

The effect of speaker age on the perception of ironic insults

Debra Jared and Alyssa Pandolfo

University of Western Ontario

Author Note

Debra Jared and Alyssa Pandolfo, Department of Psychology, University of Western Ontario. Alyssa Pandolfo is now at University College London, London, UK.

This work was supported by an NSERC USRA to AP and an NSERC Discovery grant to DJ.

We would like to thank Sarah Bainbridge and Huda Al-Sharafi for assistance in data collection.

Correspondence concerning this article should be sent to: Debra Jared, Department of Psychology, Brain & Mind Institute, Western Interdisciplinary Research Building, University of Western Ontario, 302 Perth Drive, London, Ontario Canada N6A 3K7

©Canadian Psychological Association, 2020. The Canadian Journal of Experimental Psychology home page can be found here: <https://www.apa.org/pubs/journals/cep>. This article may not exactly replicate the final version published in the CPA journal. It is not the copy of record. Please do not copy or cite without author permission.

Abstract

We investigated a cue that readers may use in determining whether a remark such as “You are so helpful!” is intended as a compliment or an ironic insult. The cue was the age of the speaker. Remarks were preceded by a sentence that either invited a literal or ironic interpretation of the remark. Data were collected on the familiarity of the remark as an ironic statement, and the incongruity of the remark with the prior context. In Experiment 1, participants were asked to rate the intent of the speaker as to how ironic, mocking, polite, and funny they intended their remark to be. In Experiment 2, participants read the scenarios as their eye movements were tracked. Results showed that age of the speaker had an impact on first pass reading times when statements were not familiar as ironic statements. Our younger adult participants did not appear to immediately activate a nonliteral interpretation of an ambiguous remark made by an older adult unless they had evidence from past experience that the remark is often used as an insult. However, ratings of the ironic intent of the statements were unaffected by speaker age; the age of the speaker affects the ease of interpretation but not the final outcome. The results are consistent with constraint-based theories of sentence comprehension.

Keywords: irony comprehension, nonliteral language, eye tracking, age effects

Public Interest Statement

Sometimes when a person says “You’re a great friend!” they really mean that, but other times they mean just the opposite. This study found that knowledge about the speaker, and in particular their age, affects the ease with which young adult participants interpret such remarks when reading, especially when other cues to intent are weak. These findings are important for understanding intergenerational communication.

The effect of speaker age on the perception of ironic insults

This study is fitting for a special issue of CJEP to honour the career of Albert Katz for two reasons. One reason is because it was inspired by a paper that he published with Penny Pexman and Todd Ferretti entitled *Discourse factors that influence online reading of metaphor and irony* (Pexman, Ferretti, & Katz, 2000). Metaphor and irony are interesting forms of language because they involve statements for which the intended meaning differs from what is actually said. As an example, “children are precious gems” could be interpreted as a metaphor, meaning that children are valuable, or could be intended sarcastically to convey that children are a burden. The authors examined cues that individuals might use to decide whether such a statement should be interpreted metaphorically or ironically, and in particular, whether these cues are used online when reading. Of interest to us was that one of the cues they examined was a characteristic of the person making the utterance- their occupation. Previous research by Katz and Pexman (1997) had shown that when statements such as “children are precious gems” were preceded by a context that mentioned a speaker with a high-irony occupation, such as a comedian, ratings of sarcasm were higher than when the speaker had a high-metaphor occupation (e.g., clergyman). In the online study, which used a word-by-word moving window reading paradigm and similar materials, contexts that mentioned a character with a high-irony occupation produced faster reading times on the last word of the target statement and longer reading times just after the target statement when the context invited a sarcastically ironic reading than when it invited a nonironic meaning. The authors argued that this finding provides evidence that the social identity of a character is a cue that readers use immediately to constrain their interpretation of an ambiguous statement (see Pexman, 2005, for a review), and is consistent with interactive models of language comprehension in which all available and relevant information is continuously integrated to compute the best interpretation (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994). Our study sought to extend this finding to a different aspect of social identity, and focused on the interpretation of verbal irony. As

evident in the example above, verbal irony is a nonliteral utterance that can be used to highlight a discrepancy between expected or desired outcomes and reality (Pexman, 2008).

The second reason why this study is fitting for this special issue is because we chose to examine speaker age as a cue when interpreting nonliteral language. As Albert Katz has been a faculty member at Western for over 40 years, and he is now in his senior years, this variable may be of personal interest to him. There is a stereotype that younger people, particularly teenagers, are more likely to use verbal irony than older people. If younger people do not expect older individuals to use verbal irony, particularly ironic insults, then they may have more difficulty recognizing irony in statements made by older individuals than by their peers. Howman and Filik (2020) provided evidence in support of this stereotype; older (65 years+) participants in their study reported using sarcasm significantly less often than young adult participants. Furthermore, Phillips, Allen, Bull, Hering, Kliegel, and Channon (2015) found age effects in the comprehension of verbal irony. They presented ironic and literal stories to participants ranging in age from 18 to 86 in a sentence-by-sentence manner. The last sentence of the story was a speaker uttering either an ironic or literal statement. For example, one ironic story involved a woman taking her friend Jean to a play that turned out not to be very good, and Jean says “That was a fantastic play you took me to see!”. After each story, the participant was asked a multiple-choice question to assess whether they perceived the statement literally or ironically. The researchers found that older adults had poorer comprehension of sarcastic remarks compared with younger and middle-aged adults, while there was no difference between the groups’ performance for the literal remarks. This difficulty of older adults in understanding verbal irony may account for why they report using it less frequently than younger adults. If they do not often make ironic comments, older individuals may be misunderstood by young adults when they do use this type of language.

