The cognitive and neural correlates of irony

Elizabeth Valles-Capetillo[1], Cristian D. Ibarra[1], Magda Giordano[1*]

[1] Departamento de Neurobiología Conductual y Cognitiva

Instituto de Neurobiología, Universidad Nacional Autónoma de México

Boulevard Juriquilla 3001, Querétaro, 76230, México.

email: elizabeth.valless@comunidad.unam.mx, cristian.ibarra@uabc.edu.mx, giordano@unam.mx*

*Corresponding author

Abstract

One of the most difficult forms of communication is irony (Wilson & Sperber, 1981), therefore it has been proposed that it can be a useful indicator of communication abilities (Caillies et al., 2014). Irony is a statement that transmits the opposite meaning of its literal counterpart (Grice, 1975). The cognitive process of mentalization, executive functions and holistic processing style may participate in irony detection. The brain areas that have been associated with detection of irony are the medial prefrontal cortex, inferior frontal gyrus (IFG), the posterior superior temporal gyrus, precuneus, the inferior parietal lobule, among others. The aim of this study was to analyze the cognitive processes involved in irony identification and its neural correlates in a sample of Mexican adults. A total of 45 participants performed an irony identification task during MRI acquisition, while cognitive processes were assessed with a psychometric battery. From the statements evaluated (i.e. ironic, literal, unrelated and white lies), irony was the hardest intention to detect, having the lowest percentage of correct classification and the slowest classification time. The cognitive process that correlated with irony detection were the tasks that evaluated ToM, inhibitory control and verbal abilities. The identification of ironic statements recruited brain areas associated with ToM, language, and inhibitory control networks.

keywords: irony, fMRI, Theory of Mind, inhibitory control, perceptual style

1. Introduction

Language is used to communicate ideas, thoughts, and feelings, through the combination of sounds, gestures, and symbols (Balconi, 2010). Nevertheless, understanding the sounds that are part of a language, or the meaning of the words, is not enough to understand the richness of language's meaning in social interactions (Martin & McDonald, 2003), because the same statement can have different communicative intentions and interpretations depending on the context. The ability to infer communicative intentions is a key process in human communication (Bambini et al., 2011), allowing people to distinguish between the possible alternative interpretations of the same statement (Bosco et al., 2017).

One of the most difficult forms of communication is irony (Wilson & Sperber, 1981), therefore it has been proposed that it can be a useful indicator of communication abilities (Caillies et al., 2014). According to the standard pragmatic view, irony is a statement that transmits the opposite meaning of its literal counterpart (Grice, 1975).

Pexman (2008) proposed in the *parallel constraint-satisfaction* framework that the correct selection of the intended meaning is supported by the cognitive process of mentalization, the ability to infer the mental states and emotions of others, and executive functions (e.g. inhibitory control). The author also suggested that the experience that the listener has with irony plays an important role (Pexman, 2008). Another cognitive process that has an important role in the detection of the intended

meaning is the holistic perceptual processing style, which integrates different types of information from the context, e.g. linguistic information, facial expressions, among others (Martin & McDonald, 2003).

Several pathologies are associated with deficits in communication abilities, for example, Autism Spectrum Disorder (ASD). The questionnaire Autism Spectrum Quotient (AQ), designed to evaluate symptoms associated with ASD (i.e. social skills, communication, attention to detail, attention switching, and imagination), can also detect individual differences in the neurotypical population. Differences have been found in relation to the type of profession of the participants (i.e. scientist, humanities, and engineering) (Baron-Cohen, Wheelwright, Skinner, et al., 2001). These results seem to indicate that the variability in social abilities, increases when the sample evaluated includes participants with different types of professions, especially those related to science, humanities, and engineering.

The brain areas that have been associated with irony identification are the medial prefrontal cortex (MPFC) (Shibata et al., 2010; Spotorno et al., 2012; H. Uchiyama et al., 2006; H. T. Uchiyama et al., 2012; Wakusawa et al., 2007; Wang et al., 2006). In the left hemisphere, the inferior frontal gyrus (IFG) (Spotorno et al., 2012), the inferior parietal lobe (IPL) (Rapp et al., 2010; Spotorno et al., 2012), pre-supplementary and motor area (Shibata et al., 2010), precentral gyrus (Rapp et al., 2010), and precuneus (Shibata et al., 2010). In the right hemisphere, the precentral gyrus (Rapp et al., 2010; Shibata et al., 2010), the dorsolateral prefrontal cortex (dlPFC) (Spotorno et al., 2012), the inferior frontal gyrus (Spotorno et al., 2012), the temporal superior gyrus (TSG) (Wang et al., 2006) and the cingulate gyrus (Rapp et

al., 2010; H. Uchiyama et al., 2006).

The results described above come from articles with great methodological variability concerning the native language of the participants, the modality of the stimuli, and the writing system, and none of them included a Spanish-speaking population. Despite this methodological variability, a meta-analysis showed that the dorsomedial prefrontal cortex (dmPFC), rostromedial prefrontal cortex (rmPFC), inferior frontal gyrus (IFG), and the posterior superior temporal gyrus (pSTG) presented consistent activity during irony identification (Reyes-Aguilar et al., 2018).

The aim of the present study was to analyze the cognitive processes involved in irony identification, such as mentalization, perceptual style, and executive functions, and to identify the neural correlates in a sample of Mexican adults. To increase the variability in the sample, according to previous literature (Baron-Cohen, et. al, 2001), people from different professional categories (i.e. science, humanities, and engineering) were selected.

2. Method

2.1. Participants

The experiment was completed by 45 participants (22 female) with a mean age of 26.69 ± 5.83. Of the 45 participants, 15 were studying or had a degree related to science (7 female), 12 non-scientific disciplines (e.g. humanities, 6 female), and 18 engineering (9 female). The participants reported no psychiatric or neurological history, were right-handed and signed a written informed consent to participate in the study. The handedness was measured with the Edinburgh handedness inventory

(Oldfield, 1971). The experimental protocol was approved by the Ethical Committee of the Neurobiology Institute of the Universidad Nacional Autónoma de México (#047.H.RM).

2.2. Behavioral paradigm

To assess irony identification was created a task for Mexican adults (Valles-Capetillo, Ibarra, D. C., Martinez, D., Giordano, M., submitted). Briefly, the task evaluated the identification of four types of intentions: irony, white lies, literal and unrelated. In total, 14 contexts per type of intention were created to shape the interpretation of the statement. Fourteen statements were created and used with all contexts. Thus, the same statement would have a different intention depending on the context in which it was presented. The task was created in Psychopy (3.0.1 version) (Peirce, 2007), and the stimuli were presented in the auditory modality without changes in prosody (see Figure 1).

