.3	141	22.8
	Total times	611	28.3	665	41.3	682	47.0	620	25.0
Spillover	First-pass	319	19.3	325	18.0	309	17.5	318	17.9
	Regression path	370	19.8	390	36.7	404	28.4	407	64.1
	Second-pass	62	9.8	89	18.7	56	8.8	59	10.5
	Total times	372	20.4	398	24.6	356	20.0	352	18.6

Pre-critical region "(s)he is(n't)": There were no significant effects in first-pass or regression path reading times (*F*s < 2.3, *p*s > .14).

There was no main effect of polarity in either second-pass or total reading times, *F*s < 1. Second-pass reading times showed a significant main effect of context, $F1(1, 39) = 4.79$, $p < .05$; $F2(1, 23) = 4.55$, $p < .05$, but total reading times did not, *F*s < 2.9, *p*s > .09. Both measures of reading time showed an interaction, second-pass:

*F*1(1, 39) = 12.17, *p* < .01, *F*2(1, 23) = 4.29, *p* = .05; total reading times: *F*1(1, 39) = 9.07, p < .01; *F*2(1, 23) = 3.53, *p* = .07. The fact that effects were only observed in later measures of reading time would suggest that these effects are due to participants re-reading the beginning of the sentence in conditions in which they experienced difficulty.

Analysis of simple main effects showed that for literal sentences, affirmative targets had shorter reading times than negative targets in both second-pass, *F*1(1, 39) = 4.75, *p* < .05; *F*2(1, 23) = 1.84, *p* = .19, and total reading times, *F*1(1, 39) = 5.86, *p* < .05; *F*2(1, 23) = 2.17, *p* = .16 (significant by participants but not by items). This finding is consistent with the predictions of the Defaultness Hypothesis, in that default interpretations had shorter reading times than nondefault counterparts. Note, however, that this result could also be explained by theories of negation which would predict more difficulty for negative than affirmative materials (e.g., Clark & Clark, 1977; Horn, 1989).

For sarcastic sentences, negative materials tended to have shorter reading times than affirmative materials in both second-pass, *F*1(1, 39) = 5.91, *p* < .05, *F*2(1, 23) = 3.40, *p* = .08 (significant by participants only), and total reading times, *F*1(1, 39) = 3.52, *p* = .07, *F*2(1, 23) = 1.46, *p* = .24 (approaching significance by participants but not significant by items). This finding is consistent with the predictions of the Defaultness Hypothesis (since default interpretations received shorter reading times than nondefault interpretations), but cannot be explained by Direct Access theories of sarcasm comprehension (which would not predict differences between negative and affirmative sarcastic materials presented in equally strong contexts).

Both measures showed that for affirmative sentences, there were significantly shorter reading times for literal materials than sarcastic counterparts, second-pass: $F1(1, 39) = 11.14$, $p < .01$; $F2(1, 23) = 8.67$, $p < .01$; total reading times: $F1(1, 39) = 7.66$, $p < .01$; $F2(1, 23) = 5.25$, $p < .05$, a finding which is consistent with all theories under consideration (except for the Direct Access View). For negative sentences there were no differences (Fs < 1).

