The role of visual cues in detecting irony

Beatrice GIUSTOLISI — *University of Milan - Bicocca*
Francesca PANZERI — *University of Milan - Bicocca*

Abstract. We present four studies that aimed at investigating the contribution of purely visual cues for the detection of irony. In Study 1-3, we presented, without any preceding context, remarks (criticisms and compliments) uttered with sincere and with ironic intent, in three modalities: in the V modality participants could read the comment and see speakers' facial expressions and bodily movements; in the A modality they could only hear the audio tracks of the uttered sentences; in the VA modality, both visual and auditory information were present. We found that purely visual cues were sufficient to discriminate the ironic intent of the speakers. In Study 4 we presented comments in the V modality, without showing the content of the remark: Accuracy in the detection of sarcasm dropped. We discuss that irony in Study 1-3 might have been recognized indirectly, by comparing the polarity of the remark with the polarity of the actors' attitude, and we interpret Study 4 data as casting some doubts on the idea that there exist visual cues that specifically convey the speaker's ironic intent.

Keywords: Irony recognition, Irony markers, Facial expressions, Ironic tone of voice, Ironic compliments.

1. Introduction

Ann is in a pub with her friend Leo. He is bringing two mugs of beer to the table, but he stumbles and spills the beer on Ann. She exclaims: "You did a great job!". Ann's comment is obviously ironic: Ann is showing her contemptuous attitude towards the idea that Leo did a great job (Wilson & Sperber, 2012), or towards whoever would be so silly to have this idea (Clark & Gerrig, 1984). For the communication to be successful, Leo needs to recognize Ann's communicative intent and thus realize that she is ironic and that therefore she thinks quite the opposite of what she said. Ironic forms of language constitute about 8% of conversational turns among friends (Gibbs, 2000) and, to avoid misunderstandings, ironists should alert their addressees that their utterances should not be taken at face value. Ironic remarks are always inappropriate relative to the situation, and thus the incongruity between the remark and the context can be viewed as a necessary ingredient (Attardo, 2000). Signals of irony, on the other hand, fall into two main categories: verbal, such as the choice of particular morphosyntactic constructions (Seto, 1998) and linguistic expressions (e.g., extreme adjectives, Kreuz & Roberts, 1995), or paralinguistic, involving modulations of the speech, particular facial expressions and gestures (Hancock, 2004). Prosodic signals include variations in the tone of voice (such as prolonged articulation, increased intensity and stress), as well as other acoustic features (e.g., nasalization). Kinesic cues consist primarily of facial expressions, such as smiles, eyebrow raising, or winks (see Burgers & van Mulken, 2017 for an overview).

A different but related question is whether the presence of these irony markers permits the recognition of ironic remarks. Several studies found that the incongruity between the context and the statement (Attardo's 'irony factor') constitutes the most powerful cue for the correct discrimination of irony (Deliens et al., 2018, a.o.). Even if it is not possible to identify prosodic features that uniquely characterize ironic remark, because they vary among types of irony

(Anolli, Ciceri & Infantino, 2002), and among languages (Cheang & Pell, 2009; Lœvenbruck et al., 2013; Scharrer & Christmann, 2011), many scholars often implicitly assume that the presence of the so-called ironic tone of voice permits the discrimination of ironic utterances. Some studies found that interlocutors can correctly detect ironic remarks relying on purely auditory information (Rockwell, 2000; Bryant & Fox Tree, 2002), at least in their own language (Cheang & Pell, 2011). The possible contribution of facial expressions, on the other hand, is under-studied.

To the best of our knowledge, only Deliens and colleagues (2018) empirically investigated the question of whether facial expressions constitute reliable cues for irony detection, even if this was not the primary goal of their research. In their first experiment, they asked 127 participants to evaluate on a 7-point scale how ironic a speaker was, relying mainly either on visual or auditory cues. Tested items were pronounced by professional actors and comprised literally positive statements (such as "Yes, I love tea") in a sincere and in an ironic interpretation, and literally negative remarks ("No, I hate tea") with a sincere intent only. To obtain items in which the visual cues were more salient, the actors were asked to pronounce the statements with a monotonous tone of voice, so as to suppress prosodic cues; analogously, to highlight the role of auditory cues, actors had to maintain a neutral facial expression, while pronouncing the statements with congruent (ironic, literal positive or literal negative) intonation. Deliens and colleagues found that participants could correctly discriminate between ironic and literal statements, even in the absence of preceding context, simply by focusing on the actors' facial expressions (with flat intonation), or on their tone of voice (with neutral expressions). Nevertheless, the prosodic contours and facial expressions that accompanied literal statements with positive content were judged to be more ironic than their literal negative counterparts. The authors thus concluded that, compared to contextual cues, the auditory and visual signals are more ambiguous as to the detection of the speaker's meaning.

