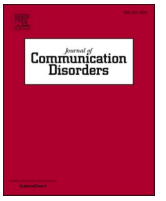
Journal of Communication Disorders 118 (2025) 106587

Contents lists available at ScienceDirect

Journal of Communication Disorders

journal homepage: www.elsevier.com/locate/jcomdis



Design and validation of a sarcasm recognition task

Pablo J. Ocampo [a], Nicolás Vassolo [a], Bautista Elizalde Acevedo [a,b], Mariana Bendersky [c,d,1], Lucía Alba-Ferrara [a,c,*,1]

[a] Facultad de Ciencias Biomédicas, Universidad Austral, Pilar, Buenos Aires, Argentina
[b] Instituto de Investigaciones en Medicina Traslacional (IIMT), CONICET-Universidad Austral, Derqui-Pilar, Buenos Aires, Argentina
[c] Unidad Ejecutora para el Estudio de las Neurociencias y Sistemas Complejos (ENyS), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina
[d] Laboratorio de Anatomía Viviente, 3ra Cátedra de Anatomía Normal, Facultad de Medicina, Universidad de Buenos Aires, Buenos Aires, Argentina

ARTICLE INFO

Keywords:
Sarcasm
Social cognition
Mentalization
Pragmatic language
Task validation

ABSTRACT

Pragmatic language, which includes expressions of non-literal meaning, is crucial for social cognition, yet current assessment techniques are inadequate for the neurotypical population. This study was designed to develop and validate a sarcasm recognition task, a domain of pragmatic language, in a sample of 84 healthy adults. The task was specifically created for Spanish-speakers, highlighting its cultural and linguistic relevance.

In experiment 1, forty original vignettes containing sarcastic or literal content were evaluated by 49 subjects to select reliable stimuli, resulting in a subset with high reliability. In experiment 2, the new task and a neuropsychological screening battery were administered to 35 healthy adults. Cronbach's alpha was used to analyze reliability, and correlation indices with gold standard social cognition tests were computed.

Experiment 1 yielded 40 reliable stimuli (Cronbach's Alpha: 0.818), which constituted the final version of the task. These same 40 items were administered in Experiment 2, where the task demonstrated significant correlations with established social cognition tests. The findings indicate that the novel sarcasm recognition task is brief (40 total items), reliable and robust, without ceiling effects on the neurotypical population, and its design will allow for evaluation in parallel with fMRI scans.

1. Introduction

The ability to interact effectively with others is sine qua non for successful social interaction. Social interaction is inherent to humans, and it is considered essential for the proper development of nervous, endocrine, and immune systems (Repetti et al., 2007). There is a broad range of skills supporting complex human social interaction that involves not only the full spectrum of basic cognitive functions but also a set of specific abilities to accurately process social messages to respond appropriately (Beaudoin & Beauchamp, 2020). These specific processes are crucial for social performance and constitute what is called social cognition, understood as "the ability to recognize, manipulate, and behave regarding socially relevant information" (Adolphs, 2001, p.231).

A key component of social cognition is Theory of Mind, *the ability to attribute mental states to others, recognizing that these may differ*

* Corresponding author. Buenos Aires, Argentina.
E-mail address: lalba@austral.edu.ar (L. Alba-Ferrara).
[1] Both authors share senior authorship

https://doi.org/10.1016/j.jcomdis.2025.106587
Received 26 May 2025; Received in revised form 9 September 2025; Accepted 26 October 2025
Available online 28 October 2025

from one's own—for instance, others may possess knowledge that one lacks, or lack knowledge that one possesses (Premack & Woodruff, 1978; Beaudoin & Beauchamp, 2020). Since this is not directly perceptible through sensory channels, it is predicated on inferential processes; it is a deduction, a theory.