Character cues and verbal irony: Rating tasks

The initial work on the impact of a speaker’s occupation on the comprehension of nonliteral language used statements that could be interpreted metaphorically or ironically but are unlikely to be

interpreted literally. Subsequent research by Pexman and Olinek (2002) investigated whether a speaker's occupation influenced the interpretation of remarks such as "you are a wonderful friend" that could be interpreted either as a literal or as an ironic remark. They gave participants passages that consisted of two context sentences followed by a remark, and asked them to rate whether the speaker was being sarcastic, polite, or mocking. Passages varied in whether the speaker had an occupation that was rated as having a high or low likelihood of using sarcasm, and in the positivity of the context and the positivity of the remark. Sarcasm ratings were higher with greater incongruity between the context and remark (see also Ivanko & Pexman, 2003), but were not influenced by speaker occupation. Mocking ratings were, however, higher for speakers with high compared to low sarcasm occupations, particularly in weak contexts. In a second experiment, neutral contexts were used, and this time they found that sarcasm ratings were significantly higher for remarks made by speakers from high sarcasm occupations than from low sarcasm occupations, suggesting that readers use expectations about the likelihood of a person using sarcasm as a cue to assess speaker intention, particularly when other cues to intent are absent.

Another characteristic of a speaker that affects sarcasm perception is gender, with ratings of sarcasm being higher when produced by male speakers than by female speakers (Katz, Piasecka, & Toplak, 2001), and sarcastic comments rated as more likely to be produced by a male than a female (Colston & Lee, 2004). Pexman and colleagues (Pexman, Whalen, & Green, 2010; Pexman & Zvaigzne, 2004) explored a further characteristic of speakers, the closeness of their relationship, and found that it did not affect ratings of irony in ironic criticisms but did have an impact on some ratings of the pragmatic functions of irony, such as humour. These rating studies provide evidence that speaker characteristics affect the reader's perception of a character's communicative intent when they make a potentially ironic remark.

Verbal irony and online reading tasks

There are few studies of effects of speaker characteristics on verbal irony interpretation that have

used online reading tasks. Such studies are informative regarding whether readers use cues about speakers immediately to constrain their interpretation of ambiguous remarks such as “you are a wonderful friend”. Katz et al. (2001; see Katz et al., 2004) reported that in a word-by-word reading study, reading times on the last words of sarcastic statements were longer when the statement was made by a female character than a male character, particularly when the comment was directed at another female. Also in a word-by-word reading task with ironic criticisms, Pexman et al. (2010) found some impact of the closeness of the relationship between the speakers in the region right after the remark. Regel, Coulson, and Gunter (2010) manipulated the proportion of ironic comments made by two characters in short passages and observed a congruency effect in the P200 of event-related potential (ERP) data, that is, participants produced a larger P200 for the ironic speaker on ironic remarks than on literal remarks and the reverse for the nonironic speaker.

The methodologies in the aforementioned studies involved presenting stimuli one word at a time (for the whole passage in the moving window studies and just for the remark in the ERP studies), which is not like natural reading. There have been a number of studies of verbal irony that have used eye tracking, and they provide evidence that differences can be detected in eye movement measures between remarks that are preceded by contexts that invite literal vs. nonliteral interpretations. Furthermore, they provide information as to the measures and conditions in which such effects are likely to be observed. Kaakinen and colleagues (Kaakinen, Olkonieni, Kinnari, & Hyona, 2014; Olkonieni, Ranta, & Kaakinen, 2016) found that participants were more likely to immediately reread remarks during first pass reading when they followed a context that invited an ironic compared to a literal interpretation and were more likely to look back to the former than the latter. In research by Filik and colleagues (Filik, Leuthold, Wallington & Page, 2014; Filik & Moxey, 2010; Turcan & Filik, 2016), longer reading times were observed on remarks in an ironic than in a nonironic context, primarily when ironic statements were unfamiliar as ironic statements. In the Turcan and Filik study there was no effect of the degree of incongruity in the context on the irony effect (in contrast to Ivanko

& Pexman, 2003, who found some evidence of an incongruity effect on the final word of a remark in a word-by-word reading study). Filik, Howman, Ralph-Nearman, and Giora (2018) further demonstrated that remarks that were placed in a context that invited an ironic insult interpretation had longer reading times than the same remark in a literal context, particularly when it was an affirmative rather than negative statement (e.g., ...he **is/is not** the most popular hairdresser). Eye tracking, therefore, appears to be a sensitive measure of processing that may be able to detect subtle influences on the interpretation of nonliteral language as it is being read under fairly natural conditions.