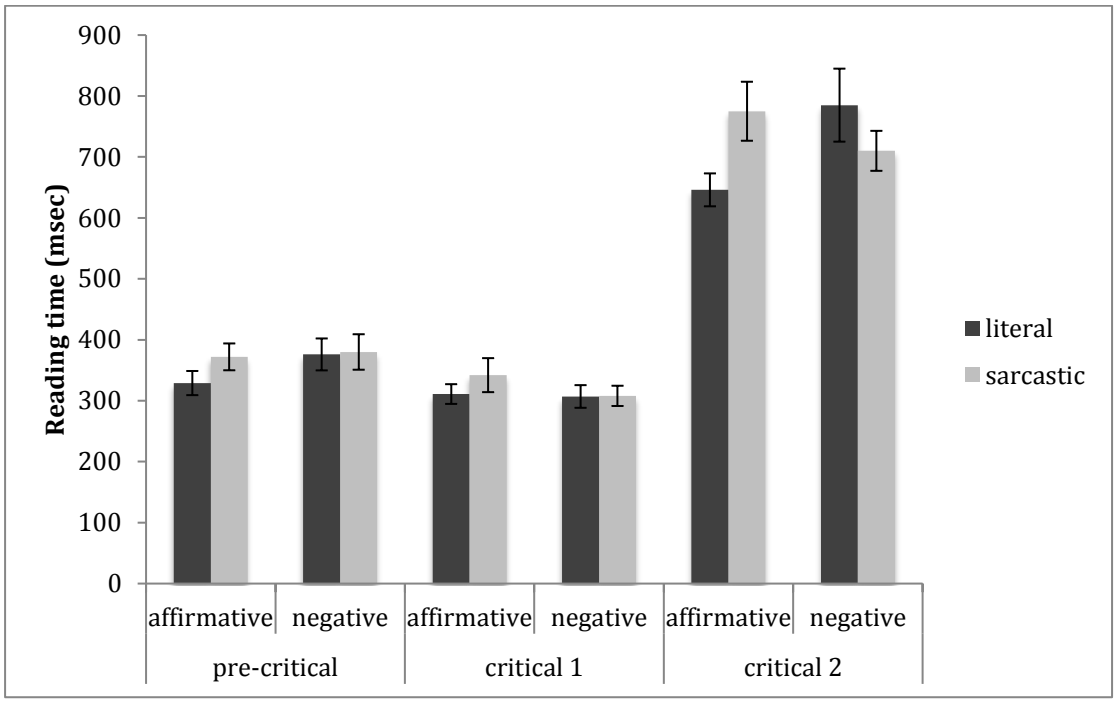

Figure 1: Regression path reading times for pre-critical, critical 1, and critical 2 regions (error bars show standard errors of the mean).

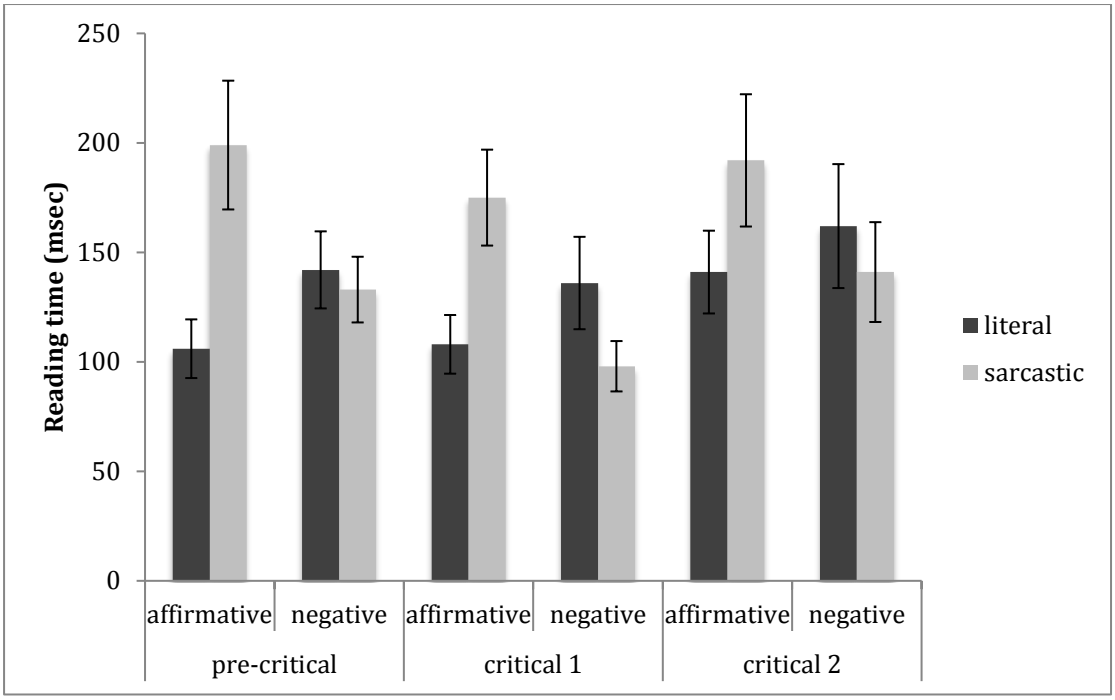

Figure 2: Second-pass reading times for pre-critical, critical 1, and critical 2 regions (error bars show standard errors of the mean).