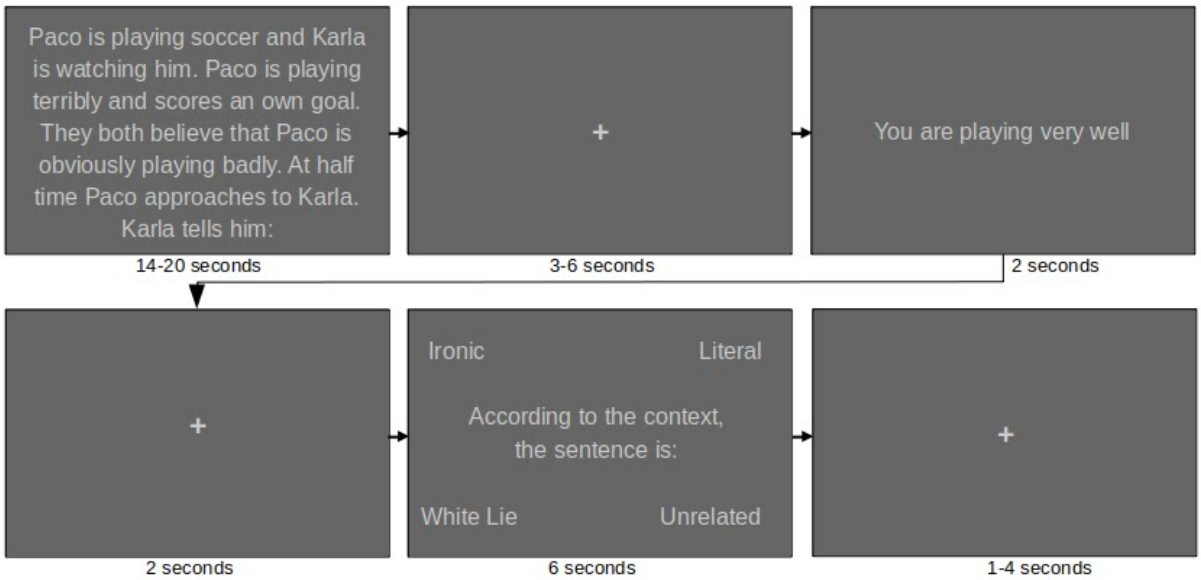

Figure 1.- Irony identification task. The context and statement were presented in audio without changes in prosody (text was not presented). The classification

screen was presented in text modality. Stimuli were presented in semi-random order (for further details see: Valles-Capetillo, Ibarra, D. C., Martinez, D., Giordano, M., submitted).

2.3. Psychometric tests

Cognitive variables were assessed using a battery of psychometric tests. Cognitive variables of interest were mentalization, perceptual style, and executive functions. Mentalization was assessed with the Short Story Task (SST) (Dodell-Feder et al., 2013) and the Reading the Mind in the Eyes Test (RMET) (Baron-Cohen, Wheelwright, Hill, et al., 2001). The perceptual style was assessed with the Perceptual Reasoning Scale of WAIS-IV (Wechsler, D., 2008) and the local-global test using Psychtoolbox-3 (Brainard, 1997). For executive functions the Psychology Experiment Building Language software (PEBL) was used (Mueller & Piper, 2014). The executive functions evaluated were: updating, which was assessed with the n-back test; planning, assessed with the Tower of London; go-no-go test was used to assess inhibitory control; and digit span, for working memory.

Additional cognitive variables were assessed with the verbal comprehension index (VCI) from the Wechsler Intelligence Scale for Adults, fourth edition (WAIS-IV; Wechsler, D., 2008). Verbal fluency was assessed with the BANFE (Flores et al., 2008). To assess the degree to which an individual is likely to use sarcasm, the sarcasm self-report scale was used (SSS) (Ivanko et al., 2004). The Autism Spectrum Quotient (AQ) was selected to assess the degree of social abilities, communication, attention to detail, attention switching, and imagination; this inventory has proven to detect individual differences in the neurotypical population

(Baron-Cohen, Wheelwright, Hill, et al., 2001). To evaluate psychological distress, the Symptom Check-List 90 was selected (SCL-90) (González-Santos et al., 2007). Finally, empathy was assessed with the Interpersonal Reactivity Index, which has four scales: fantasy, perspective-taking, empathic concern, and personal distress (Fernández et al., 2011).

The first group of the participants was evaluated before the sanitary restrictions due to the pandemic COVID-19. For these participants, the tests were completed in person, in one or two sessions, at the facilities of the Neurobiology Institute. The other group of participants was evaluated during the COVID-19 pandemic. To keep in-person contact to a minimum, some tests were applied via Zoom (v5.7.7; Zoom Video Communications, Inc., 2021). The tests that could only be applied in person, took place during the MRI session (i.e. the perceptual reasoning scale from WAIS-IV, local-global and, nback).

The results of the irony detection task were not normally distributed ($p < 0.01$), according to a Shapiro test. Levene's test showed that the variances between the statement types were not equal ($p < 0.01$). Therefore, non-parametric statistics were used. The percentages and standard deviation are presented for the classification accuracy, and the median and interquartile range (IQR) for the classification time. The comparison between statement types was performed using a Friedman test for repeated measures, followed by a Durbin-Conover pairwise comparison analysis with Bonferroni correction. Finally, a Spearman correction with FDR correction was performed. Statistical analysis were performed using R (version 3.6.3) and RStudio (version 1.1.4.4.7).

2.4. Functional Magnetic Resonance Imaging

For the MRI scanner session, the task presentation was optimized using OPTSEQ (Dale, 1999) and divided into three runs. Each run had a duration of 10 minutes. Participants were asked to infer the intention of the speaker when they heard the statement. Participants were asked to try to identify the intention of the statement when listening to it, according to the context. In order for the participants to become familiar with the task, before starting the task within the scanner, a test task was applied in which they classified 4 stimuli (one per statement type). The stimuli of the training task were different from those used in the task inside the scanner.

The MRI session was performed on a 3-Tesla GE MR750 scanner with a 32-channel head coil. The acquisition parameters were, for T1w: TE of 0.003 seconds, TR of 0.008, flip angle of 12°, and a 256 x 256 matrix; for the functional sequence: TE of 0.04, TR of 2 seconds, 90° flip angle, and 64 x 64 matrix. After the MRI session participants were asked to rate how ironic the stimuli were using a Likert-type scale from 1 to 7 (1 = least ironic, 7 = very ironic).