The current study

We examined whether the age of speakers is information that is used by readers in comprehending ambiguous remarks such as “You are so helpful!”, which could be interpreted either literally as compliments or nonliterally as ironic insults. Experiment 1 is a rating study and Experiment 2 is an eye tracking study. Context sentences were written to invite either a literal or nonliteral interpretation of the subsequent remark. The age of the speaker was manipulated in the context sentence by using names that are more typical of older individuals (e.g., Stanley, Edna) or more typical of younger individuals (e.g., Braydon, Alexa). That is, each remark was preceded by one of four types of context sentence. All participants were university-aged students. Of interest was whether an interaction between Context Type and Speaker Age would be observed, and whether any such effects would be modulated by the familiarity of the remark or the incongruity of the remark with the context.

Experiment 1

The aim was to determine whether the age of the speaker influences the perception of the communicative goal of the speaker. Participants rated the extent to which speakers were attempting to be ironic, mocking, funny, and polite.

Method

Participants. The participants were 48 undergraduate students (M age = 18.9 years, SD = 1.3). All indicated that English was their first and best-known language, and was the language they were

exposed to almost exclusively (98.5% of the day). Participants earned course credit.

Materials. There were 120 experimental scenarios, each of which had four versions. Scenarios were either adapted from those found in the literature or were created by the authors. All were two sentences in length. The first sentence was the context and the second was a remark. The remarks were identical for all the members of a quadruple. They were all four words in length, followed by an exclamation point, the word “said”, and the speaker’s name (e.g., “You are so helpful!” said Brenda). Remarks were in the affirmative and were positive except for two (e.g., “That is a shame!”). The remarks were directed at people in 89 scenarios and at situations in 31 scenarios. Context sentences either invited an ironic interpretation of the speaker’s remark (e.g., Brenda came home to find her brother hadn’t taken out the garbage like she’d asked him to.) or invited a literal interpretation (e.g., Brenda came home to find her brother had taken out the garbage like she’d asked him to.). The ironic interpretation was an insult or criticism. All contexts were written in past tense and were age-neutral, with the characters doing activities like painting, housework, and getting their hair cut. In 45 context sentences only the speaker’s name was mentioned and in the remaining 75 context sentences a second speaker in the same age category was mentioned. Additionally, the wording and length of the context sentences were matched as closely as possible across conditions.

For each quadruple, the character names in one ironic context and one nonironic context were typical of older individuals, and the character names in the other two context sentences were typical of younger individuals. Character names were chosen using names from the USA Social Security Administration (<https://www.ssa.gov/oact/babynames>). Names in the older condition were popular baby names from the 1900s to the 1950s (e.g., Gilbert, Wilma) and names in the younger condition were popular baby names from the 1990s to the present (e.g., Brayden, Alexa). Speaker names in the older condition were matched with speaker names in the younger condition based on gender and name length. Each name was entered into a name age calculator, based on U.S. social security data, to determine the age range and median age of people with each name (<http://rthiever.github.io/name-age->

calculator). The older speaker names had an average median age of 63 years and an average age range from 51 to 72 years, while the younger speaker names had an average median age of 13 years, and an average age range from eight to 20 years at the time of the study. The additional names that appeared in 75 context sentences were chosen using the same database.

Four counterbalanced lists of 120 stimuli were created such that each list had one member of the stimulus quadruple, and each list had an equal number of stimuli in each condition. The lists were then divided in half, to create eight lists of 60 items each. These lists were printed with five stimuli per page in 12 page booklets. Each stimulus appeared on the left side, and to the right were four 7-point rating scales (mocking, polite, ironic, funny) with endpoints “not at all” and “very”. Instructions on the cover page asked participants to read each scenario and to rate what they thought the speaker in the second sentence intended by their comment. It was noted that an ironic statement is one in which the intended meaning is opposite to the stated meaning.

A language questionnaire was used to collect information about participants’ language history.

Procedure. Participants were tested in small groups with the experimenter present. They first completed the scenario rating task and then the language questionnaire. Six participants completed each of the eight booklet versions.

To collect information on the familiarity of the remarks as ironic statements and on the incongruity of the contexts with the remarks, two additional sets of ratings of the experimental stimuli were collected from individuals who did not take part in the main study, and who were fluent speakers of English and similar in age. One set presented the 120 remarks and asked 20 participants to rate them on 7-point scales first for positivity and then for familiarity with their use as an ironic or sarcastic statement. Seventy filler remarks that were neutral or negative were mixed in with the experimental items in the first section. The mean familiarity of the remarks as ironic statements was 3.30 ($SD = 0.58$) and the mean positivity of the remarks was 5.53 ($SD = 0.53$). The second set presented the context sentences only and asked 80 participants to rate them for positivity, again on a 7-point scale. For the

latter ratings, four counterbalanced lists were created, such that each list had one member of a stimulus quadruple, and each list had an equal number of stimuli in each condition. Twenty participants were given each list. The mean positivity of the context statements for each group were: young-literal ($M = 4.46$, $SD = 0.84$), young-nonliteral ($M = 1.64$, $SD = 0.75$), old-literal ($M = 4.42$, $SD = 0.86$), old-nonliteral ($M = 1.65$, $SD = 0.72$).