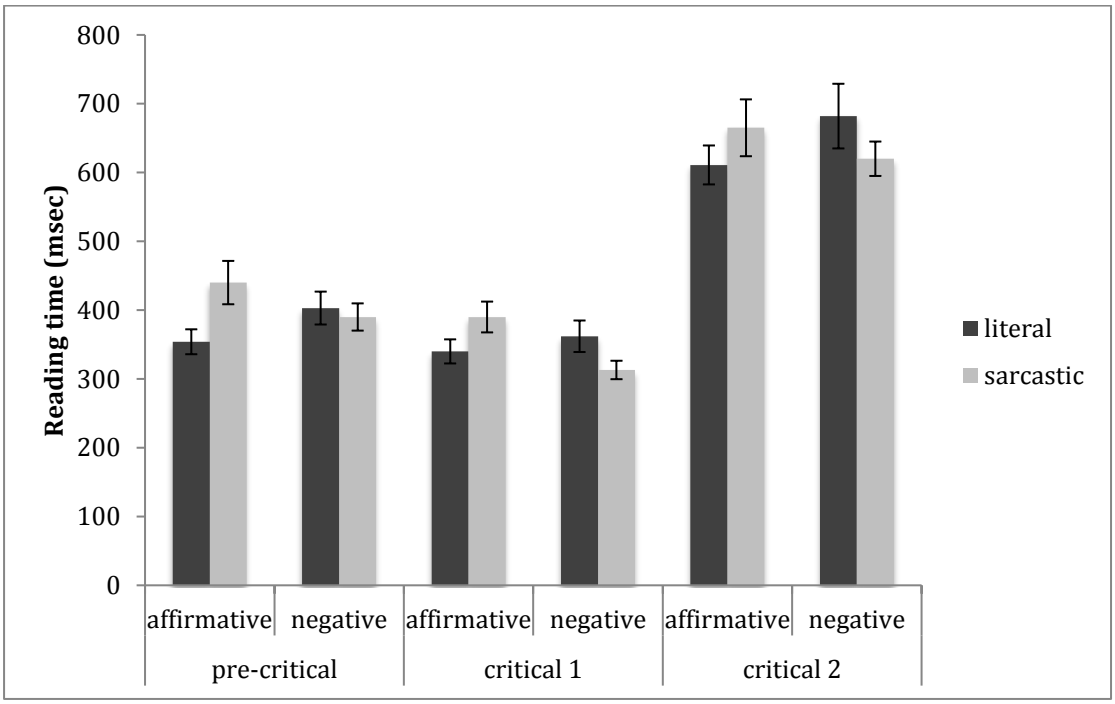

Figure 1: Total reading times for pre-critical, critical 1, and critical 2 regions (error bars show standard errors of the mean).

First critical region "the most": There were no significant effects in first-pass or regression path reading times, *F*s < 1 .9, *p*s > .17.

There were no significant main effects in second-pass or total reading times *F*s < 2.2, *p*s > .15. However, both measures revealed a significant interaction, second-pass: $F1(1, 39) = 13.38$, $p < .01$; $F2(1, 23) = 6.49$, $p < .05$; total reading times: $F1(1, 39) = 8.08$, $p < .01$; $F2(1, 23) = 7.07$, $p < .05$. Finding effects in late but not early measures would again suggest that effects are due to participants looking back and re-reading earlier portions of the sentence in certain conditions.

Simple main effects for both measures showed that for literal sentences, there were no differences between affirmative and negative conditions, *F*s < 1.6, *p*s > .22. However, for sarcastic materials, there were shorter reading times for negative than affirmative sentences, second-pass: $F1(1, 39) = 11.56$, $p < .01$, $F1(1, 23) = 7.02$, $p < .05$; total reading times: $F1(1, 39) = 11.80$, $p < .01$, $F2(1, 23) = 8.67$, $p < .01$. This finding would only be predicted by the Defaultness Hypothesis, according to which default interpretations (here, negative sarcasm) will supersede nondefault counterparts (here, affirmative sarcasm).

For affirmative sentences, there were shorter reading times for literal interpretations than sarcastic alternatives, second-pass: $F1(1, 39) = 10.13$, $p < .01$; $F2(1, 23) = 3.95$, $p < .06$; total reading times: $F1(1, 39) = 3.85$, $p < .06$; $F2(1, 23) = 2.48$, $p = .13$, whereas for negative sentences there was a trend for shorter reading times for sarcastic materials than for literal counterparts, second-pass: $F1(1, 39) = 3.42$, $p = .07$; $F2(1, 23) = 2.02$, $p = .17$; total reading times: $F1(1, 39) = 4.45$, $p < .05$; $F2(1, 23) = 3.84$, $p = .06$. Again, this pattern of effects (although not robustly significant) is fully compatible with the predictions of the Defaultness Hypothesis, in that there were shorter reading times for default than nondefault interpretations.