The dicom to nifti format conversion was performed using dcm2niix v1.0.20190902 (X. Li et al., 2016). The preprocessing of the MRI data was performed using the fmriprep pipeline (Esteban et al., 2019). Subsequent analyzes were performed with the FSL 5.0.9 program (Jenkinson et al., 2012). For the first (run), second (participant), and third (group) level analysis, a general linear model (GLM) analysis was performed using the feat tool (Woolrich et al., 2004).

In the first level analysis, the BOLD signal was analyzed during the presentation of the event of interest (i.e. statements). To analyze the events, a GLM was performed using the four statement types (i.e. irony, literal, unrelated, and white lies) as regressors; also, the first fixation cross was used as a regressor. For this analysis, a smoothing of 6 mm was applied and the significance threshold was adjusted at the cluster level ($z = 2.3$ and $p > 0.05$). A total of 22 contrasts were analyzed: 4 contrasts comparing each statement type with the fixation cross (e.g. irony > fixation cross), 4 comparing each statement type with the other three (e.g. irony > white lies + literal + no relation), 12 comparing pairs of statement type (e.g. irony > lie), and two comparing irony with two statement type (e.g. irony > lie+literal, irony > literal+no relation). In addition, 11 fmripri confounds regressors were used as physiological regressors: cerebrospinal fluid, white matter, global signal, dvars, frame wise displacement, and 6 motion regressors (translation and rotation of the x, y, z coordinates). For the second level analysis, a fixed effects GLM was performed. And for the third-level analysis, a mixed-effects GLM (Flame 1) per subject (45 subjects) was performed. For both analyses, the threshold parameters of the first level analysis were maintained ($z = 2.3$, $p > 0.05$).

3. Results

3.1. Psychometric tests

The lowest percentage of correct answers was found for the ironic statements (71.43 ± 28.16), followed by white lies (78.57 ± 20.04), literal (85.71 ± 13.52), and the highest was unrelated (92.86 ± 14.64). The highest classification time (seconds) was found for ironic statements (Md = 1.91, IQR = 1.01), followed by literal (Md =

1.72, IQR = 0.87), white lies (Md = 1.49, IQR = 1.20), and unrelated (Md = 1.31, IQR = 0.84).

The results show that there were significant differences in the percentage of correct answers ($X^2_{Friedman}(3) = 45.68$, $p < 0.001$) and classification time ($X^2_{Friedman}(3) = 31.81$, $p < 0.001$). In percentage of correct answers, significant differences were found between ironic and literal ($p < 0.05$), unrelated ($p < 0.001$) and white lies ($p < 0.001$); also, between the unrelated statements with literal ($p < 0.001$) and white lies ($p < 0.001$; see Figure 13a). In classification time, significant differences were found between the ironic with the unrelated ($p < 0.001$) and white lies ($p < 0.001$), also in literal with unrelated ($p < 0.001$) and white lies ($p < 0.001$; see Figure 2).

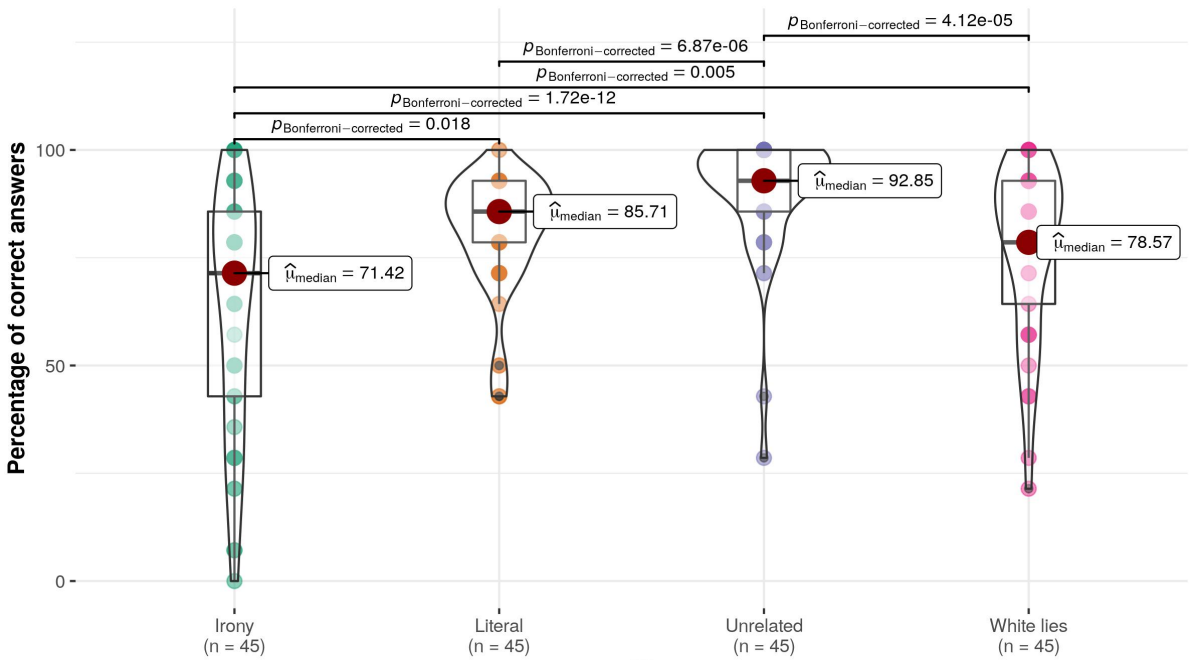

Figure 2.- Percentage of correct answers for each statement type. Ironic statements resulted in a significantly lower percentage of correct answers, followed by white lies, literal, and unrelated which had the highest percentage of correct answers. The plots show the density curves and the box plots show the median (red

circle), mean (thick line), interquartile range (rectangle), and the lower/upper adjacent values (black lines stretched from the rectangle), and scatter plot. The significant differences between statement types are reported with the p-value.

Regarding the cognitive process, the tests that evaluate ToM, the RMET test, showed a significant and positive correlation with the percentage of correct irony detection (rho = 0.35, p = 0.02*, p = 0.08); irony classification time correlated with the explicit inference from the SST (rho = 0.36, p = 0.03*, p = 0.14), however, these correlations did not survive after correction. The tests that evaluate perceptual processing did not correlate with accurate detection of irony. Regarding the executive functions, the go-no-go test, which evaluates inhibitory control, showed a negative correlation with irony classification time (rho = -0.35, p = 0.02*, p = 0.09). The linguistic abilities, evaluated with the VCI, showed a positive correlation with accurate irony detection (rho = 0.38, p > 0.05). Also, the fantasy subscale of the IRI was negatively correlated with irony classification time (rho = -0.46, p > 0.01). Finally, the AQ test showed a positive correlation with irony classification time (rho = 0.49, p > 0.01).