Results

The Ordinal package in R (Christensen, 2015) was used to analyze the ratings data. Cumulative link mixed models (CLMMs) were run, fitted with the Laplace approximation. Fixed factors included two categorical variables: Context Type (literal, nonliteral) and Speaker Age (young, old). Mean ratings are presented in Table 1. Prior to running the models, R-default treatment contrasts for these variables were changed to sum-to-zero contrasts. Two continuous variables were also included: familiarity of the remark as an ironic statement and the degree of incongruity between the context and remark. The incongruity score was calculated by subtracting the positivity rating of the context from the positivity rating of the remark. The score for the nonliteral condition was assigned to its literal mate. These continuous variables were centred and scaled. Random effects included Participant (random intercept and random slope adjustments for Context Type and Age) and Item (random intercept). Specifically, models were: $\text{Rating} \sim \text{Context_Type} * \text{Speaker_Age} * \text{scale}(\text{Familiarity}) * \text{scale}(\text{Incongruity}) + (1 + \text{Context_Type} + \text{Age} \mid \text{Participant}) + (1 \mid \text{Item})$. Full model outputs are presented online as Supplementary Materials.

There was a main effect of Context Type on ratings of the speaker's intent to be ironic, $z = -19.66$, mocking, $z = -17.44$, funny, $z = -13.17$, and polite, $z = 14.92$, all $ps < .001$. The speakers in the nonliteral versions of the scenarios were rated as intending to be more ironic, mocking, and funny, and less polite than speakers in the literal versions. None of the effects of Speaker Age or interactions of Context Type and Speaker Age were significant. There were significant interactions of Context Type and Incongruity on ratings of intent to be ironic ($z = -5.23$, $p < .001$), mocking ($z = -5.70$, $p < .001$), and

polite ($z = 7.23, p < .001$). As well, there were significant interactions of Context Type and Familiarity on ratings of intent to be mocking, $z = 2.67, p < .01$, funny, $z = 4.27, p < .001$, and polite, $z = -3.54, p < .001$, but familiarity did not interact with Context Type for ratings of intent to be ironic ($z = 1.06$).

Discussion

The ratings data provide evidence that the literal and nonliteral versions of the scenarios were perceived to differ in how ironic, mocking, funny, and polite the speaker intended their remark to be. Of particular relevance here, remarks produced by speakers with older names were perceived as intending to be just as ironic as speakers with younger names, regardless of the familiarity of the remark as an ironic statement or degree of incongruity between the ironic context and the remark. The ratings data are untimed responses. Experiment 2 monitored eye movements as participants read to examine whether age of the speaker influences the ease with which ironic statements are understood.

Experiment 2

Method

Participants. The participants were 40 undergraduate students (M age = 20.9 years, $SD = 1.7$). All indicated that English was their first and best-known language, and was the language they were exposed to almost exclusively (98.5% of the day). Participants earned \$15.

Materials. The experimental stimuli were the same as in Experiment 1. The format of the second sentence in each scenario (the remark, followed by “said” and the speaker’s name) had been chosen to avoid having the remark region at the end of the sentence where it might be skipped or subject to sentence or scenario wrap-up effects. In addition, three practice and 120 filler stimuli were developed to disguise the purpose of the study. Names used in the filler stimuli were difficult to classify as belonging to an older or younger person. Filler stimuli were two sentences long and all were non-ironic. Four counterbalanced lists were created such that each list had one member of a stimulus quadruple, and each list had an equal number of stimuli in each of the four conditions. The same practice and filler stimuli appeared on all lists. Each list had a different random order of the stimuli.

Yes/no comprehension questions were created for one quarter of both the experimental and filler stimuli to assess whether participants were reading for meaning. For the experimental stimuli, these questions were based on the contextual sentence to avoid drawing attention to the speaker remark.

To determine whether participants perceived the speaker names as typically belonging to older or younger people, we created a checklist in which participants were asked to judge whether the names belonged to a person older or younger than 40 years of age. The checklist included all speaker names used in the experimental stimuli. The language questionnaire was also given.

Procedure. Eye movement data were acquired using an Eyelink 1000 eye-tracking system (SR Research, Ontario, Canada) with a sampling rate of 1000 Hz. A chin and forehead rest were used to minimize head movements. Eye movements were recorded from the participant's right eye only. Stimuli were presented on a 21-inch Sony Trinitron monitor located 60 cm from the participant's eyes. Calibration used a nine-point grid. Participants were told that they would be shown short passages, and that they were to read them silently and press the "A" button on the game controller when they were finished reading. They were informed that after some trials they would be asked a comprehension question, and that they were to press the left and right triggers on the game controller for NO and YES answers, respectively. During each trial, a fixation point appeared on the centre of the screen to check calibration, followed by a black square, halfway down the left side of the screen. The passage appeared when participants fixated on the square. Passages were presented in 18-point black Courier New font on a light-grey background. After three practice trials, the participants read the 240 sentences on one of the lists. They were given a break halfway through the experiment, and were recalibrated as needed. After the eye-tracking task, participants filled out the language questionnaire and the name checklist.