Second critical region "popular hairdresser.": There were no significant first-pass effects, *F*s < 3.2, *p*s > .08. Regression path reading times showed no main effects, *F*s < 1.2, *p*s > .28, but there was a significant interaction, $F1(1, 39) = 6.14$, $p < .05$; $F2(1, 23) = 4.75$, $p < .05$. This finding would suggest that this was the first point in the critical sentence at which participants experienced processing difficulty (which they responded to by engaging in re-reading). Simple main effects showed that for literal materials, negative sentences had longer reading times than affirmative sentences, $F1(1,39) = 5.37$, $p < .05$; $F2(1,23) = 3.05$, $p = .09$, a finding which is most readily explained by the Defaultness Hypothesis. For sarcastic materials, there were no differences, *F*s < 2.1, $p > .16$.

For affirmative sentences, there were shorter reading times for literal than sarcastic conditions, $F1(1,39) = 8.97$, $p < .01$; $F2(1,23) = 6.39$, $p < .05$, which would be compatible with all theories except for the Direct Access View. For negative sentences there were no differences, *F*s < 1.8, *p*s > .19. There were no significant second-pass, *F*s < 2.2, *p*s > .15, or total time effects, *F*s < 3.1, *p*s > .08.

Spillover region "You're right.": There were no significant first-pass effects, *F*s < 2.1, *p*s > .16. Regression path reading times showed no significant main effects, *F*s < 1. The interaction was significant by items but not by participants, $F1 < 1$; $F2(1, 23) = 4.45$, $p < .05$. Second-pass reading times showed no significant effects, *F*s < 2.6, *p*s > .11. Total reading times showed a marginal effect of polarity, $F1(1, 39) = 3.92$, $p < .06$, $F2(1, 23) = 4.57$, $p < .05$. There were no other significant effects, *F*s < 1.

Since the current study relates to incremental interpretation of phrases, rather than accessing the meanings of individual words, it is also informative to report findings that relate more to the end-point of this process, that is, total reading times

for the entire utterance (e.g., 'he is/isn't the most popular hairdresser'). Total reading times for the entire utterance also showed a significant interaction, $F1(1, 39) = 13.41$, $p < .005$; $F2(1, 23) = 4.37$, $p < .05$.

For literal sentences, affirmative materials ($M = 1225$, $SE = 51.8$) were faster to process than negative materials ($M = 1366$, $SE = 79.5$), $F1(1, 39) = 4.25$, $p < .05$; $F2(1, 23) = 2.14$, $p = .16$, significant by participants but not by items. For sarcastic sentences, negative materials ($M = 1253$, $SE = 49.8$) were faster to process than affirmative materials ($M = 1414$, $SE = 85.5$), $F1(1, 39) = 6.60$, $p < .05$, $F2(1, 23) = 4.46$, $p < .05$.

For affirmative sentences, there were shorter reading times for literal than sarcastic conditions, $F1(1, 39) = 9.59$, $p < .005$; $F2(1, 23) = 3.20$, $p = .09$, significant by participants but not by items. For negative sentences there was a trend towards shorter reading times for sarcastic than literal materials, $F1(1,39) = 3.11$, $p = .09$; $F2(1, 23) = 1.78$, $p = .20$.

In sum, taken together, results were generally most compatible with the predictions of the Defaultness Hypothesis, in that reading times were shorter for materials receiving default interpretations than for those receiving nondefault interpretations. This would suggest that it is characteristics of the comment itself (here relating to non/defaultness) that play a key role in how it is both processed on-line and interpreted off-line (as evidenced by the results of the pre-tests). In the next experiment we also consider the possibility that characteristics of the perceiver may influence the on-line processing of similar materials.

Experiment 2

The use of sarcasm has been argued to serve a number of social functions, such as eliciting a particular emotional response in the recipient (e.g., Brown &

Levinson, 1987; Colston, 1997; Dews & Winner, 1995; Leech, 1983; see Filik, Brightman, Gathercole, & Leuthold, 2017, for recent discussion). Thus, it seems likely that individual difference factors may play a substantial role in how it is processed and understood (see e.g., Blasko & Kazmerski, 2006). Indeed, a number of recent eye-tracking studies have examined the role of individual variation in the on-line processing of written sarcasm. For example, some studies suggest that an individual's working memory capacity and emotional processing abilities (see, e.g., Olkoniemi et al., 2016) as well as whether they are autistic or not (Au-Yeung et al., 2013) can influence eye movements during reading of sarcastic comments.

Given this evidence, it seems relevant to consider which individual differences might influence an individual's tendency to interpret certain phrases as sarcastic by default. A recent study examining individual differences in the use of verbal irony was conducted by Bruntsch and Ruch (2017). They found that an individual's tendency to use irony was higher when they also scored highly on certain personality factors, such as their tendency to use ridicule aggressively. Thus, in the current experiment, we examine whether an individual's tendency to use malicious humor themselves may influence their on-line processing of negative statements that are ambiguous between a literal and sarcastic interpretation.