We believe that Deliens and colleagues' study cannot properly establish the contribution of visual cues in the discrimination of irony. First of all, they did not isolate phonological and visual markers, since the actors were asked to pronounce the remarks with a flat intonation or to maintain a neutral expression, but this artifact condition does not isolate visual from auditory cues (and, moreover, flat intonation and blank face have been connected to irony expression, see Attardo et al., 2003). In addition to that, they only tested the recognition of ironic criticisms, that is literally positive statements that are ironically used to blame. Ironic compliments, that is literally negative remarks that can be uttered ironically to convey a reversed, and thus positive, meaning, constitute a rarer form of irony, and have been proven to be more difficult to grasp (Kreuz & Glucksberg, 1989; Kreuz & Link, 2002). Still, since the participants of Deliens and colleagues' first experiment rated literally positive statements as more ironic than literally negative comments, it is relevant to establish whether irony ratings are influenced not only by speakers' communicative intent (being sincere or ironic), but also by their attitude (making a compliment or a criticism).

Various proposals have been put forth to account for the asymmetry in the production and comprehension of ironic criticisms and ironic compliments. Focusing on the social functions of irony, Dews & Winner (1995) proposed the Tinge hypothesis, claiming that irony attenuates the perceived affective attitude of the speaker: ironic criticisms mute the aggressiveness of literal criticisms and ironic compliments are veined with some meanness compared to literal

compliments. Since speakers' attitudes can be recognized through their facial expressions (see Keltner et al., 2003 for a review) and their tone of voice (Truesdale & Pell, 2018; Fish, Rothermich & Pell, 2017), it is highly relevant to verify if and how the expression of the speakers' negative or positive attitude through their modulation of speech and facial expressions interacts with the recognition of their sincere or ironic intent. We thus designed a series of experiments that aimed at assessing the role of visual cues in the detection of ironic remarks, also in comparison to the contribution of auditory information, testing sincere and ironic utterances of sentences with positive and negative content.

2. The studies

The general goals of the four studies we conducted were to verify whether purely visual cues constitute reliable cues to detect ironic remarks, and whether their contribution is comparable to that one of purely acoustical cues. In Study 1 and 3, we presented pairs of the same remark pronounced by the same actor once with a sincere and once with an ironic intent, in three different modalities: with only visual cues (participants watched the videos with no audio), with only acoustical cues (participants heard the audio tracks of the remarks), and when all the information was present (videos with audio). Participants' task was to recognize the stimulus in which the actor was ironical. Since, as discussed in the introduction, ironic criticisms differ in many respects from ironic compliments, we tested pairs of literal compliments / ironic criticisms in Study 1, and pairs of literal criticisms / ironic compliments in Study 3. The other two studies were run to control for possible confounds: in Study 2 we checked whether the results obtained in the first study could be extended to more ecological situations, in which participants are presented with only one remark at a time. Study 4 investigated the possible contribution of the linguistic material in the recognition of irony, to establish whether participants rely on visual irony markers, or whether the detection of the ironic intent is only indirectly obtained.

In all the four studies we conducted, the critical items were pairs of the same remark (e.g., "The party was really funny!") uttered once sincerely and once ironically. Our goal was to obtain remarks that were pronounced in a spontaneous way, and to do that we used a Discourse Completion Task (Félix-Brasdefer, 2010). Four Italian speaking adults (two females) have been recruited on a voluntary basis. They participated in two different elicitation sessions: in each session, they were presented with written scenarios introducing a context, for instance the organization of a party or the online renting of a holiday house. After reading the context, the actors were asked to pronounce a final remark, using some specific words (e.g., "party" and "funny", or "ugly" and "house"), that they could combine in the way that sounded more natural to them. Actors were then videotaped while they were narrating the story (in their own words) and pronouncing the final remark. In one session, the scenario was congruent with the remark (i.e., the party was successful with all the guests having fun; the rented house was a run-down hovel), and thus a sincere interpretation of the remark was obtained (a literal compliment when both the scenario and the comment were positive, as in "The party was really funny!"; a literal criticism when they were both negative, as in "What an ugly house!"). In a second session, separated by an interval of two weeks at least, the scenarios had a reversed polarity (e.g., the planned party turned out to be a failure with extremely bored guests, and the rented house turned out to be a wonderful villa), and again the actors were asked to narrate the story, that had to end with exactly the same remark they chose to utter in the first session. This session

presented scenarios that were incongruent with the final remarks, which thus received an ironic interpretation – as ironic criticisms (when the scenario was negative and the remark positive) or as ironic compliments (when the context was positive and the comment negative). In each session, we presented 10 stories, that ended up with a remark that was literally positive in 5 cases, and literally negative in the other 5 scenarios. Since the 10 remarks were uttered by 4 actors in two different sessions (once ironically and once sincerely), we thus obtained a total of 80 remarks, 40 minimal pairs, that constituted 20 literal compliments and 20 ironic criticisms, and 20 literal criticisms and 20 ironic compliments. The material was then edited to obtain three versions: the original version with video and audio (VA), one with only the video (V), and one with only the audio (A).