Results

The mean accuracy on the comprehension questions was 93% ($SD = 4.3$), indicating that participants were reading for meaning. There were no data for 6.1% of trials due to blinks, track loss, or failure to read the second sentence. Furthermore, data were excluded from trials in which the

participant did not categorize the age of the speaker's name as we intended on the checklist administered after the eye-tracking task (13.5% of trials). Many measures can be derived from eye tracking data; here we present three. The region of primary interest was the speaker's remark in the second sentence (e.g., "I am so smart!"), which was exactly the same in all four experimental conditions. First pass (FP) is the sum of the fixations, starting from first landing on the region until leaving it. Data from this measure tell us whether the age of the speaker immediately influences reading times. Total time (TT) is the sum of all fixations made on the region and is informative regarding comprehension repair processes. We also examined first pass reading times on the remainder of the sentence (e.g., said Brenda), a measure we call spillover (SO). We excluded trials with reading times in the remark region (4 words) less than 200 ms or greater than 3 *SDs* from the mean (2.8% of trials), and for the spillover region (2 words), we excluded trials with first pass times less than 100 ms or greater than 3 *SDs* from the mean (4.9% of trials). Regressions Out of each region were examined but there were no effects of interest. Means are presented in Table 1.

Data were analyzed using generalized linear mixed models (GLMMs) in R (Baayen, 2008; Baayen, Davidson & Bates, 2008; R Development Core Team, 2017). These models are an alternative to linear mixed models, which assume a normally distributed dependent variable and possible data transformation to achieve that (see Lo & Andrews, 2015). GLMMs do not assume a normal distribution, but rather allow the user to specify a frequency distribution that fits skewed latency data (the Gamma distribution was used here). Lo and Andrews argue that this method is more appropriate for interpreting interactions than using linear mixed effects models with a data transformation. The lme4 package, version 1.1-13 (Bates, Mächler, Bolker, & Walker, 2015) was used. The models were fit by maximum likelihood with the Laplace approximation technique. For the main analyses, the variables were the same as in Experiment 1, except that Trial Order was also included as a control variable. Specifically, the model was $\text{Fixation} \sim \text{Context_Type} * \text{Speaker_Age} * \text{scale(Familiarity)} * \text{scale(Incongruity)} + \text{Order} + (1 + \text{Context_Type} + \text{Speaker_Age} | \text{Participant}) + (1 | \text{Item})$. The random

effects structure was the maximum that would converge. In some subsequent analyses, a simpler structure was used when necessary for convergence. We present t values for model estimates; values of $t > 1.96$ are significant. Full model outputs are in the Supplementary Materials.

There was a significant effect of Context Type (FP: $t = -3.57$; TT: $t = -5.53$; SO: $t = -2.63$).

Reading times were longer when they were preceded by a context that invited a nonliteral interpretation than by one that invited a literal interpretation. The interaction of Context Type and Age was not significant (FP: $t = -0.30$; TT: $t = -0.62$; SO: $t = -1.81$). However, the interaction of Context Type, Age, and Incongruity was significant in the remark region (FP: $t = 2.29$; TT: $t = 5.66$), and that of Context Type, Age, and Familiarity was significant in the spillover region (SO: $t = 3.59$). Furthermore, there was a significant interaction of Context Type, Speaker Age, Incongruity, and Familiarity (FP: $t = -2.39$; TT: $t = -2.97$; SO: $t = -2.15$). To understand this 4-way interaction, analyses were conducted separately on high and low familiarity remarks, based on a median split. In Filik et al. (2014), remark familiarity influenced the size of irony effects. For high familiarity remarks, there was an effect of Context Type (FP: $t = -2.23$; TT: $t = -2.16$; SO: $t = -2.60$), but no significant interactions. In contrast, for low familiarity remarks, there was a significant interaction of Context Type, Speaker Age, and Incongruity in the remark region (FP: $t = 4.30$; TT: $t = 5.38$), and a significant interaction of Context Type and Speaker Age in the spillover region ($t = -2.11$). See Figure 1. Speaker Age, therefore, had more impact when remarks were less familiar as ironic statements. When speakers were young, there was a significant effect of Context Type in the remark region (FP: $t = -2.67$; TT: $t = -3.90$). However, when speakers were older, there was no effect of Context Type (FP: $t = -0.93$; TT: $t = -1.11$; SO: $t = 0.37$), although in Total Time Context Type interacted with Incongruity (TT: $t = -3.35$). Note that Figure 1 shows a trend for faster remark reading times for older speakers with greater incongruity, which is not expected for statements in literal contexts because they are not incongruent (they were assigned their nonliteral mate's score). However, statements that were more incongruent were also more positive. Indeed, when Incongruity was replaced with Positivity ratings for the remark in the models for older

speakers, there was a robust effect of Positivity (FP: $t = -3.02$; TT: $t = -3.62$), perhaps reflecting an expectation by younger participants that older individuals speak positively.

As a nod to Albert Katz's study on gender (Katz et al., 2001), we examined whether the effect of Context Type was influenced by the gender of the speaker (74 females, 46 males). Caution is warranted in interpreting the findings because male and female speakers were not in matched contexts. The fixed effects were the same as the main analyses above, but with Speaker Gender instead of Speaker Age. Slopes by participants were not included as random effects so models would converge. The interaction of Context Type and Speaker Gender was significant (FP: $t = 3.63$; TT: $t = 2.05$; SO: $t = 2.60$), with a larger effect of Context Type for female speakers than male speakers, as in Katz et al.'s study. The interaction of Context Type, Speaker Gender, Incongruity, and Familiarity was also significant (FP: $t = -3.30$; TT: $t = -2.55$; SO: $t = 4.84$). As for Speaker Age, Speaker Gender effects were more evident for remarks that were less familiar as ironic statements.