To measure participants' level of malicious humor, we used the Indirect Aggression Scale Aggressor Version (IAS-A; Forrest, Eatough, & Shevlin, 2005), which is a reliable and valid 25-item scale measuring the level of three facets of indirect aggression ('Malicious Humor', 'Social Exclusionary', and 'Guilt Induction'). Examples of the nine items that make up the malicious humor subscale are: 'Made fun of people in public', 'Used sarcasm to insult them', 'Criticized them in

public'. Participants respond to these statements by indicating how often they have used each behavior against others, from 1 - 'Never' to 5 - 'Regularly'.

In Experiment 1, we used context to provide disambiguating information for the subsequently presented target utterance, and then analyzed participants' reading behavior on the target utterance in relation to its fit with the prior context. In contrast, in Experiment 2, participants' eye movements were monitored while they read short scenarios containing negative statements that were later disambiguated as being intended literally or sarcastically. For example, a comment such as "He isn't the most popular hairdresser" could be followed by a statement such as "I see what you mean… A few of the other hairdressers are more popular" (indicating a literal interpretation in which the hairdresser is popular, but a small number of others are more popular) or "I see what you mean… Many of the other hairdressers are more popular" (indicating a sarcastic interpretation in which he is not popular, since a large number of others are more popular - see Table 3 for full examples in each condition). This approach allowed participants time to establish an interpretation of the ambiguous negative utterance, which would later be confirmed or disconfirmed by subsequently presented text, allowing us to directly tap into their interpretations of the target utterance. Following the eye-tracking study, participants completed the indirect aggression scale.

If participants' tendency to use malicious humor influences their on-line processing and comprehension of ambiguous phrases, then we would expect to observe a relationship between reading times and their scores on the malicious humor subscale. Specifically, in order to calculate ease of processing of the sarcastic interpretation of the negative phrases, we will subtract reading times for the sarcastic condition from the literal condition for each measure of reading time for the analysis

regions highlighted in Table 3, with the rationale that a larger number is indicative of a stronger preference for the sarcastic interpretation. Thus, if someone's tendency to use malicious humor can influence their processing of ambiguous phrases, we might expect to observe a positive relationship between these (literal minus sarcastic) reading times, and scores on the malicious humor subscale of the indirect aggression questionnaire.

Method

Participants: Forty-eight native English-speakers from the University of Nottingham community (39 females, mean age 26.79, *SD* = 8.52) participated in the experiment. They received a small monetary award as an inconvenience allowance for participating.

Materials and Design: Twenty-eight experimental materials were constructed based on those used in Experiment 1 (see Table 3 for an example; the full set is available from the first author).

Table 3. Example Material Experiment 2 (Forward Slashes Denote Analysis Regions).

Literal	Joan and Sarah looked into the window of the local hairdresser's salon./ "He isn't the most popular hairdresser," commented Joan./ "I see what you mean," replied Sarah, "A few of the/ other hairdressers$^{\text{pre-critical}}$/ **are more popular."$^{\text{critical}}$**/ They debated$^{\text{spillover}}$/ where to go for lunch./
Sarcastic	Joan and Sarah looked into the window of the local hairdresser's salon./ "He isn't the most popular hairdresser," commented Joan./ "I see what you mean," replied Sarah, "Many of the/ other hairdressers$^{\text{pre-critical}}$/ **are more popular."$^{\text{critical}}$**/ They debated$^{\text{spillover}}$/ where to go for lunch./

Procedure: The eye-tracking procedure was identical to Experiment 1. Following the eye-tracking task, the IAS-A questionnaire (Forrest et al., 2005) was administered via Qualtrics.

Data analysis. Materials were divided into analysis regions as shown in Table 3. The *pre-critical region* comprised the phrase "other X" (e.g., "other hairdressers"). The *critical region* was the point in the sentence at which the participant should be able to decide whether the previous comment was intended literally, or sarcastically, and typically comprised phrases such as "are more popular". The *spillover region* was the following two words.

The same data pooling procedures were used as in Experiment 1. Specifically, we removed trials where two or more adjacent regions had zero first-pass reading times, accounting for 9.30 % of the data. The same four measures of reading behavior are reported as in Experiment 1. When reading times were zero for a particular region, the relevant point was excluded from the analysis, and means were calculated from the remaining data points in the design cell. In the pre-critical region, this procedure accounted for 11.15% of first-pass and regression path data, and 3.44% of total time data; in the critical region, 0.57% of first-pass and regression path and 0.16% of total time data, and in the spillover region, 11.39% of first-pass and regression path and 0% of total time data.