Study 1, 3 and 4 were implemented on QualtricsXM, a platform to conduct online surveys. Participants were instructed to run the experiment on a computer screen in a quiet room and/or using earphones, to ensure that they could all watch the videos and hear the audio tracks in an optimal way. Study 2 was implemented on E-Prime3 and administered in the Psycholinguistics Laboratory of our Department under the supervision of the first author. All participants were recruited through the Sona System of our Department and received credits for their participation. The system ensured that the same individual could participate in only one of the four studies. The research project was approved by the Ethical committee of our Department, and informed consent was obtained from all participants.

2.1. Study 1

Our general aim was to verify whether purely visual cues constitute reliable signals to detect ironic remarks, and whether their contribution is comparable to that one of purely acoustical cues. Since, as discussed in the introduction, ironic criticisms are the most widespread form of irony, in this first study we tested only pairs of literal compliments / ironic criticisms. The material thus consisted of a total of 20 pairs of literally positive remarks, pronounced once ironically and once sincerely, presented in three blocks: in the first block, only the pairs of videos, without audio (V), were presented, but participants could read on the screen the remark that was pronounced; the second block consisted of pairs of audio tracks (A); in the third and final block participants could watch the videos with the audio (VA). The blocks were always presented in the same order (V, A, and VA); within each block the items were presented in randomized order. The participants' task was to indicate in which of the videos or audio tracks the actor was ironic. An example is given in Figure 1. Eighty-one Italian native speakers (61 females, 20 males, mean age = 23 years, *SD* = 5 years) took part in Study 1.

Il commento che viene fatto nei prossimi video è:

"QUESTA FESTA E' STATA DIVERTENTE"

Video 1:

Video 2:

In quale video l'attore è ironico?

○ Video 1

○ Video 2

Figure 1. Example of an item of the first block (V) of Study 1: on the top of the screen, participants could read "The remark that is pronounced in the following videos is 'The party was funny'"; the participants' task was to indicate in which video the actor was ironical.

In Study 1 it was our purpose to analyze if participants, presented with pairs of literal compliments / ironic criticisms, recognized more easily the ironic comment with visual or auditory cues, and if accuracy in these artifact conditions decreased when all the information was present. In other words, the VA modality served as a baseline, and we thus first checked whether the ironic remarks could be correctly detected in this last block, since the effective recognizability of the pairs of sincere/ironic remarks is the *sine qua non* condition. We removed two items in which overall VA accuracy was less than 51% correct. Subsequent analyses were then run on 18 items. Since the experiment was run online, we also wanted to ensure that participants were not responding at random. Again, we focused only on the third VA block, and a binomial calculation indicated that participants responding to more than 12/18 items correctly (67%) were not performing the task at random ($p < 0.05$). Seven participants (8%), with a mean accuracy in VA items between 50% and 67% were thus removed. The final analysis was therefore conducted on 74 participants. As depicted in Figure 2, accuracy was

higher in the VA modality (91%) compared to the other modalities, and slightly higher in the V modality (84%) compared to the A modality (79%).
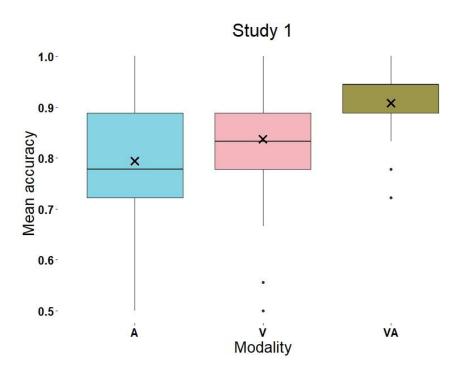

Figure 2. Boxplots representing accuracy in the three modalities. Here and in the following figures the straight line indicates the median, whereas the cross the mean.

The dichotomous variable accuracy was analyzed by means of generalized linear mixed models, with random intercepts for subject and items, and by-subject and by-item random slopes for the effect of modality. We found that modality affected accuracy ($\chi^2 = 21.67$, $p < 0.0001$), and specifically accuracy was significantly lower in the V modality compared to the VA modality ($\beta = -0.73$, $SE = 0.16$, $z = -4.64$, $p < 0.0001$) and in the A modality compared to the VA modality ($\beta = -1.14$, $SE = 0.26$, $z = -4.40$, $p < 0.0001$). The difference between the A and the V modality was not significant ($p = 0.14$). Considering the A and the V modality, we calculated the number of participants who performed above chance level (more of 67% correct). In the A modality, 14 participants (19%) performed at chance, and 60 (81%) above chance level. In the V modality, 7 (9%) participants performed at chance and 67 (91%) participants above chance level.