The classification time for literal (rho = 0.32, p = 0.03*, p = 0.11) and unrelated (rho = 0.30, p = 0.04*, p = 0.15) statements was positively correlated with explicit inference of the SST, similar to what we found in relation to irony detection, these correlations did not survive after correction. The literal statements showed no other significant correlations with cognitive measures. The percentage of correct answers of unrelated showed a positive correlation with the VCI of the WAIS IV (rho = 0.34, p > 0.05), and with with SSS (rho = 0.31, p = 0.035*. p = 0.13), however this last correlation did not survive after correction.

Similar to what was found with irony detection, the white lies statements presented a negative correlation between classification time and the go-no-go test, but the correlation did not survive after correction (rho = -0.36, p = 0.01*, p = 0.06). Also, the fantasy subscale of the IRI presented a negative correlation with white lies classification time (rho = -0.48, p > 0.01). The percentage of correct answers to white lies was positively correlated with the VCI of the WAIS IV (rho = 0.46, p > 0.01) and the GCI of the WAIS IV (rho = 0.43, p > 0.05). No correlations were observed with tests assessing ToM or perceptual processing style.

3.1. Functional Magnetic Reasoning Imaging

3.1.1. Contrast of one intention versus the others

The contrast of irony > literal+unrelated+white lies, showed significant activity in the precuneus, in Brodmann's area 31 (BA, MNI coordinates -4, -60, 24). The contrast of literal > irony+unrelated+white lies showed significant activity in three clusters: the right premotor area (BA 6, MNI coordinates 50, -6, 32); the left primary motor area (BA 4, MNI coordinates -38, -20, 38), extending to the primary sensory

area (BA 1, MNI coordinates -40, -26, 48); and the right putamen (MNI coordinates 26, 16, -6), extending to the caudate (MNI coordinates 14, 20, -4). The unrelated contrast > irony+literal+white lies showed significant activity in three clusters: the left premotor area (BA 6, MNI coordinates -42, -4, 44), extending to the anterior cingulate gyrus (BA 24, MNI coordinates -2, -4, 48); the left visual-motor area (BA 7, MNI coordinates -32, -46, 42), extending to the anterior supramarginal gyrus (BA 40, MNI coordinates -62, -32, 32); and the right primary motor area (BA 4, MNI coordinates 40, -18, 46), extending to the premotor area (BA 6, MNI coordinates 40, 0, 48). The white lies > irony+literal+unrelated contrast did not show significant activity (see Figure 3).

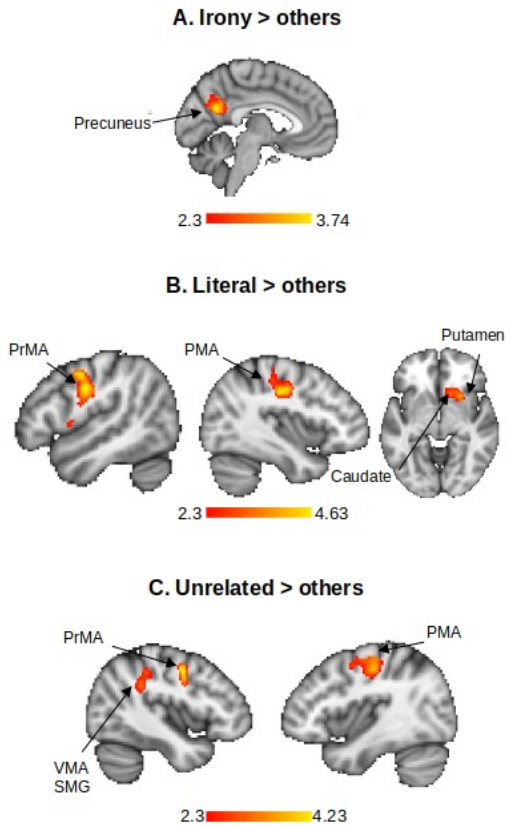

Figure 3.- Contrast of one intention versus the others. Panel A showed the contrast of irony > literal+unrelated+white lies, this contrast has significant activity in the Precuneus. Panel B showed the contrast literal > irony+unrelated+white lies, which presented significant activity in the right premotor area, left primary motor area (BA 4) and the right putamen. Panel C showed the contrast unrelated > irony+unrelated+white lies, which had significant activity in the left premotor area, the left visual-motor area, and the right primary motor area. The contrast white lies > irony+literal+unrelated did not show significant activity. Maps were thresholded at $z < 2.3$, $p < 0.05$. The cluster extension, MNI coordinates are listed in Supplementary Table 1.

3.1.2. Contrast by pairs

The irony > literal contrast showed significant activity in three clusters, the precuneus (MNI coordinates -2, -62, 44), extending to the left motor visual area (BA 7, MNI coordinates -6, -68, 34) and the right visual-motor area (BA 7, MNI coordinates 8, -64, 36); the left angular gyrus (BA 39, MNI coordinates -58, -50, 10), extending into the medial temporal gyrus (BA 31, MNI coordinates -58, -48, 6) and posterior superior temporal gyrus (BA 22); and the right IFG (pars triangularis, BA 45, MNI coordinates 50, 24, 14), extending into the dorsolateral prefrontal cortex (BA 9 and 46, MNI coordinates 50, 28, 26) and IFG (pars opercularis, BA 44, MNI coordinates 40, 18, 16) (see Figure 4 B).

Irony > unrelated contrast showed significant activation in two clusters, the right cuneal cortex (BA 18, MNI coordinates 6, -84, 16), extending to the visual associative area (BA 18, MNI coordinates 22, -90, 12), and lingual gyrus (BA 17, MNI

coordinates 2, -88, 0), and secondary visual (BA 18, MNI coordinates 16, -74, 0); the precuneus (BA 31, MNI coordinates -2, -60, 24), extending to the ventral cingulate gyrus (BA 23, MNI coordinates -8, -50, 24) and the lingual gyrus (BA 18; MNI coordinates -10, -52, 2) (see Figure 4 C).