Results and Discussion

Data for each region were subjected to two 2 *interpretation* (literal vs. sarcastic) paired-samples t-tests, one treating participants ($t1$) and one treating items ($t2$) as random variables (see Table 4 for descriptive statistics).

Table 4. Descriptive statistics for Experiment 2.

Region	Measure (msec)	Literal *M*	Literal *SE*	Sarcastic *M*	Sarcastic *SE*
Pre-critical	First-pass	334	14.1	337	15.4
	Regression path	382	18.9	373	17.1
	Second-pass	85	9.3	91	9.6
	Total times	398	16.2	405	18.1
Critical	First-pass	486	23.0	504	24.0
	Regression path	659	24.0	681	28.5
	Second-pass	144	16.5	127	14.2
	Total times	629	26.0	630	25.7
Spillover	First-pass	282	13.3	279	14.2
	Regression path	329	19.2	343	24.5
	Second-pass	124	14.2	120	12.8
	Total times	388	20.4	375	19.6

There were no significant effects in any measure of reading time in the pre-critical (ts < 1.2, ps > .26), critical (ts < 1.5, ps > .16), or spillover region (ts < 1). Thus, it seems that when averaged across all participants, there was not a clear preference for either interpretation. We next examined whether participants' likelihood of adopting a sarcastic interpretation is influenced by their own tendency to use malicious humor.

Malicious Humor Scores

Each participant's Malicious Humor score was computed (Cronbach's alpha = .75). Scores could range from 9 (never uses malicious humor against others) to 45 (regularly uses malicious humor against others). Participants' scores ranged from 10 to 26 (M = 14.48, SD = 3.82).

Examining the Relationship Between Reading Time and Malicious Humor

In order to examine the relationship between participants' tendency to interpret a negative phrase sarcastically and their malicious humor score, we first needed to calculate a measure of their tendency towards a sarcastic interpretation. To do this, for each measure of reading time on each region of text, participants' reading times for sarcastic materials were subtracted from their reading times for literal materials, with the rationale that larger resulting scores would indicate a greater tendency to interpret the phrases sarcastically. Next, in order to investigate whether these reading scores were related to malicious humor, bivariate correlations were performed between the malicious humor scores and participants' literal minus sarcasm scores for each measure of reading time for the pre-critical, critical, and spillover regions (see Table 5).

Results showed that both the regression path and total reading time literal minus sarcasm scores in the critical region were positively related to the level of use of malicious humor (see Figure 4, for illustration).

Table 5. Bivariate Correlations (2-tailed) for Malicious Humor and Eye-tracking Scores (literal minus sarcastic) Experiment 2.

Region	Measure	Malicious Humor *r*
Pre-critical	First-pass	.05
	Regression path	.07
	Second-pass	.04
	Total times	.04
Critical	First-pass	.14
	Regression path	.32*
	Second-pass	.25
	Total times	.32*

Spillover	First-pass	
	Regression path	
	Second-pass	
	Total times	

Note. * = $p = 0.03$

To investigate whether participants' use of malicious humor may predict their reading behavior, we performed two simple regressions with participants' malicious humor score as the predictor variable: one with the regression path literal minus sarcasm scores in the critical region as the criterion variable, and another with the total reading time literal minus sarcasm scores in the critical region as the criterion variable. Results showed that more regular use of malicious humor predicted larger regression path literal minus sarcasm scores (i.e., a greater tendency to interpret materials sarcastically) in the critical region and larger total reading time literal minus sarcasm scores in the critical region (see Tables 6 and 7 for regression results).

Table 6. Simple Regression for Total Reading Times (L-S) Score in Critical Region for Experiment 2.

Model	*B*	*SE b*	*ß*	*t*	Sig.
Constant	-156.67	70.80		-2.21	.032
Malicious Humor	10.78	4.73	.32	2.28	.027

Note. $R^2 = .08$

Table 7. Simple Regression for Regression Path (L-S) Score in Critical Region for Experiment 2.

Model	*B*	*SE b*	*ß*	*t*	Sig.
Constant	-211.02	85.17		-2.48	.017
Malicious Humor	13.10	5.69	.32	2.30	.026

Note. $R^2 = .10$

Figure 4: Scatter plot illustrating the relationship between literal minus sarcastic regression path reading times and malicious humor scores.