While the terms *mentalization* and *Theory of Mind* are often used interchangeably, in this manuscript we adopt a hierarchical perspective situating them within broader socio-cognitive constructs. Social cognition is understood as the overarching domain encompassing processes by which humans interpret and respond to others in social contexts (Schurz et al., 2014). Within this domain, *mentalization* denotes the broad capacity to interpret behavior in terms of underlying mental states—such as desires, intentions, emotions, or beliefs (Forgács, 2024). In contrast, Theory of Mind is conceptualized as a more specific subcomponent, primarily concerned with attributing cognitive or epistemic states to others, namely what they know, believe, or think about the world (Premack & Woodruff, 1978). Evidence supports a hierarchical and fractionated organisation of socio-cognitive functions, where broader constructs give rise to more specialized processes (Schurz et al., 2014). Pragmatic language can be viewed as a downstream manifestation of socio-cognitive processing: while Theory of Mind and mentalization provide inferential mechanisms necessary for pragmatics, pragmatic language extends those mechanisms into communicative practice, relying on contextual cues, speaker intentions, and non-literal language use (Bosco et al., 2018; Ouerchefani et al., 2024). Thus, we conceptualize pragmatic language as nested within social cognition and dependent upon Theory of Mind and mentalization capacities.

Theory of Mind and mentalization enable other fundamental socio-cognitive processes, which in turn constitute the umbrella construct of social cognition, including pragmatic language. Pragmatic language refers to the ability to attribute meaning to verbal or written communications not necessarily restricted to the literal content of what is conveyed (Monetta et al., 2009), thus transcending semantics or syntax (Uchiyama et al., 2006). This allows for highly complex uses of language, including, for example, the ability to produce fiction for various purposes, humorous uses of language, or the comprehension and production of ironic or sarcastic messages (Monetta et al., 2009). Mastery of these pragmatic uses of language is important for effective social interaction, and deterioration in these capacities is associated with significant difficulties in daily life, as observed in individuals with Autism Spectrum Disorders (Wang et al., 2006) or in those with neurological injuries to the lobes (Shamay-Tsoory et al., 2005) or white matter connective tracts (Davis et al., 2016).

The development of this capacity is gradual and presents marked evolutionary stages; in the case of sarcasm and irony, the detection of non-literal meaning begins to develop around the age of six, while the underlying pragmatic intention becomes recognizable only around nine to ten years of age (Glenwright & Pexman, 2010). It is also important to distinguish between these two subdomains; although both depart from the literal meaning, irony is characterized by conveying meanings opposite to it, while sarcasm—although it often adopts opposite meanings—is directed toward people, often with the intention of criticism or with biting implications (Filik et al., 2019; Uchiyama et al., 2006) or non-aggressive humorous intentions, and can even be used to mitigate the social or confrontational consequences of a critical expression (Davis et al., 2016). A phrase like "What a beautiful day" during a rainy afternoon would be ironic, yet to be sarcastic it should be directed to a specific person with a more biting connotation, a teacher could use it towards a student who arrives late to class by saying something like "Nice of you to be on time" or "Glad your watch is working well". However, the social uses of sarcasm are not limited to criticism. Recent studies demonstrate previously unknown functions, such as the ability to use sarcasm as an adaptation mechanism, a pragmatic and resilient use of sarcasm to cope with challenging situations, improving well-being (Rothermich et al., 2021). This use is associated with humorous functions, frequently linked to reducing symptoms of anxiety or depression (Melton, 2016; Wellenzohn et al., 2016).

Neuroanatomically, the capacity for non-literal language use has consistently been associated with the right hemisphere (Sotillo et al., 2005; Faust & Mashal, 2007; Schmidt & Seger, 2009; Alba-Ferrara et al., 2011; Bendersky et al., 2021; Hauptman et al., 2023). The implication of non-literal language uses in everyday communication is such that individuals with damage to this hemisphere notably struggle to identify the central points of a social communicative exchange (Hough, 1990) and to distinguish lies and jokes (Winner et al., 1998). In the current state of scientific discussion, the role of each cerebral hemisphere in processing non-literal language functions is still under debate. While the involvement of the right hemisphere has been consistently found in numerous studies, the greater consensus at present points to the involvement of both hemispheres, thus suggesting a bilateral processing of non-literal language functions (Jung-Beeman, 2005; Stringaris et al., 2007; Zempleni et al., 2007; Papagno & Romero Lauro, 2010; Bambini et al., 2011; Bohrn et al., 2012; Hagoort & Levinson, 2014; Reyes-Aguilar et al., 2018; Hauptman et al., 2023).