The irony > white lies contrast showed significant activity in two clusters, the right supramarginal gyrus (BA 40, MNI coordinates 46, -50, 46), extending to the posterior superior temporal gyrus (BA 22), precuneus (BA 31, MNI coordinates -4, -62, 22), cingulate gyrus (BA 31, MNI coordinates -6, -54, 30) and the IFG (pars opercularis, BA 44, MNI coordinates 46, 14, 14); and the left primary sensory (BA 1, MNI coordinates -42, -20, 48), extending to BA 5 of the primary association (MNI coordinates -30, -34, 48). (see Figure 4 A).The contrast of irony > literal + white lies showed significant activity in three clusters, the precuneus (BA 31, MNI coordinates -4, -60, 24), extending to the motor visual area (BA 7, MNI coordinates -2, -62, 42); the IFG (pars triangularis, BA 45, MNI coordinates 50, 24, 14), extending to the IFG (pars opercularis, BA 44, MNI coordinates 38, 18, 16) and the dorsolateral prefrontal cortex (BA 9, MNI coordinates 48, 28, 28); and the angular gyrus (BA 39, MNI coordinates 44, -50, 44) (see Figure 4 D). Irony contrast > literal + unrelated showed activity in one cluster, the precuneus or change for IPCC (BA 31, MNI coordinates -2, -60, 24), extending to the posterior cingulate gyrus (BA 31, MNI coordinates -6, -54, 30) and the ventral posterior cingulate gyrus (BA 23, MNI coordinates -16, -56, 22) (see Figure 4 E).

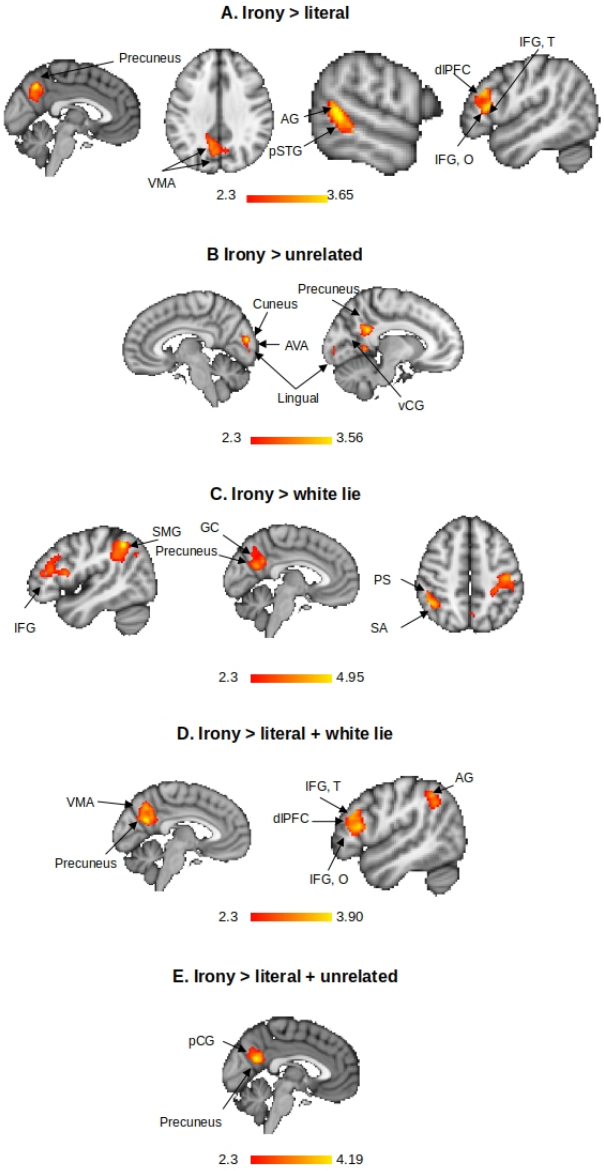

Figure 4.- Contrast by pairs. Panel A showed the contrast irony > literal; panel B, irony > unrelated; panel C, irony > white lies; panel D, irony > literal + unrelated; and panel E, irony > literal + unrelated. Maps were thresholded at $z < 2.3$, $p < 0.05$. The cluster extension, MNI coordinates are listed in Supplementary Table 2.

4. Discussion

The aim of this study was to evaluate the cognitive processes and the neural areas associated with the identification of ironic statements. To increase the variability in the cognitive and neural processing, based on what is currently known (Baron-Cohen, Wheelwright, Hill, et al., 2001), we selected a sample that had variability with respect to the academic background of the participants (i.e. humanities, engineering and sciences).

The variability of the academic background of the sample was positively reflected in the results of the cognitive evaluation. As expected (Bohrn et al., 2012; Pexman, 2008; Spotorno et al., 2012), ironic statements presented a positive correlation with tasks that assess ToM (i.e. RMET and SST). Also, the literal and unrelated statements showed a positive correlation with SST. Nevertheless, these

correlations did not survive after correction, even so, these results seem to show that the ability to infer the intention of a statement (i.e. ironic, literal or unrelated) within a social context is important. This relationship could be better appreciated with an even more varied sample.

Regarding the other processes evaluated, inhibitory control presented a negative correlation with the classification latency of ironic and white lies statements, however, these correlations did not survive the correction. Nevertheless, these associations show the importance of inhibitory control in the identification of irony and lies, it could be allowing the inhibition of a preponderant interpretation, for example, a direct interpretation, in favor of another, such as irony and white lies (Gaudreau et al., 2015). The literal and unrelated statements showed no relationship with inhibitory control, even before correction.

Verbal skills showed a positive correlation with the percentage of correct answers for irony, white lies and unrelated. These correlations suggest that ironic, white lies, and unrelated statements, being indirect messages require greater linguistic abilities to infer the message (Grice, 1975). This was also observed in the lack of correlation found in relation to the statements that transmits a direct message, i.e. literal.

The classification latency of ironic and white lies statements had a negative correlation with the fantasy scale of the IRI. This scale assesses the ability that people have to imagine themselves in fictitious situations (Fernández et al., 2011). This seems to show that people that are faster at detecting indirect statements, i.e.

ironic and white lies, have a greater tendency to imagine themselves in fictional situations. Activities that benefit the development of imagining oneself in fictitious situations, such as reading fiction, have been reported to benefit social communication skills (Kidd & Castano, 2013). The literal and unrelated statements showed no significant correlation with the fantasy scale of the IRI.

Finally, the AQ test showed a positive correlation with the classification latency of the ironic statements. This test assesses social skills, communication, attention to detail, attentional shift, and imagination; higher scores indicate difficulty in the areas assessed (Baron-Cohen, Wheelwright, Hill, et al., 2001). This correlation shows that when people take longer to detect ironic statements, have greater difficulties in the areas evaluated by the AQ, e.g. social skills and communication, reflecting the importance of these abilities in irony detection.

Regarding the brain regions, as expected (Reyes-Aguilar et al., 2018) the ironic statements presented significant activity in areas associated with the ToM network (i.e. precuneus and pSTG). Also, presented consistent activity in areas associated with the language network (i.e. IFG). Contrary to what was expected, the MPFC, associated with ToM, did not present significant activity. However, the precuneus, which has also been associated with ToM, presented consistent activity in the ironic statements.