General Discussion

In summary, results from Experiment 1 showed that readers experienced relative ease in processing: 1) default affirmative literal items compared to nondefault negative literal items and affirmative sarcastic items. They also experienced relative ease of processing for (2) default negative sarcastic items compared to nondefault negative literal items and affirmative sarcastic items. A number of these results can be explained by traditional accounts of irony processing. Specifically, the finding that affirmative sarcasm is more difficult to process than affirmative literal materials can be readily explained by the Graded Salience Hypothesis, which would predict longer reading times for sarcastic than literal materials, due to the extra steps involved in understanding novel ironies. In addition, the finding that negative literal materials are

more difficult to process than affirmative literal materials can be accounted for by theories of negation. However, the full pattern of effects is best accommodated by the Defaultness Hypothesis, since it was the case that readers experienced more processing difficulty with nondefault (i.e., affirmative sarcastic and negative literal) than default (i.e., affirmative literal and negative sarcastic) interpretations.

In relation to the empirical literature, the current findings both support and extend those of Giora et al. (2015a), who also found this pattern of results in a self-paced reading study, using Hebrew materials. The results of the current experiment demonstrate that these findings generalize to English materials. In addition, with the use of eye-tracking during reading, we were able to examine in more detail the timecourse of processing. From the current eye-tracking results we were able to determine that the first point at which readers appear to experience processing difficulty is in regression path reading times for the part of the sentence where the intended interpretation is ultimately disambiguated as being literal or sarcastic (i.e., the second critical region which contained words such as "popular hairdresser"). Interestingly, there were no differences in early reading time measures on the first critical region (e.g., the region containing words such as "the best/the most"), which is the first point at which readers may have been able to construct the intended interpretation. This may reflect the more incremental nature of the interpretative processes involved in the current study, as opposed to some previous studies that have found effects in early reading measures, when investigating issues relating to lexical interpretation of salient versus nonsalient ironic expressions (e.g., Filik et al., 2014). Finally, it appears that readers go back and re-read the beginning of the sentence as evidenced by longer second-pass and total reading times on early portions of the

target sentence, which provides evidence that some degree of reanalysis may be required in the case of nondefault interpretations.

Results from Experiment 2 suggested that a reader's tendency to interpret certain phrases sarcastically can be influenced by factors of their personality, specifically, their tendency to use malicious humor against others. That is, the more regularly participants reported using malicious humor towards others (as indicated by their scores on the indirect aggression scale), the higher their reading times scores for literal minus sarcastic conditions, indicating a greater tendency to interpret the scenarios sarcastically. This provides some evidence that individual difference factors can influence on-line processes in text comprehension. This finding would add to recent evidence from other eye-tracking studies showing that factors such as working memory capacity, emotional processing ability (e.g., Olkoniemi et al., 2016), as well as the presence of autistic traits (e.g., Au-Yeung et al., 2013) can influence the on-line processing of sarcasm. It is important to note that some aspects of the results from Experiment 2 (i.e., no overall difference in reading times between literal and sarcastic conditions) could be taken as support for the Direct Access View (which would predict such a null result). However, the Direct Access View cannot explain the results from Experiment 1, nor the influence of personality factors in Experiment 2.

In conclusion, we conducted two eye-tracking experiments in which we investigated the on-line processing of written sarcasm. Results from Experiment 1 were most compatible with predictions of the Defaultness Hypothesis (Giora et al., 2015a), in that reading times suggested that participants experienced relative processing ease with default than nondefault interpretations of ambiguous phrases. Results from Experiment 2 suggest that aspects of a reader's personality, in particular,

their tendency to use malicious humor against others, can influence their on-line processing and interpretation of ambiguous phrases.

Acknowledgements

The work reported in this paper was supported by a grant from the British Academy International Partnership and Mobility Scheme (PM140296) awarded to Ruth Filik and Rachel Giora.

References

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.

Beardsley, C. M. (1958). *Aesthetics*. New York: Harcourt, Brace and World.

Blasko, D. G., & Kazmerski, V. A. (2006). ERP Correlates of individual differences in the comprehension of nonliteral language. *Metaphor and Symbol*, *21*, 267-284.

Brown, P., &, Levinson, S. (1987). *Politeness*. Cambridge, UK: Cambridge University Press.

Bruntsh, R., & Ruch, W. (2017). The role of humor-related traits and broad personality dimensions in irony use. *Personality and individual differences*, *112*, 139-143.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics* (Vol. 515). J. Kagan (Ed.). New York, NY: Harcourt Brace Jovanovich.