To sum up, in Study 1 we showed that when presented with pairs of literally positive comments, one uttered as an ironic criticism and one as a literal compliment, ironic remarks could be correctly detected not only by relying on the ironic tone of voice (audio tracks, A), but also when only visual cues were available (video with no audio, V). Granted that facial expressions and bodily movements do enable interlocutors to recognize irony above chance level, the second goal of the study was to assess the relative contribution of auditory and visual irony markers. We found a higher accuracy rate and slightly more participants above chance level in the V modality compared to the A modality, but since the difference between the two modalities

did not reach statistical significance, we cannot conclude that visual cues are more effective than auditory cues.

2.2. Study 2

In the first study, participants were presented with pairs of the very same remark, pronounced by the same actor once with a sincere and once with an ironic intent. This situation does not occur in ordinary circumstances, when interlocutors have to reconstruct the communicative intent of a single remark, without having the possibility of comparing it with its counterpart. To verify whether the results of the previous experiment could indeed reflect, in a laboratory setting, the process underlying irony recognition, we ran a second experiment in which participants were presented with one single remark at the time, that they had to label as ironic or sincere. In this second study, we expected general lower accuracy rates, but if the pattern of responses reflects the results of the first study, we can gain evidence of the validity of the methodology adopted in Study 1.

The material was the same one as Study 1 (20 minimal pairs of literally positive remarks, pronounced once ironically and once sincerely), which was again presented in three blocks: only video (V, with the pronounced remark written in the screen); only audio (A), and video plus audio (VA). In this experiment, though, the remarks were presented one at a time, and for each remark, participants had to indicate whether the actor was sincere or ironic. Ordering of the blocks was fixed (V, A, VA); remarks within the blocks were randomized. Since the experiment required more time to be completed, the test was run in a laboratory to better control for participants' attentiveness. Twenty-six participants (21 females, 5 males, mean age = 23 years, SD = 5 years) took part in the experiment.

In Study 2 it was our purpose to verify whether the results we obtained in the first study (accuracy above chance level in both the A and in the V modality, with visual cues being cues at least as reliable as auditory cues) were replicable also in a more ecological setting, with remarks presented one at a time. As in Study 1, the VA modality was considered the baseline. Eight remarks (4 literal compliments and 4 ironic criticisms) with an accuracy of less than 51% correct were removed. The final analysis included 32 remarks. A binomial calculation indicated that participants responding to more than 21/32 remarks correctly (66%) were not performing the task at random (p = 0.025). One participant, with a mean accuracy in VA remarks of 66% was removed. The final analysis was therefore conducted on 25 participants. Differently from the previous study, in which participants saw pairs of comments, here the remarks were presented in isolation, and thus we could calculate separately the accuracy in the recognition of literal and ironic remarks. The results are plotted in Figure 3: accuracy was overall higher in the VA modality (86% for literal and ironic meaning) with respect to the other modalities. Moreover accuracy was slightly lower in the A compared to the V modality, and in ironic comments compared to literal comments (A modality – literal: 78% / ironic: 67%; V modality – literal: 82% / ironic: 75%).
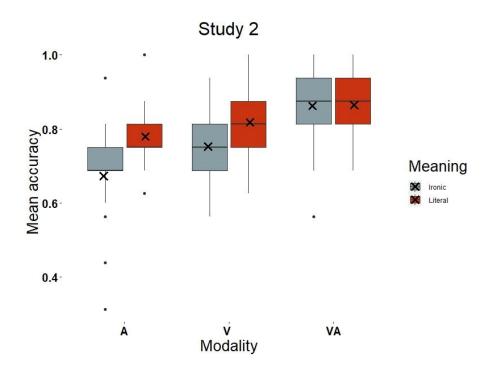
Figure 3. Boxplots representing accuracy in the three modalities in ironic and literal comments.

The dichotomous variable accuracy was analyzed by means of generalized linear mixed models, with random intercepts for subject and items, by-subject random slopes for the effects of modality and meaning and by-items random slopes for the effect of modality. We found that the effect of the modality by meaning interaction did not significantly affect accuracy ($\chi^2 = 4.89$, $p = 0.09$), nor did the main effect of meaning ($\chi^2 = 1.51$, $p = 0.22$). In other words, the fact that the remark was ironic or sincere did not have a significant impact on the rate of correct recognition. We found that modality affected accuracy ($\chi^2 = 20.25$, $p < 0.0001$), and specifically accuracy was significantly lower in the V modality compared to the VA modality ($\beta = -0.75$, $SE = 0.16$, $z = -4.55$, $p < 0.0001$) and in the A modality compared to the VA modality ($\beta = -0.98$, $SE = 0.28$, $z = -3.44$, $p = 0.0006$). The difference between the A and the V modality was not significant ($p = 0.44$). Also in this case, we focused on the A and the V modality, and we calculated the number of participants who performed above chance level (more of 66% correct). All participants responded above chance level in the V modality, whereas 4 (16%) participants responded at chance in the A modality.