The inherently complex discussion regarding the hemispheric localization of pragmatic language processing includes the ambitious task of distinguishing areas responsible for processing each subdomain. This is because even the most similar subdomains, such as irony and sarcasm (often grouped as synonyms in popular discourse), present differentiated neural networks for their processing (Filik et al., 2019). Specifically, understanding sarcasm is particularly complex because it is associated with multiple contextual factors, including factors such as age and educational level (Matsui et al., 2016), intonation or prosody in verbal communications (Davis et al., 2016), or the presence of eye contact (Williams et al., 2009), among others. The current lack of a single use task to assess pragmatic language in an fMRI setting makes it even more complicated to get neuroimaging results that are replicable. The use of different, non-validated tasks to elicit pragmatic language during fMRI scans results in variability, hence the need for a validated task.

A study on irony comprehension applied the contextual discrepancy paradigm (Valles-Capetillo et al., 2022). In this task, participants read short written scenarios (30–40 words) that set up an interaction between two characters, followed by a brief statement (3–6 words). The same statement could be interpreted differently depending on the context, yielding four conditions: literal, ironic, unrelated, or white lie. Thus, an identical utterance (e.g., "You are playing very well") could shift in meaning depending on the written context provided (e.g., after a poor performance, a flawless one, or when the speaker wished to conceal the truth). The study found that contextual discrepancy led participants to identify irony more effectively than the other cues, yielding higher classification accuracy and classification time. While the study also aimed to investigate the relation between irony comprehension and specific cognitive

functions, such as Theory of Mind, social skills or imagination, it found no consistent correlations. Finally, the researchers did not compare the instrument with gold standard tests.

Another recent study (Nakamura et al., 2022) used a task specifically designed for sarcasm detection based on auditory stimuli in a Japanese sample. In this study, participants listened to short contexts followed by a target sentence. The interpretation of the target depended on both the preceding context and its prosody, making the same utterance sound literal or sarcastic. Typically, participants judge the degree of sarcasm in the target sentence, allowing the dissociation of contextual and prosodic contributions to pragmatic meaning. For the design of this task, they did not conduct a validation process, nor did it receive adaptations to Spanish. Furthermore, it differs from our sarcasm recognition task as Nakamura and colleagues used auditory stimuli involving prosodic intonation. This format is overall more demanding for subjects than visual stimuli given the way it relies on working memory to hold the initial hypothetical scenario to then contrast it with the congruent or incongruent comment.

In 2016, another team of Japanese researchers conducted a similar study, performing functional magnetic resonance imaging in parallel with sarcasm recognition tasks, seeking information about the underlying mechanisms of auditory sarcasm comprehension (Matsui et al., 2016). In this paradigm, participants were presented with short parent–child scenarios consisting of four phases: (1) a background situation; (2) the parent's instruction to the child; (3) the child's response; and (4) the parent's final comment. The first three phases were shown through written text accompanied by illustrations, while the parent's final comment was presented either as written text or as an auditory stimulus recorded with positive or negative prosody. Participants then judged whether the parent's comment was meant sincerely or sarcastically. For instance, a child is instructed to tidy up toys before eating but instead begins to snack without cleaning. The parent then comments, "You did a great job tidying the toys!"—a remark that, in this context, is interpreted as sarcasm. Much like the previous research (Nakamura et al., 2022), the authors did not perform statistical analysis of behavioral data to validate the stimuli and there was no adaptation to other languages nor were replication studies to ascertain the validity of the instrument. Tables 1 and 2