Colston, H. L. (1997). Salting a wound or sugaring a pill: The pragmatic functions of ironic criticisms. *Discourse Processes*, *23*, 24–45.

Coulson, S. (2005). Sarcasm and the space structuring model. In S. Coulson and B. Lewandowska-Tomaszczyk (Eds.) *The Literal and Nonliteral in Language and Thought*, (pp. 129–144). Frankfurt am Main, Germany: Peter Lang.

Dews, S., & Winner, E. (1995). Muting the meaning: A social function of irony. *Metaphor and Symbolic Activity*, *10*, 3-19.

Fein, O, Yeari, M., & Giora, R. (2015). On the priority of salience-based interpretations: The case of irony. *Intercultural Pragmatics*, *12*, 1-32.

Filik, R., Brightman, E., Gathercole, C., & Leuthold, H. (2017). The emotional impact of verbal irony: Eye-tracking evidence for a two-stage process. *Journal of Memory and Language, 93,* 193-202.

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.

Filik, R., & Moxey, L. M. (2010). The on-line processing of written irony. *Cognition, 116,* 421-436.

Forrest, S., Eatough, V., & Shevlin, M. (2005). Measuring adult indirect aggression: The development and psychometric assessment of the indirect aggression scales. *Aggressive Behavior, 31,* 84-97.

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge, UK: Cambridge University Press.

Gibbs, R. W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics, 34,* 457–486.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics, 8,* 183-206.

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

Giora, R., Drucker, A., Fein, O., & Mendelson, I. (2015b). Default sarcastic interpretations: On the priority of nonsalient interpretations. *Discourse Processes, 52,* 173-200.

Giora, R., & Fein, O. (1999). Irony: Context and salience. *Metaphor and Symbol, 14,* 241-257.

Giora, R., Fein, O., Laadan, D., Wolfson, J. Zeituny, M., Kidron, R., Kaufman, R., &

Shaham, R. (2007). Expecting irony: Context vs. salience-based effects. *Metaphor and Symbol, 22,* 119-146.

Giora, R., Givoni, S., & Fein, F. (2015a). Defaultness reigns: The case of sarcasm. *Metaphor and Symbol*, *30*, 290-313.

Giora, R., Givoni, S., Heruti, V., & Fein, O. (2017). The role of defaultness in affecting pleasure: The optimal innovation hypothesis revisited. *Metaphor & Symbol*, *32,* 1-18.

Giora, R., Jaffe, I., Becker, I. & Fein, O. (2018). Strongly attenuating highly positive concepts: The case of default sarcastic interpretations. *Review of Cognitive Linguistics*.

Giora, R., Livnat, E. Fein, O., Barnea, A., Zeiman, R., & Berger, I. (2013). Negation generates nonliteral interpretations by default. *Metaphor & Symbol*, *28*, 89-11.

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

Horn, L. R. (1989). *A natural history of negation*. Chicago, IL: University of Chicago Press.

Kaakinen, J. K., Olkoniemi, H., Kinnari, T., and Hyönä, J. (2014). Processing of written irony: An eye movement study. *Discourse Processes, 51,* 287-311.

Leech, G. N. (1983). *Principles of pragmatics*. London: Longman.

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.

Partington, A. (2011). Phrasal irony: Its form, function and exploitation. *Journal of Pragmatics, 43,* 1786–1800.

Searle, J. R. (1979), *Expression and meaning: Studies in the theory of speech and acts.* Cambridge, UK: Cambridge University Press.

Searle, J. (1993). Metaphor, in A. Ortony (Ed.) *Metaphor and thought* (pp. 83-111), Cambridge, UK: Cambridge University Press.

Tobin, V., & Israel, M. (2012). Irony as a viewpoint phenomenon. In B. Dancygier & E. Sweetser (Eds.), *Viewpoint in Language: A Multimodal Perspective* (pp. 25-46). Cambridge: Cambridge University Press.

Țurcan, A., and Filik, R. (2016). An eye-tracking investigation of written sarcasm comprehension: The roles of familiarity and context. *Journal of Experimental Psychology: Learning, Memory and Cognition, 42,* 1867-1893.

Turcan, A., & Filik, R. (in press). Investigating sarcasm comprehension using eye-tracking during reading: What are the roles of literality, familiarity, and echoic mention? In A. Athanasiadou & H. Colston (Eds.), *Irony in language use and communication*. FTL series. Amsterdam: John Benjamins.

Utsumi, A. (2000). Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from nonirony. *Journal of Pragmatics, 32,* 1777-1806.