Results of Study 2 confirmed that of Study 1, expanding them to a more ecological scenario: when presented with a single comment, either an ironic criticism or a literal compliment, participants could recognize if the comment was pronounced with a sincere or with an ironic intent. Accuracy decreased from the VA modality to the V and the A modality; the V modality led to better results in terms of accuracy and of percentages of participants responding above chance level compared to the A modality, but, again, the difference between these two was not significant. Importantly, all participants were above chance in recognizing the speaker's meaning in the V modality.

2.3. Study 3

The first two studies analyzed literally positive statements, that constitute compliments when the speaker is sincere, and criticisms when the speaker's communicative intent is ironical. As discussed in the introduction, ironic criticisms, or sarcasm, constitute the most widespread form of irony, and thus they can be somehow conventionalized in their expression. The purpose of Study 3 was then to check whether the visual recognition of irony, that we found in the case of ironic criticisms, was indeed possible also for ironic compliments (that had to be distinguished from literal criticisms), since they constitute a much rarer form of irony. The material for this study consisted of 20 pairs of literally negative statements (for instance "What an ugly house!") pronounced once sincerely (literal criticisms) and once ironically (ironic compliments). As in Study 1, participants were asked to indicate in which video or audio the actor was ironical, in three blocks: V, A and VA. We recruited 101 participants (81 females, 20 males, mean age = 23, $SD = 5$).

As for Study 1 and Study 2, the VA modality was considered the baseline. No item was removed because no item had an accuracy of less than 51% correct. The final analysis was therefore conducted on 20 items. We performed a binomial calculation to determine chance level. Participants responding to more than 14/20 items correctly (70%) were not performing the task at random ($p = 0.02$). Two participants, with a mean accuracy in VA items between 65% and 70% were removed. The final analysis was therefore conducted on 99 participants. Results are presented in Figure 4. Accuracy was higher in the VA modality (95%) compared to the other modalities, and slightly lower in the V modality (89%) and lowest in the A modality (74%).
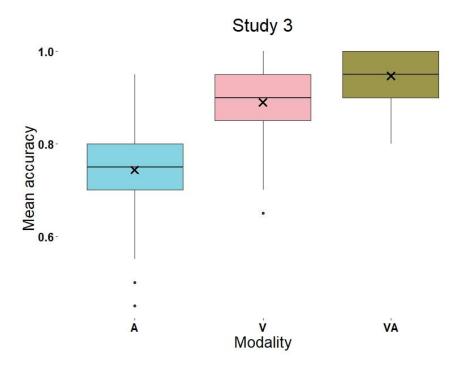

Figure 4. Boxplots representing accuracy in the three modalities.

With a generalized linear mixed models analysis with random intercepts for subject and items, and by-subject and by-item random slopes for the effect of modality, we found that accuracy was significantly affected by modality (χ^2 = 20.46, p < 0.0001). Specifically, accuracy in the A modality was significantly lower than accuracy in the VA modality (β = -2.07, SE = 0.39, z = -5.34, p < 0.0001) and significantly lower than accuracy in the V modality (β = -1.24, SE = 0.45, z = -2.76, p = 0.006). Accuracy in the V modality was significantly lower than accuracy in the VA modality (β = -0.83, SE = 0.27, z = -3.08, p = 0.002). As for participants performing at chance, they were 36 (36%) in the A modality and 6 (6%) in the V modality.

The procedure of Study 3 was like that of Study 1. Crucially, instead of having pairs of ironic criticisms and literal compliments, pairs of Study 3 were composed of ironic compliments and literal criticisms. The pattern of response in this study was similar to the first two studies: as expected, the VA modality had higher accuracy than the A and the V modality; the V block was easier than the A block both for the accuracy and for the number of participants who performed above chance level; this time, contrary to what was the case for Study 1 and Study 2, the difference in accuracy between the V and the A block reached statistical significance.

Quite unexpectedly, the overall accuracy of Study 3 looks comparable to that of Study 1 and in fact mean percentages are slightly higher than in Study 1: this is surprising because this study used pairs of literally negative remarks, that end up being literal criticisms when the actor is sincere, and ironic compliments when the meaning is reversed. Since ironic praises are much rarer than the type of irony (sarcasm) tested in Study 1 and different studies found that in general ironic compliments are harder to appreciate also for adults, we were expecting a lower accuracy in this study.