Another recent investigation was carried out in 2019 and aimed to discern the underlying neurological mechanisms of sarcasm processing and non-sarcastic irony using functional magnetic resonance imaging in parallel with a task developed by the researchers (Filik et al., 2019). The applied instrument consisted of sixty short stories, between two characters. In this task, participants read short dialogues in which the final remark could be interpreted as sarcastic irony, non-sarcastic irony, or literal, depending on the preceding context. For example, after Caitlin chose to print her application on pink paper, Bernice commented, "Very formal!"—a statement understood as sarcastic criticism in this context. The instrument used for this study allowed researchers to detect brain activations of networks associated with mentalization and semantic abilities, while also distinguishing activation margins based on each stimulus condition. However, there was no exhaustive task validation process, nor was there an attempt to develop a valid instrument for use in Spanish-speaking populations.

To sum up, although pragmatic language has been assessed with several widely used measures, none are currently available in Spanish or adapted for Argentine populations. Moreover, tasks commonly employed in other languages have often not been validated at the level of individual stimuli, and they typically capture broad pragmatic abilities rather than focusing specifically on irony and sarcasm. To address this gap, we developed and validated a culturally appropriate task specifically targeting sarcasm comprehension for Spanish speakers.

2. Method

The development and validation of the sarcasm recognition task involved a two-experiments research design, initially designing and evaluating the task, and later validating it through the statistical correlation with gold standard instruments and the analysis of its internal consistency.

Table 1
Sample details.

Gender	Amount	Sample percentage
Male	28	57.1
Female	21	42.9
Total	**49**	**100**
Age	**Amount**	**Sample percentage**
50 – 65	1	2.1
40 – 49	6	12.2
30 – 39	12	24.3
20 – 29	26	53.2
18 – 19	4	8.3
Total	**49**	**100**
Years of education	**Amount**	**Sample percentage**
20 – 25	4	8.2
15 – 19	29	59.1
12 – 14	16	32.7
Total	**49**	**100**

Table 2
Sample details.

Gender	Amount	Sample percentage
Male	22	62.9
Female	13	37.1
Total	**35**	**100**
Age	**Amount**	**Sample percentage**
50 – 65	2	5.7
40 – 49	4	11.4
30 – 39	9	25.7
20 – 29	13	37.1
18 – 19	7	20.1
Total	**35**	**100**
Years of education	**Amount**	**Sample percentage**
20 – 25	3	8.6
15 – 19	21	60.0
12 – 14	11	31.4
Total	**35**	**100**

2.1. Experiment 1: Task design

2.1.1. Participants

The experiment included 49 healthy, literate adult volunteers recruited by convenience sampling. Eligibility was verified in a brief interview to ensure no history of neurological disorders or vision/reading difficulties. Most participants were students or faculty members from Austral University Hospital, while additional participants were recruited from non-academic staff of the same institution (e.g., administrative, technical, and support personnel) to broaden heterogeneity in educational background.

This project received the corresponding approval from the Ethics Committee of Hospital El Cruce (HEC) in the Province of Buenos Aires, following the guidelines set forth in the Declaration of Helsinki.

2.1.2. Data collection instruments

Sarcasm Recognition Task: Forty illustrated and digitized vignettes representing interaction scenarios between people or expressions of a single individual were designed. They consist of black and white illustrations with a short text bubble containing a single sentence ranging from 2 to 7 words. The vocabulary in each vignette was chosen to be simple and easily understandable, exclusively using high-frequency words in everyday language, meaning no complex or infrequent vocabulary was used. Additionally, the situations depicted in each vignette aimed to reflect everyday experiences, ensuring adequate ecological validity, where each stimulus corresponds to a common cultural context, aiming to be suitable for use in Argentine participants. The vignettes were designed in pairs, with each pair sharing the same written expression with drawings representing different scenarios, one in which written expression and drawing were consistent (literal pair) and another variant of it, discordant with the text (sarcastic pair). The discrepancy between the drawing and the written expression in the sarcastic vignettes was exaggerated through caricature-like images considering the discrepancy between the written expression and the context as a fundamental indicator of sarcasm (Nakamura et al., 2022).