Discussion

The eye tracking data provide evidence that stereotypes associated with the names of people—here concerning their probable age—immediately influence the interpretation of remarks that are ambiguous as to whether or not they were meant to be taken literally. This information particularly affected the comprehension of remarks that are less often used as ironic statements, that is, when the speaker's intent may be harder to discern. It is important to note, however, that in Experiment 1, where participants were asked in an untimed task to evaluate the intent of the speaker, speaker age did not affect the perception of ironic intent; it appears to affect instead the ease with which that interpretation is reached. The findings are consistent with previous studies that have observed effects of speaker characteristics, (e.g. their occupation), on verbal irony interpretation, but extend that work in showing that such characteristics can have an immediate impact on reading times in a natural reading task.

General Discussion

The broad theoretical perspective we adopt is an interactive view of language comprehension, such

as connectionist neural network models, in which different sources of information (called constraints) activate possible interpretations of a message that compete with one another over time (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994). These constraints operate immediately and in parallel, and their contribution to the activation of a particular interpretation depends on the strength of their association with that interpretation from past experience. Katz and Pexman (e.g., Katz, 2005; Katz et al., 2004; Pexman, 2008; Pexman et al., 2000) claim that such models are well-suited to account for the interpretation of verbal irony. The constraints that contribute to the activation of a particular interpretation of a remark (e.g., a literal compliment, an ironic criticism) include inferences about the speaker, characteristics of the situation (e.g., the incongruity of the remark with the prior context), and prior experience with ironic language (e.g., the familiarity of the remark as an ironic statement). Together with the utterance, these cues activate the possible interpretations, and through a process of constraint satisfaction the system eventually settles on an interpretation that is the best fit for the activated information. The duration of the competition depends on whether all constraints point to the same interpretation, which would result in only brief competition, or whether there is support for each of the possible interpretations. Given that much of language is intended literally (Gibbs, 2000, reported that irony was used on just 8% of turns in conversation between friends), it is likely that the system would take more time to settle on an ironic interpretation than a literal one, and would require fairly strong cues to do so. We shall use this constraint satisfaction view to explain our findings.

In general, remarks in Experiment 2 had longer first pass reading times when they were preceded by a context that invited a nonliteral interpretation than by a context that invited a literal interpretation. This finding suggests that participants did typically activate the nonliteral interpretations of the remarks, but that extra processing time was required to settle on such an interpretation compared to the literal interpretations. More specifically, the effect of context type was observed for remarks that are often used as ironic insults; in this case high familiarity with the remark as an ironic statement would strongly activate the ironic interpretation, and other weaker cues, such as speaker characteristics, would

have little additional impact. The effect of context type was also observed for remarks that are less often used as ironic insults when the speaker had a name associated with a younger person. With familiarity only weakly activating the

ironic interpretation, knowledge that speaker was young provided further activation of the ironic interpretation. However, when remarks were less familiar as ironic insults and the speaker had a name associated with an older person, there was little evidence that the nonliteral meaning was immediately activated. There was neither an effect of Context Type nor an interaction of Context Type and Incongruity in the first pass data. Our younger adult participants, therefore, seem not to immediately activate an ironic interpretation of an ambiguous remark made by an older adult unless they have evidence from past experience that the remark is often an insult. The older name cue may even have inhibited the ironic interpretation briefly, counteracting the cue of incongruity. The total time data provide evidence that the ironic interpretation was eventually activated, especially when the remark was quite incongruent with the context.

The results of an admittedly post hoc analysis of speaker gender were consistent with the view that information from character names serves as a constraint in the interpretation of ambiguous remarks, particularly when such remarks are not often used as ironic insults, with male names providing stronger activation of the ironic interpretation than female names. Further eye-tracking research that is designed specifically to look at the influence of speaker gender will need to confirm this finding. Another aspect of a speaker's name that may affect the ease with which readers interpret ambiguous remarks is whether the name appears to be that of someone for whom the target language is a second language, because nonliteral aspects of a second language are difficult to acquire.

We interpreted our findings using the constraint satisfaction view but they have implications for other theories of irony processing. According to the direct access model (Gibbs, 1994), the nonliteral meaning of an utterance is activated in parallel with the literal meaning if it is supported by the context. Here the remark should have been processed equally quickly in the two contexts, but it was processed

more slowly in the ironic context than in the nonironic context. An alternative is the graded salience hypothesis (e.g., Giora, 2003), in which context has a more limited role; rather the most salient (frequent) interpretation of an utterance is activated first. For familiar ironic remarks, both the literal and ironic meanings are salient and are activated in parallel, whereas for less familiar ironic remarks, the literal interpretation is activated first, and the ironic interpretation is activated once a mismatch between the context and the literal interpretation is detected (see Giora & Fein, 1999). This view predicts an interaction of Context Type and Familiarity, with a larger effect of Context Type for less familiar remarks. However, in first pass reading times there was an effect of Context Type but no interaction with Familiarity. Therefore the constraint satisfaction view best accounts for our data.