The precuneus has been reported as part of the ToM network and has been associated with the processes of self-reflection, autobiographical memory, mental imagery, and construction of inferential meanings (Amodio & Frith, 2006; Cavanna &

Trimble, 2006; Frith & Frith, 2006; Gallagher et al., 2000; Mar et al., 2008; Schnell et al., 2016; Schurz et al., 2014). This region did not show activity in the contrasts in which literal, unrelated and white lies were compared with the other statements, suggesting its role specifically in irony detection.

The pSTG showed activity in the contrasts of irony > white lies, and irony > literal. This area has been associated with processes related to social communication, such as ToM (Amodio & Frith, 2006; Gallagher et al., 2000; Mar et al., 2008), action observation (Yang et al., 2015), extracting social information about a situation (Egorova et al., 2016; Redcay, 2008) and the integration of paralinguistic cues with contextual information (Egorova et al., 2016). The relationship between the process of integrating the information with the identification of irony has also been reported in studies with EEG and eye tracking. In EEG studies, an increase in the P600 potential has been reported, which has been associated with integration processes (Spotorno et al., 2013), understanding of the speaker's communicative behavior according to contextual information (Regel et al., 2010) and with the reanalysis of the meaning of the high-meaning statement (Weissman & Tanner, 2018). Finally, with eye tracking, the integration of information has been associated with an increase in regressions in the reading of ironic information, also, the increase in regressions has been associated with a reanalysis of information to decode the message (Filik et al., 2017; Filik & Moxey, 2010; Olkoniemi, Johander, et al., 2019; Olkoniemi, Strömberg, et al., 2019). In sum, the results obtained with other psychophysiological techniques seem to support the importance of information integration during irony identification.

The right IFG showed significant activity in the contrasts comparing irony with the other statement types: irony > literal (BA 45), irony > white lies (BA 44), and irony > literal + white lies (BA 44 and 45). This region has been associated with processes related to language, such as the identification of unusual or incongruous semantic fields (Jung-Beeman, 2005; Obert et al., 2016), identification of semantic violations, reanalysis of meanings (Bašnáková et al., 2014), and in the selection of the best interpretation according to the social context (Egorova et al., 2016). The importance of detecting anomalous linguistic information during irony detection has also been reported in EEG studies, in which an increase in the N400 potential has been observed (Cornejo et al., 2007; Regel et al., 2010; Regel & Gunter, 2017). This region did not show activity in the contrasts in which literal, unrelated and white lies were compared with the other statements.

The dorsolateral prefrontal cortex (dlPFC) of the right hemisphere showed significant activity in the contrasts of irony > literal, irony > lies, and irony > literal + white lies. This region has been associated with the processes of goal-oriented attention, task switching, planning, novelty seeking (Jones & Graff-Radford, 2021), problem-solving, and working memory (Jones & Graff-Radford, 2021; Mansouri et al., 2007). This region did not show activity in the contrasts in which literal, unrelated and white lies were compared with the other statements.

The primary motor cortex (BA 4) showed activity in the contrasts of literal > other, unrelated > other. This region has been associated with motor activity, such as the observation of motor tasks and integration of multiple sensory inputs, the inhibition of involuntary movement, and imagining motor sequences (Bhattacharjee

et al., 2021). This region did not present significant activity in relation to the ironic statements.

The premotor cortex (BA 6) presented activity in the contrasts of literal > others, unrelated > others. This region has been associated with the observation of goal-directed actions (Binkofski & Buccino, 2006; Chouinard & Paus, 2006; Morin & Grèzes, 2008), which allows recognition of actions carried out by others. Also, it has been associated with the mirror neuron system, involved in social interactions and attribution of intentions to others (Bhattacharjee et al., 2021; Morin & Grèzes, 2008), as well as the recognition of multimodal actions, even proposing that it could represent a precursor of language processing (Binkofski & Buccino, 2006). Similar to the primary motor cortex, this region did not show significant activity in relation to ironic statements.

In conclusion, the identification of irony seems to be supported by the cognitive process of ToM, inhibitory control, linguistic skills and the ability to imagine themselves in fictitious situations. The identification of the literal statements seems to be supported by ToM; the unrelated, by ToM and verbal skills; and white lies by inhibitory control, verbal skills and the ability to fantasize. The holistic processing style did not present a significant correlation with the statement type evaluated. For future studies, it would be interesting to explore a broader battery of psychometric tests that evaluate this process, to assess whether there is indeed no relationship between these variables, or whether it is due to the tests selected for this study. Also, a more varied sample could help to appreciate the association between the different statement types and the other cognitive process associated. Finally, regarding the

brain regions associated with irony identification, we found that areas associated with ToM network (i.e. precuneus, and pSTG), language network (i.e. IFG) and inhibitory control (i.e. dlPFC) are recruited. In comparison, the areas recruited when detecting literal and unrelated statements were mostly motor areas (e.g. primary motor area, premotor area, visual-motor area). These results indicate that these statements rely on different resources for their identification, resources such as the inhibition of involuntary movement, imagining motor sequences or observation of actions. With regard to white lies, no specific regions were recruited, possibly due to the fact that detection of white lies relies on similar brain areas as for the detection of irony. For future studies, a more varied sample, regarding the profession of the participants, could help to appreciate the variability in the identification of different statement types. Also, having a second group with difficulty in social communication (e.g. Autism Spectrum Disorder), would allow us to confirm the present findings.

Declaration of interest

Conflict of interest: none

Acknowledgements

Diola Elizabeth Valles Capetillo is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and has received CONACyT fellowship 755580. This study was supported by grants from DGAPA-PAPIIT (IN 203818) and CONACyT (Fronteras No. 225-2015 and scholarship to E. V. C. No. 755580). We greatly acknowledge their experimental guidance to Luis Concha Noyola and Juan Fernandez Ruiz. For technical support, we thank Erick Pasaye and Leopoldo González, the technical staff and the laboratory Lavis (UNAM-Juriquilla). For experimental support, we want to thank Abril Dominguez. For their careful review of the manuscript, we want to thank Jessica González Norris and Mike Jeziorski.

Author Contributions

Elizabeth Valles-Capetillo: designed research, performed research, contributed unpublished reagents/analytic tools, analyzed data and wrote the paper. **Cristian Ibarra**: performed research, contributed unpublished reagents/analytic tools, and wrote the paper. **Magda Giordano**: designed research, contributed unpublished reagents/analytic tools, wrote the paper, methodology, funding acquisition, project administration and supervision.