The sarcasm recognition task was designed through use of the software e-prime (https://pstnet.com/products/e-prime/), it presents 40 stimuli for a period of 4.5 s each in a random order, with an intertrial interval of 0.5 s. The whole task lasts for 180 s. A 4.5-

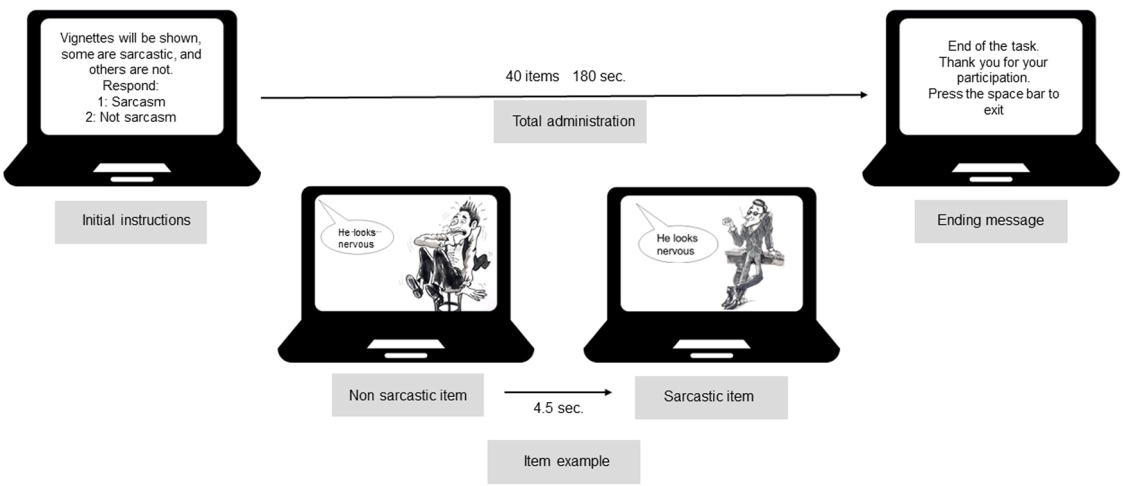


Fig. 1. Example vignettes from the Sarcasm Recognition Task illustrating literal and sarcastic.

seconds response window was adopted based on pilot testing, which showed that shorter limits increased missed responses, while this duration ensured >90 % completion. Within this time window, the participant must select using the keyboard whether they consider the stimulus to be sarcastic or literal. Responses are not forced, meaning a response is not required to move to the next stimulus. This allows for 3 response alternatives: pressing key 1 if the stimulus is considered sarcastic, pressing key 2 if it is considered literal, or not responding if the response time is exhausted. The fixed amount of time per stimulus configures an instrument whose administration takes 3 min, regardless of the subject's performance. Response time was not scored; missing responses were logged if no answer was given within the window.

All data and materials used in the analysis are publicly available at OSF (https://osf.io/ghrjc/?view_only=0d5ddb6ed088478dbf6734f966cdd242).

2.1.3. Procedure

Forty-nine participants were evaluated with the first version of the task. For each vignette, we computed the proportion of correct responses and compared it against chance level (50 %) using one-sample binomial tests. Vignettes that did not reach a proportion of correct responses significantly above chance ($p < .05$) were considered unreliable. These items were either replaced with new scenarios or modified to increase clarity and reduce ambiguity before inclusion in the final version of the task. Fig. 1

2.2. Experiment 2: Task validation

2.2.1. Participants

This experiment was conducted with a sample of 35 healthy and literate adult volunteers. They were recruited through the same method of non-probabilistic sampling as in experiment 1. Participants in Experiment 1 were not eligible to participate in Experiment 2, guaranteeing that the two samples were independent.