2.4. Study 4

The general goal of the studies we conducted was to establish whether purely visual information, such as speakers' facial expressions and bodily movements, could constitute reliable cues for the recognition of irony. Even if we did find evidence that participants could effectively detect well above chance level the ironic remarks when only visual information was available, we cannot yet conclude that there exist reliable visual markers of irony *per se*. The procedure of the experiments we designed involved the presentation of the video (without audio) of the actors pronouncing the remarks but, crucially, the comment was written on the screen (see Figure 1). This was done to maintain the parallelism with the only audio condition, in which the prosodic information (the sincere or ironic tone of voice) was not detached from the semantic content of the remark. The fact that participants knew what the actors were saying leaves open a partially different interpretation of our results. Instead of being guided by visual markers of irony, the participants of our studies could in fact be recognizing irony in an indirect way, by comparing the polarity of the remark with the polarity of the actors' attitude: when there is a clash between the two (e.g., the remark is positive but the actor is displaying a contemptuous attitude), the remark is interpreted as ironic (see also Mantovan, Giustolisi & Panzeri, 2019).

To verify whether the visual cues that facilitated the detection of irony were direct markers of irony, in Study 4 we presented 38 pairs of sincere/ironic remarks, 18 literally positive[1] (literal compliments / ironic criticisms, as in Study 1 and 2) and 20 literally negative (literal criticisms / ironic compliments, as in Study 3), only in the V modality, but this time the remark that was pronounced was not written on the screen, and so participants did not know whether the actors were uttering positive or negative statements.[2] The participants' task was to recognize the ironic remark. Ninety-three participants (76 females, 17 males, mean age = 23, SD = 4) took part in Study 4.

As shown in Figure 5, mean performance was 79% in recognizing ironic compliments, and lower, 64%, in recognizing ironic criticisms. With a generalized linear mixed models analysis with random intercepts for subject and items, and by-subject random slopes for the effect of type of irony, we found that this difference was significant (χ^2 = 7.12, p = 0.008). Specifically, accuracy was significantly lower in detecting ironic criticisms than in detecting ironic compliments (β = -0.94, SE = 0.34, z = -2.77, p = 0.006). As for participants performing at chance, they were 30 (32%) in detecting ironic compliments and 58 (62%) in detecting ironic criticisms. In this case, the majority of participants could not detect ironic criticisms, whereas the majority of participants could still detect ironic compliments.
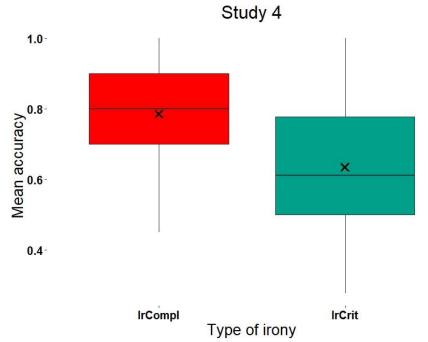

Figure 5. Boxplots representing accuracy in detecting ironic compliments (IrCompl) and ironic criticisms (IrCrit).

---

[1] Considering that in Study 1 we removed two pairs of remarks for low accuracy in the VA block, we did not use those two pairs in Study 4.

[2] We cannot exclude the possibility that some participants could lipread and thus reconstruct at least some of the remarks; however, considering the difficulty of exclusive lipreading in the absolute absence of auditory information and of a supporting context, we find this option unlikely.

3. Discussion

We aimed at verifying whether ironic remarks could be correctly identified relying on purely visual cues and whether these cues were as effective as prosodic information connected to the ironic tone of voice. We thus presented, without any preceding context, remarks uttered sincerely and ironically and participants had to recognize the ironic one. We manipulated the metacommunicative cues that were available: in the V condition, participants could only see speakers' facial expressions and bodily movements; in the A condition they could only hear the audio tracks of the uttered sentences; in the VA modality both visual and auditory information were present. In Study 1 and 3, participants were presented with pairs of the same remark and had to discriminate the ironic one; in Study 2, remarks were presented in isolation. As for the material, Study 1 and 2 tested literally positive remarks, that constituted literal compliments / ironic criticisms; in Study 3, literally negative comments were presented, corresponding to literal criticisms / ironic compliments. In all these studies, we found that purely visual cues were sufficient to discriminate the ironic intent of the speakers: participants' accuracy in the V modality was high (around 80%) and the majority of participants were responding above chance level. In fact, facial expressions and bodily movements helped participants in the detection of ironic remarks even more than purely prosodic information, even if the difference in accuracy between the V and the A modality reached statistical significance only in Study 3.