References

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268–277. https://doi.org/10.1038/nrn1884

Balconi, M. (2010). Biological Basis of Linguistic and Communicative Systems: From Neurolinguistics to Neuropragmatics. In M. Balconi (Ed.), *Neuropsychology of Communication* (pp. 3–27). Springer Milan. https://doi.org/10.1007/978-88-470-1584-5_1

Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P. M., & Pietrini, P. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bulletin*, *86*(3–4), 203–216. https://doi.org/10.1016/j.brainresbull.2011.07.015

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the Mind in the Eyes” Test Revised Version: A Study with Normal Adults, and Adults with Asperger Syndrome or High-functioning Autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241–251. https://doi.org/10.1017/S0021963001006643

Baron-Cohen, Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/a:1005653411471

Bašnáková, J., Weber, K., Petersson, K. M., Van Berkum, J., & Hagoort, P. (2014). Beyond the language given: The neural correlates of inferring speaker meaning. *Cerebral Cortex*, *24*(10), 2572–2578. https://doi.org/10.1093/cercor/bht112

Bhattacharjee, S., Kashyap, R., Abualait, T., Annabel Chen, S.-H., Yoo, W.-K., & Bashir, S. (2021). The Role of Primary Motor Cortex: More Than Movement Execution. *Journal of Motor Behavior*, *53*(2), 258–274. https://doi.org/10.1080/00222895.2020.1738992

Binkofski, F., & Buccino, G. (2006). The role of ventral premotor cortex in action execution and action understanding. *Journal of Physiology-Paris*, *99*(4), 396–405. https://doi.org/10.1016/j.jphysparis.2006.03.005

Bohrn, I. C., Altmann, U., & Jacobs, A. M. (2012). Looking at the brains behind figurative language--a quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia*, *50*(11), 2669–2683. https://doi.org/10.1016/j.neuropsychologia.2012.07.021

Bosco, F. M., Parola, A., Valentini, M. C., & Morese, R. (2017). Neural correlates underlying the comprehension of deceitful and ironic communicative intentions. *Cortex*, *94*, 73–86. https://doi.org/10.1016/j.cortex.2017.06.010

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4),

433–436.

Caillies, S., Bertot, V., Motte, J., Raynaud, C., & Abely, M. (2014). Social cognition in ADHD: Irony understanding and recursive theory of mind. *Research in Developmental Disabilities*, *35*(11), 3191–3198. https://doi.org/10.1016/j.ridd.2014.08.002

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, *129*(3), 564–583. https://doi.org/10.1093/brain/awl004

Chouinard, P. A., & Paus, T. (2006). The Primary Motor and Premotor Areas of the Human Cerebral Cortex. *The Neuroscientist*, *12*(2), 143–152. https://doi.org/10.1177/1073858405284255

Cornejo, C., Simonetti, F., Aldunate, N., Ibanez, A., Lopez, V., & Melloni, L. (2007). Electrophysiological Evidence of Different Interpretive Strategies in Irony Comprehension. *Journal of Psycholinguistic Research*, *36*, 411–430. https://doi.org/10.1007/s10936-007-9052-0

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, *8*(2–3), 109–114.

Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., & Hooker, C. I. (2013). Using fiction to assess mental state understanding: a new task for assessing theory of mind in adults. *PloS One*, *8*(11), e81279. https://doi.org/10.1371/journal.pone.0081279

Egorova, N., Shtyrov, Y., & Pulvermüller, F. (2016). Brain basis of communicative actions in language. *NeuroImage*, *125*, 857–867. https://doi.org/10.1016/j.neuroimage.2015.10.055

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I.,

Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Fernández, A. M., Dufey, M., & Kramp, U. (2011). Testing the psychometric properties of the Interpersonal Reactivity Index (IRI) in Chile: Empathy in a different cultural context. *European Journal of Psychological Assessment*, *27*(3), 179–185. https://doi.org/10.1027/1015-5759/a000065

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. https://doi.org/10.1016/j.jml.2016.09.006

Filik, R., & Moxey, L. M. (2010). The on-line processing of written irony. *Cognition*, *116*(3), 421–436. https://doi.org/10.1016/j.cognition.2010.06.005

Flores, J., Ostrosky, F., & Lozano, A. (2008). Batería de Funciones Frontales y Ejecutivas: Presentación. *Revista Neuropsicología, Neuropsiquiatría y Neurociencias*, *8*(1), 141–158.

Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/J.NEURON.2006.05.001

Gallagher, H., Happe, F., Brunswick, N., Fletcher, P., Frith, U., & Frith, C. (2000). Gallagher HL, Happé F, Brunswick N, Fletcher PC, Frith U, Frith CD. Reading the mind in cartoons and stories: an fMRI study of ‘theory of mind’ in verbal and nonverbal tasks. Neuropsychologia 38: 11-21. *Neuropsychologia*, *38*, 11–21. https://doi.org/10.1016/S0028-3932(99)00053-6

Gaudreau, G., Monetta, L., Macoir, J., Poulin, S., Laforce, R. J., & Hudon, C.

(2015). Mental State Inferences Abilities Contribution to Verbal Irony Comprehension in Older Adults with Mild Cognitive Impairment. *Behavioural Neurology*, *2015*, e685613. https://doi.org/10.1155/2015/685613

González-Santos, L., Mercadillo, R. E., Graff, A., Barrios, F. A., González-Santos, L., Mercadillo, R. E., Graff, A., & Barrios, F. A. (2007). Versión computarizada para la aplicación del Listado de Síntomas 90 (SCL 90) y del Inventario de Temperamento y Carácter (ITC). *Salud mental*, *30*(4), 31–40.

Grice, H. P. (1975). Logic and Conversation. *Speech Acts*, 41–58. https://doi.org/10.1163/9789004368811_003

Ivanko, S. L., Pexman, P. M., & Olineck, K. M. (2004). How Sarcastic Are You?: Individual Differences and Verbal Irony. *Journal of Language and Social Psychology*, *23*(3), 244–271. https://doi.org/10.1177/0261927X04266809

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jones, D. T., & Graff-Radford, J. (2021). Executive Dysfunction and the Prefrontal Cortex. *Continuum (Minneapolis, Minn.)*, *27*(6), 1586–1601. https://doi.org/10.1212/CON.0000000000001009

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, *9*(11), 512–518. https://doi.org/10.1016/j.tics.2005.09.009

Kidd, D. C., & Castano, E. (2013). Reading Literary Fiction Improves Theory of Mind. *Science*, *342*(6156), 377–380. https://doi.org/10.1126/science.1239918

Mansouri, F. A., Buckley, M. J., & Tanaka, K. (2007). Mnemonic Function of

the Dorsolateral Prefrontal Cortex in Conflict-Induced Behavioral Adjustment. *Science*, *318*(5852), 987–990. https://doi.org/10.1126/science.1146384

Mar, R., Djikic, M., & Oatley, K. (2008). Effects of reading on knowledge, social abilities, and selfhood: Theory and empirical studies. *Directions in Empirical Literary Studies*, 127–137.