2.2.2. Stimuli

The number of vignettes was kept constant across experiments. Items that did not perform reliably in Experiment 1 were removed, and new items—with different illustrations and revised sentences—were created for Experiment 2.

2.2.3. Data collection instruments

In order to validate the sarcasm recognition task, a battery of gold standard tests evaluating Theory of Mind and pragmatic language was employed alongside the newly developed task. This consisted primarily of subtests from the Montreal Communication Evaluation Protocol (MCE), a psychometric tool validated in Argentina and developed for the evaluation of different language subdomains in clinical populations with lesions in the right hemisphere. Among other language subdomains, it includes comprehension of metaphors, semantic associations, and direct and indirect speech acts, and is considered a gold standard protocol for assessing pragmatic language (Ferreres et al., 2007). From this protocol, the following subtests were taken: Metaphors - Semantic Judgments - Speech Acts. Other tests in the battery include the NART (Del Ser et al., 1997), a brief, validated, and commonly used test employed to obtain an estimate of verbal IQ; the Digit Span (Weschler & Stone, 1945), two brief tests universally used to assess attention span and working memory capacity; and finally, the Reading the Mind in the Eyes Test (Abbreviated and Adapted Spanish Version RME-18), a locally adapted version of a widely validated instrument for assessing Theory of Mind capacity.

2.2.4. Procedure

The sarcasm recognition task resulting from experiment 1 was applied to the 35 participants of this experiment alongside the battery of instruments for neuropsychological screening, a Theory of Mind evaluation, and a series of subtests from the Montreal Communication Evaluation Protocol to assess pragmatic language. The results of these administrations were analyzed using the SPSS statistical program, evaluating reliability coefficients using Cronbach's alpha, and analyzing correlations based on the Pearson coefficient.

3. Data analysis

Experiment 1: The data obtained from the implementation of the sarcasm recognition task were analyzed by evaluating the accuracy rate for each trial through a one sample binomial test.

Experiment 2: Reliability tests were conducted for all stimuli of the sarcasm recognition task, evaluating the Cronbach's alpha coefficient for the entire task as well as for each item. Additionally, Pearson correlation tests were carried out to demonstrate significant positive correlations between the sarcasm recognition task and the gold standard tests. This analysis included the results stemming from the administration of the sarcasm recognition test and the battery of instruments selected to screen the subjects, evaluate Theory of Mind, and assess different subdomains of pragmatic language.

4. Results

4.1. Experiment 1

The initial administration of the task was used for item analysis and refinement. For each vignette, we computed the proportion of

correct responses and compared it against chance level (50 %) using one-sample binomial tests. Vignettes that did not reach accuracy significantly above chance ($p < .05$) were replaced or modified (e.g., new illustrations and/or wording). From the original 40 vignettes, 14 were identified as not significantly above chance level ($p < .05$) in Experiment 1 and were subsequently revised or replaced with new illustrations and wording. The final version of the task comprised 40 vignettes, all of which performed significantly above chance. Table 3 reports item-level accuracy for the final set. Internal consistency of the final version was also adequate (Cronbach's α = [0.818]).

4.2. Experiment 2

The reliability analysis of the sarcasm recognition task was measured using Cronbach's internal consistency index. The obtained value for the total scale was 0.818, reflecting the consistency of the stimuli and the task.

Pearson correlation coefficients were calculated using the SPSS statistical program. Two tailed significant correlations were found between the sarcasm recognition task and both response alternatives in the Speech Acts subtest, with the free response variant in the Metaphors subtest, and with the Theory of Mind test, the RME-18. The results of this analysis are included in Table 4.

Positive correlations were found between the sarcasm task and the Metaphors 1, Speech Acts 1, and Speech Acts 2 subtests of the MEC protocol. Regarding the remaining 3 subtests, no significant correlations were found.

A significant correlation was also found between the sarcasm recognition task and the RME-18 test. Finally, no correlation was found between the sarcasm recognition task and the NART test.