The results of Study 4, though, indicate that when the interlocutors do not know the content of the remark (as they did in the first three studies), they have more difficulties in recognizing the ironic comment. In particular, the rate of recognition of irony drastically dropped in the case of ironic criticisms: in Study 1, in the V modality, they were recognized 84% of the times, with 91% of participants above chance level, when the remark was written on the screen; when this information was not present (Study 4), the performance lowered to 64%, with only 38% of participants responding above chance level. These data suggest that the strategy employed to detect sarcasm in Study 1 and 2 was not grounded on a direct recognition of specific visual markers of irony, but involved the computation of the remarks' meaning. We hypothesized that in the first two studies participants could correctly detect the ironic remarks only indirectly, because they were comparing the positive polarity of the remark with the emotional attitude expressed by the speaker via specific facial expressions: when they converged, the remark was interpreted as sincere, when they diverged, it was read as ironic.

This hypothesis does not necessarily speak in favor of a two-stage computation of figurative meaning. There is a lively debate about the processing of ironic statements, that is whether interlocutors have to first process the literal meaning (Grice, 1975), or whether they have a direct access to the (reversed) ironic meaning (Gibbs, 1994), or whether the presence of specific contextual cues or linguistically conventionalized forms makes one reading more salient than the other (Utsumi, 2000; Giora, 2003). Our studies do not investigate how interlocutors process the meaning of comments, since they require a metalinguistic judgment that might follow a different inferential path. Indeed, as Deliens and colleagues (2018) argued, the simple discrimination of the ironic/sincere intent of a message does not necessarily coincide with its full comprehension. Future studies, then, might further explore the role of visual irony markers in the online processing and comprehension of irony.

The fact that the rate of recognition of ironic criticisms was rather low in the last experiment, with the majority of participants responding at chance level questions the existence of visual markers of "pure" irony. Focusing on ironic criticisms, we might hypothesize that a key role was played by the detection of the speaker's negative attitude, displayed by particular facial expressions. In fact, also in the case of intonation, Bryant & Fox-Tree (2005) found that the dripping sarcastic utterances were rated not only as more ironic, but also as more angry and they concluded that "an ironic tone of voice could also be described as an angry tone of voice". This is consistent with Wilson & Sperber (2012)'s account, that views sarcastic speakers as expressing their negative, contemptuous attitude towards the thought expressed by the ironic comment. We thus speculate that in the first two studies, when participants knew that the speaker was in fact uttering a literally positive remark, they could detect the speaker's contemptuous attitude, which clashed with the positivity of the remark, and they eventually inferred that the speaker was ironic. On the other hand, in Study 4, participants were asked to discriminate in which video the speaker was being ironic, without knowing the content of the remark. When presented with pairs of videos in which speakers were either uttering a literal compliment and an ironic criticism, we believe that participants could mistake the speaker's positive stance in making a literal compliment with the jocularity typically associated with irony. Recall, in fact, that also in Deliens and colleagues (2018)'s first experiment participants rated literal positive statements as more ironic than their literal negative versions, suggesting that the positive attitude conveyed by praises might be confused with the ironic intent.

Another somehow surprising result of our studies concerns ironic compliments: in Study 3, accuracy in the detection of ironic praises was unexpectedly high, especially in the visual mode (89% in the V modality; 74% in the A modality). Ironic compliments are much rarer than sarcasm, and they are harder to understand, especially if they do not echo a preceding statement (Kreuz & Glucksberg, 1989), yet they were easily discriminated from their literal counterparts. A possible explanation of this result is consistent with the hypothesis that our participants equated irony with the expression of a positive attitude, and thus when they were presented with videos of speakers uttering an (ironic) compliment or a (literal) criticism, they identified irony with the expression of a positive stance. This assumption would also account for the relatively high accuracy in the detection of ironic compliments in the last study, when participants did not know the content of the uttered remark. To account for the even higher accuracy in Study 3, when the pronounced remark was written on the screen, though, we might follow a suggestion put forth by Mauchand, Vergis & Pell (2020). They presented audio tracks of literally positive and negative statements uttered sincerely and ironically, and asked participants to rate the friendliness of the speaker, in three modalities: with no particular focus, focusing on the prosody, and focusing on the content. They found that utterances with negative content elicited negative ratings of friendliness, even when the speaker's intonation was ironical, and thus should have conveyed a compliment. They thus conclude that the negativity of the propositional content would alert participants against potential threats expressed by demeaning judgments, and lead them to rate the speaker as unfriendly. In our case, though, the relevance of the negative stimulus might have rendered our participants more attentive towards the facial expressions displayed by the speakers, and thus recognize even better the positive attitude conveyed in the ironic utterances.

Before concluding, we should recognize that a fine grained analysis of our video material is needed to tackle the question of which particular facial expressions (if any) are connected to

the expression of irony and/or of the speaker's attitude. We are planning to conduct this analysis to verify the speculative discussion of the present paper.