Martin, I., & McDonald, S. (2003). Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain and Language*, *85*(3), 451–466. https://doi.org/10.1016/S0093-934X(03)00070-1

Morin, O., & Grèzes, J. (2008). What is “mirror” in the premotor cortex? A review. *Neurophysiologie Clinique = Clinical Neurophysiology*, *38*(3), 189–195. https://doi.org/10.1016/j.neucli.2008.02.005

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods*, *222*, 250–259. https://doi.org/10.1016/j.jneumeth.2013.10.024

Obert, A., Gierski, F., Calmus, A., Flucher, A., Portefaix, C., Pierot, L., Kaladjian, A., & Caillies, S. (2016). Neural Correlates of Contrast and Humor: Processing Common Features of Verbal Irony. *PLOS ONE*, *11*(11), e0166704. https://doi.org/10.1371/journal.pone.0166704

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Olkoniemi, H., Johander, E., & Kaakinen, J. K. (2019). The role of look-backs in the processing of written sarcasm. *Memory & Cognition*, *47*(1), 87–105. https://doi.org/10.3758/s13421-018-0852-2

Olkoniemi, H., Strömberg, V., & Kaakinen, J. K. (2019). The ability to recognise emotions predicts the time-course of sarcasm processing: Evidence from eye movements. *Quarterly Journal of Experimental Psychology (2006)*, *72*(5), 1212–1223. https://doi.org/10.1177/1747021818807864

Peirce, J. W. (2007). PsychoPy--Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pexman, P. M. (2008). It's Fascinating Research. *Current Directions in Psychological Science*, *17*(4), 286–290. https://doi.org/10.1111/j.1467-8721.2008.00591.x

Rapp, A. M., Mutschler, D. E., Wild, B., Erb, M., Lengsfeld, I., Saur, R., & Grodd, W. (2010). Neural correlates of irony comprehension: The role of schizotypal personality traits. *Brain and Language*, *113*(1), 1–12. https://doi.org/10.1016/j.bandl.2009.11.007

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: Implications for the emergence of autism. *Neuroscience & Biobehavioral Reviews*, *32*(1), 123–142. https://doi.org/10.1016/j.neubiorev.2007.06.004

Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research*, *1311*, 121–135. https://doi.org/10.1016/j.brainres.2009.10.077

Regel, S., & Gunter, T. C. (2017). Don't Get Me Wrong: ERP Evidence from Cueing Communicative Intentions. *Frontiers in Psychology*, *8*, 1465. https://doi.org/10.3389/fpsyg.2017.01465

Reyes-Aguilar, A., Valles-Capetillo, E., & Giordano, M. (2018). A Quantitative Meta-analysis of Neuroimaging Studies of Pragmatic Language Comprehension: In Search of a Universal Neural Substrate. *Neuroscience*, *395*, 60–88. https://doi.org/10.1016/j.neuroscience.2018.10.043

Schnell, Z., Varga, E., Tényi, T., Simon, M., Hajnal, A., Járai, R., & Herold, R. (2016). Neuropragmatics and irony processing in schizophrenia – Possible neural correlates of the meta-module of pragmatic meaning construction. *Journal of Pragmatics*, *92*, 74–99. https://doi.org/10.1016/J.PRAGMA.2015.11.004

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/J.NEUBIOREV.2014.01.009

Shibata, M., Toyomura, A., Itoh, H., & ichi Abe, J. (2010). Neural substrates of irony comprehension: A functional MRI study. *Brain Research*, *1308*, 114–123. https://doi.org/10.1016/j.brainres.2009.10.030

Spotorno, N., Cheylus, A., Van Der Henst, J.-B., & Noveck, I. A. (2013). What's behind a P600? Integration operations during irony processing. *PloS One*, *8*(6), e66839. https://doi.org/10.1371/journal.pone.0066839

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J. B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, *63*(1), 25–39. https://doi.org/10.1016/j.neuroimage.2012.06.046

Uchiyama, H., Seki, A., Kageyama, H., Saito, D. N., Koeda, T., Ohno, K., & Sadato, N. (2006). Neural substrates of sarcasm: A functional magnetic-

resonance imaging study. *Brain Research*, *1124*(1), 100–110. https://doi.org/10.1016/j.brainres.2006.09.088

Uchiyama, H. T., Saito, D. N., Tanabe, H. C., Harada, T., Seki, A., Ohno, K., Koeda, T., & Sadato, N. (2012). Distinction between the literal and intended meanings of sentences: A functional magnetic resonance imaging study of metaphor and sarcasm. *Cortex*, *48*(5), 563–583. https://doi.org/10.1016/j.cortex.2011.01.004

Wakusawa, K., Sugiura, M., Sassa, Y., Jeong, H., Horie, K., Sato, S., Yokoyama, H., Tsuchiya, S., Inuma, K., & Kawashima, R. (2007). Comprehension of implicit meanings in social situations involving irony: A functional MRI study. *NeuroImage*, *37*(4), 1417–1426. https://doi.org/10.1016/j.neuroimage.2007.06.013

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Developmental changes in the neural basis of interpreting communicative intent. *Social Cognitive and Affective Neuroscience*, *1*(2), 107–121. https://doi.org/10.1093/scan/nsl018

Wechsler, D. (2007). WISC-IV: Escala de Inteligencia de Wechsler (2a ed.). Madrid: TEA

Weissman, B., & Tanner, D. (2018). A strong wink between verbal and emoji-based irony: How the brain processes ironic emojis during language comprehension. *PLOS ONE*, *13*(8), e0201727. https://doi.org/10.1371/journal.pone.0201727

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, *21*(4), 1732–1747. https://doi.org/10.1016/j.neuroimage.2003.12.023

Wilson, D., & Sperber, D. (1981). Irony and the use-mention distinction. *1981*. https://www.academia.edu/17030837/Irony_and_the_use_mention_distinction

Yang, D. Y.-J., Rosenblau, G., Keifer, C., & Pelphrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience & Biobehavioral Reviews*, *51*, 263–275. https://doi.org/10.1016/J.NEUBIOREV.2015.01.020