Our results have practical implications for the design of research on nonliteral language processing. They provide evidence that associations based on a character's name can have an impact on the results when a sensitive processing measure such as eye tracking is used, and therefore careful consideration should be given to the choice of names, and to matching names across critical conditions designed to investigate the influence of other variables on nonliteral language processing.

Conclusion

As noted, this study was inspired by one of Albert Katz's publications and is just one example of the influence that his work has had on the field of nonliteral language processing; others are in this special issue. Our results suggest that if Albert uses statements that he intends to be interpreted ironically, a younger person may initially have difficulty perceiving that intent unless there are other strong cues in the context. However, the theoretical perspective that he favours suggests that those individuals who know Albert well, and who therefore know the likelihood of him uttering an ironic insult, will be able to use this additional constraint to assist with their interpretation of his remarks.

References

- Baayen, R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48.
- Christensen, R. H. B. (2015). Ordinal: Regression models for ordinal data. R package version 2015.6-28 [Computer software]. Retrieved from <http://www.cran.r-project.org/packageordinal/>
- Colston, H. L. & Lee, S. Y. (2004). Gender differences in verbal irony use. *Metaphor and Symbol*, *19*, 289-306.
- Filik, R. Howman, H., Ralph-Nearman, C., & Giora, R. (2018). The role of defaultness and personality factors in sarcasm interpretation: Evidence from eye-tracking during reading. *Metaphor and Symbol*, *33*, 148-162.
- 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, & Cognition*, *40*, 811-828.
- Filik, R. & Moxley, L. M. (2010). The on-line processing of written irony. *Cognition*, *116*, 421-436.
- Fox, J., & Weisberg, S. (2011). *An {R} Companion to Applied Regression* (2nd edn.). Thousand Oaks, CA: Sage.
- Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol*, *15*, 5-27.
- Gibbs, R. W. (2004). Figurative thought and figurative language. In M. A. Gernsbacher (Ed), *Handbook of psycholinguistics* (pp. 411-446). San Diego: Academic Press.
- Giora, R. (2003). *On our mind: Salience, context, and figurative language*. Oxford: Oxford University Press.

- Giora, R., & Fein, O. (1999). Irony: Context and salience. *Metaphor and Symbol, 14*, 241-257.
- Howman, H. E., & Filik, R. (2020). The role of emoticons in sarcasm comprehension in younger and older adults: Evidence from an eye-tracking experiments. *Quarterly Journal of Experimental Psychology*. <https://journals.sagepub.com/doi/full/10.1177/1747021820922804>
- Ivanko, S. L. & Pexman, P. M. (2003). Context incongruity and irony processing. *Discourse Processes, 35*, 241-279.
- Kaakinen, J. K., Olkonieni, H., Kinnari, T., & Hyönä, J. (2014). Processing of written irony: An eye movement study. *Discourse Processes, 51*, 287-311.
- Katz, A. N. (2005). Discourse and sociocultural factors in understanding nonliteral language. In H. L. Colston & A. N. Katz (Eds.), *Figurative language comprehension: Social and cultural influences* (pp. 183-207). Mahwah, NJ: Erlbaum.
- Katz, A. N., Blasko, D. G., & Kazmerski, V. A. (2004). Social influences on sarcastic language processing. *Current Directions in Psychological Science, 13*, 186-189.
- Katz, A. N. & Pexman, P. M. (1997). Interpreting figurative statements: Speaker occupation can change metaphor to irony. *Metaphor and Symbol, 12*, 19-41.
- Katz, A. N., Piasecka, I., & Toplak, M. (2001, November). *Comprehending the sarcastic comments of males and females*. Poster presented at the 42nd annual meeting of the Psychonomic Society, Orlando, FL.
- Lo, S. & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology 6*:1171.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676-703.
- Olkonieni, 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, & Cognition, 42*, 433-450.

- Pexman, P. M. (2005). Social factors in the interpretation of verbal irony: The roles of speaker and listener characteristics. In H. L. Colston & A. N. Katz (Eds.), *Figurative language comprehension: Social and cultural influences* (pp. 209-232). Mahwah, NJ: Erlbaum.
- Pexman, P. M. (2008). It's fascinating research: The cognition of verbal irony. *Current Directions in Psychological Science, 17*, 286-290.
- Pexman, P. M., Ferretti, T. R., & Katz, A. N. (2000). Discourse factors that influence online reading of metaphor and irony. *Discourse Processes, 29*, 201-222.
- Pexman, P. M., & Olinek, K. M. (2002). Understanding irony: How do stereotypes cue speaker intent? *Journal of Language and Social Psychology, 21*, 245-274.
- Pexman, P. M., Whalen, J. M., & Green, J. J. (2010). Understanding verbal irony: Clues from interpretation of direct and indirect ironic remarks. *Discourse Processes, 47*, 237-261.
- Pexman, P. M., & Zvaigne, M. T. (2004). Does irony go better with friends? *Metaphor and Symbol, 19*, 143-163.
- Phillips, L. H., Allen, R., & Bull, R., Hering, A., Kliegel, M., & Channon, S. (2015). Older adults have difficulty in decoding sarcasm. *Developmental Psychology, 51*, 1840-1852.
- R Core Development Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Regel, S., Coulson, S. & Gunter, T. C. (2010). The communicative style of the speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research, 1311*, 121-135.
- Turcan, A. & Filik, R. (2016). An eye-tracking investigation of written sarcasm comprehension: The roles of familiarity and context. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 42*, 1867-1893.