References

Anolli, L., Ciceri, R., & Infantino, M. G. (2002). From "blame by praise" to "praise by blame": Analysis of vocal patterns in ironic communication. *International Journal of Psychology*, 37(5), 266-276.

Attardo, S. (2000). Irony markers and functions: Towards a goal-oriented theory of irony and its processing. *Rask*, 12(1), 3-20.

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor*, 16(2), 243-260.

Bryant, G. A., & Fox Tree, J. E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and symbol*, 17(2), 99-119.

Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice?. *Language and speech*, 48(3), 257-277.

Burgers, C., & van Mulken, M. (2017). Humor markers. In: S. Attardo (Ed.) *The Routledge handbook of language and humor*, pp. 385-399. New York: Routledge.

Cheang, H. S., & Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. *The Journal of the Acoustical Society of America*, 126(3), 1394-1405.

Cheang, H. S., & Pell, M. D. (2011). Recognizing sarcasm without language: A cross-linguistic study of English and Cantonese. *Pragmatics & Cognition*, 19(2), 203-223.

Clark, H. H., & Gerrig, R. J. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General*, 113(1), 121–126.

Deliens, G., Antoniou, K., Clin, E., Ostashchenko, E., & Kissine, M. (2018). Context, facial expression and prosody in irony processing. *Journal of memory and language*, 99, 35-48.

Dews, S., & Winner, E. (1995). Muting the meaning a social function of irony. *Metaphor and Symbol*, 10(1), 3-19.

Félix-Brasdefer, J.C. (2010). Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In: A. Martínez-Flor, E. Uso-Juan (Eds.), *Speech Act Performance: Theoretical, Empirical, and Methodological Issues*, pp. 41-56. Amsterdam/Philadelphia: John Benjamins.

Fish, K., Rothermich, K., & Pell, M. D. (2017). The sound of (in) sincerity. *Journal of Pragmatics*, 121, 147-161.

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

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and symbol*, 15(1-2), 5-27.

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

Grice, H.P. (1975). Logic and conversation. In P. Cole, J.L. Morgan (Eds.), *Syntax and Semantics, vol. 3*, pp. 41-58. New York: Academic Press.

Hancock, J. T. (2004). Verbal irony use in face-to-face and computer-mediated conversations. *Journal of Language and Social Psychology*, 23(4), 447-463.

Keltner, D., Ekman, P., Gonzaga, G. C., & Beer, J. (2003). Facial expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences*, pp. 415–432. New York: Oxford University Press.

Kreuz, R. (1996). The use of verbal irony: Cues and constraints. In: J. S. Mio & A. N. Katz (Eds.), *Metaphor: Implications and applications*, pp. 23-38. Mahwah, NJ: Lawrence Erlbaum.

Kreuz, R. J., & Glucksberg, S. (1989). How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of experimental psychology: General*, 118(4), 374-386.

Kreuz, R. J., & Link, K. E. (2002). Asymmetries in the use of verbal irony. *Journal of Language and Social Psychology*, 21(2), 127-143.

Kreuz, R. J., & Roberts, R. M. (1995). Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Metaphor and symbol*, 10(1), 21-31.

Lœvenbruck, H., Jannet, M. A. B., d'Imperio, M., Spini, M., & Champagne-Lavau, M. (2013). Prosodic cues of sarcastic speech in French: slower, higher, wider. In *Proceedings of 14th Annual Conference of the International Speech Communication Association* (Interspeech 2013), Lyon, France, pp. 3537-3541.

Mantovan, L., Giustolisi, B., & Panzeri, F. (2019). Signing something while meaning its opposite: The expression of irony in Italian Sign Language (LIS). *Journal of Pragmatics*, 142, 47-61.

Mauchand, M., Vergis, N., & Pell, M. D. (2020). Irony, prosody, and social impressions of affective stance. *Discourse Processes*, 57(2), 141-157.

Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic research*, 29(5), 483-495.

Scharrer, L., & Christmann, U. (2011). Voice modulations in German ironic speech. *Language and speech*, 54(4), 435-465.

Seto, K.-i. (1998). On non-echoic irony. In R. Carston & S. Uchida (Eds.), *Relevance Theory: Applications and Implications*, pp. 239-255. Amsterdam: Benjamins.

Truesdale, D. M., & Pell, M. D. (2018). The sound of Passion and Indifference. *Speech Communication*, 99, 124-134.

Utsumi, A. (2000). Verbal irony as implicit display of ironic environment: Distinguishing ironic utterances from non-irony. *Journal of Pragmatics*, 32(12), 1777-1806.

Wilson, D. & Sperber, D. (2012). Explaining irony. In: *Meaning and relevance,* pp. 123-145. Cambridge, MA: Cambridge University Press.