Table 1
Mean Scenario Ratings from Experiment 1 and Mean Fixation Data from Experiment 2 (SEs in parentheses)

Measures	Literal				Nonliteral			
	Young		Old		Young		Old	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Ratings (/7)								
Ironic	1.52	(0.05)	1.49	(0.05)	6.14	(0.05)	6.07	(0.05)
Mocking	1.54	(0.06)	1.48	(0.04)	5.51	(0.07)	5.47	(0.07)
Funny	1.63	(0.06)	1.64	(0.05)	3.69	(0.07)	3.69	(0.07)
Polite	5.47	(0.06)	5.42	(0.06)	2.61	(0.06)	2.66	(0.06)
Eye Tracking								
<i>Remark Region</i>								
First Pass (ms)	618	(7.25)	630	(7.55)	649	(8.02)	648	(7.65)
Total Time (ms)	681	(8.67)	692	(8.90)	723	(9.44)	715	(9.03)
Regressions Out (proportion)	.10	(.01)	.09	(.01)	.08	(.01)	.09	(.01)
<i>Spillover Region</i>								
First Pass (ms)	366	(7.74)	382	(7.81)	383	(8.06)	393	(8.21)
Regressions Out (proportion)	0.40	(.02)	0.41	(.02)	.41	(.02)	.39	(.02)

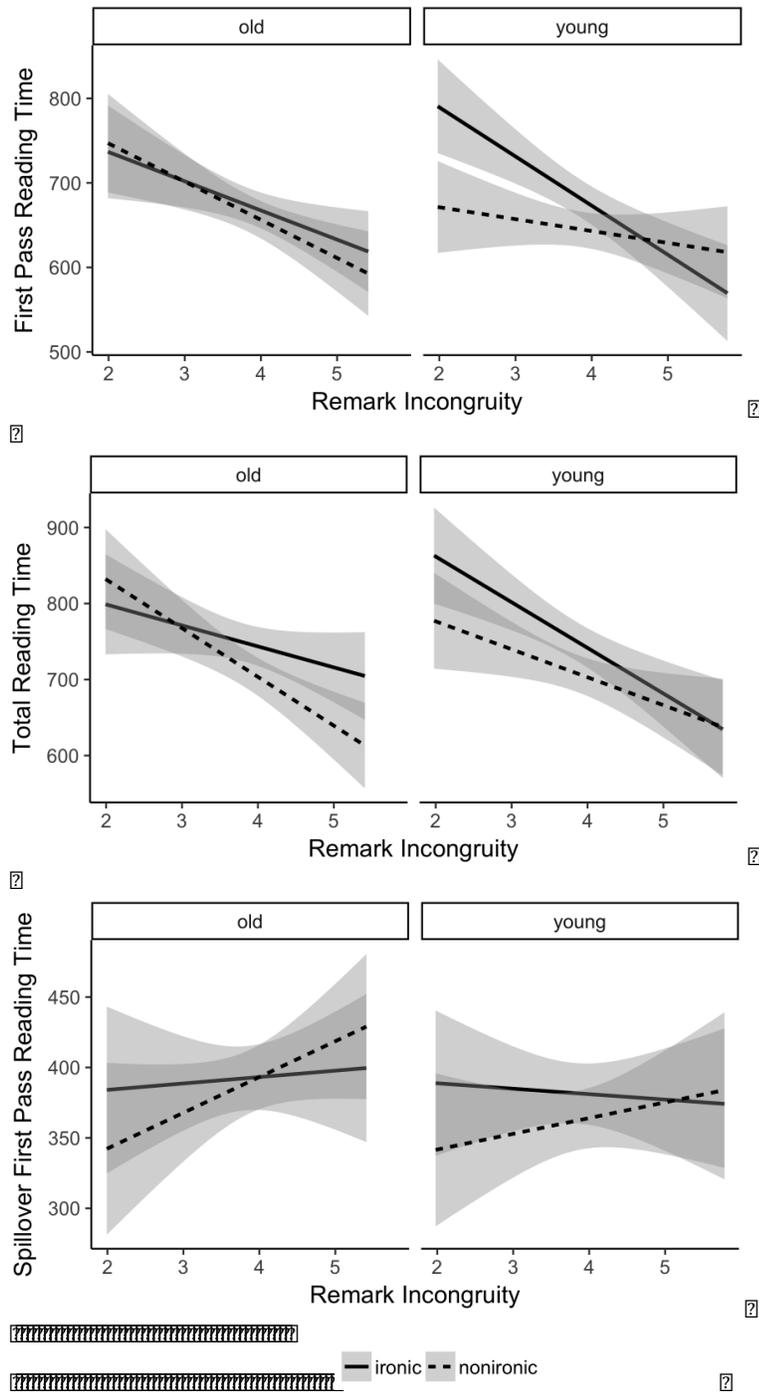


Figure 1. The impact of Speaker Age, Context Type and Incongruity on reading times (in ms) for remarks with low familiarity as an ironic statement. Observed values were plotted. Shaded areas represent 95% confidence intervals.