5. Discussion

This investigation generated a novel task to measure sarcasm processing skills, which demonstrated robust psychometric properties, including adequate internal consistency among items. The clarity and standardization of the vignettes likely facilitated consistent interpretation, underscoring the methodological rigor of the task and aligning with prior research emphasizing the pivotal role of contextual frameworks in sarcasm recognition (Nakamura et al., 2022). Moreover, these findings reaffirm the theoretical understanding that sarcasm comprehension depends not only on linguistic cues but also on shared contextual knowledge, speaker intent, and social norms (Nakamura et al., 2022). The task's design, therefore, contributes to the growing body of evidence supporting the multifaceted nature of sarcasm as a social cognitive skill.

The task's design also highlights practical advantages. The use of visual stimuli provides a brief, accessible, and adaptable means to evaluate sarcasm recognition. Unlike auditory tasks, which can be time-consuming and cognitively demanding due to their sequential presentation of cues, visual stimuli present all necessary information simultaneously, allowing for quicker and more intuitive decision-making (Cohen, Horowitz & Wolfe, 2009). This efficiency makes the task particularly suitable for neuroimaging settings, where time and cognitive load are critical considerations (Davis et al., 2016; Laval & Bert-Erboul, 2005).

Experiment 2 showed associations between the sarcasm recognition task and theory of mind as measured by the (RME-18), and with the Metaphors 1, Speech Acts 1, and Speech Acts 2 subtests of the MEC protocol which assess pragmatic language.

The correlation between the sarcasm recognition task and three of the MEC protocol subtests reflects the construct validity of the

Table 3
Accuracy per stimulus.

Stimulus	Accuracy	Stimulus	Accuracy
L1	0.80	S1	0.91
L2	0.71	S2	0.80
L3	0.80	S3	0.80
L4	0.80	S4	0.80
L5	0.60	S5	0.80
L6	0.68	S6	0.77
L7	0.51	S7	0.80
L8	0.77	S8	0.66
L9	0.82	S9	0.83
L10	0.65	S10	0.86
L11	0.71	S11	0.83
L12	0.85	S12	0.74
L13	0.51	S13	0.63
L14	0.77	S14	0.88
L15	0.74	S15	0.71
L16	0.74	S16	0.68
L17	0.88	S17	0.85
L18	0.91	S18	0.77
L19	0.91	S19	0.77
L20	0.51	S20	0.88
N:35			

Note: L1 to L20 represent the Literal stimuli. S1 to S20 represent the Sarcastic stimuli.
Sarcasm Recognition Task final version.

Table 4
Pearson correlation coefficients.

		Sarcasm Recognition Task
NART	*Correlation*	−0.10
	Sig.	.56
RME-18	*Correlation*	.46**
	Sig.	.01
Metaphors 1*	*Correlation*	.66**
	Sig.	.00
Metaphors 2*	*Correlation*	−0.08
	Sig.	.64
Speech Acts 1*	*Correlation*	.64**
	Sig.	.00
Speech Acts 2*	*Correlation*	.51**
	Sig.	.00
Semantic Judgement 1*	*Correlation*	.05
	Sig.	.77
Semantic Judgement 2*	*Correlation*	.20
	Sig.	.24
	M	0.8
	DT	0.1
	N	35

** . Significant correlation at the 0.01 level.
* Subtests from the Montreal Communication Evaluation Protocol (MEC).

instrument, associated with other pragmatic language tests, although they do not exactly address the same subdomain. While sarcasm is a form of pragmatic language, there are other aspects to this function, such as the comprehension of metaphors or speech acts, skills addressed by the MEC protocol and its subtests (Monetta et al., 2009). The correlation between the sarcasm recognition task and the RME-18 test demonstrates a partial overlap of Theory of Mind capacity in sarcasm comprehension and in pragmatic language skills (Premack & Woodruff, 1978; Schurz et al., 2014; Beaudoin & Beauchamp, 2020), both contributing to the construct of Social Cognition (Adolphs, 2001). Sarcasm comprehension relies on the ability to infer the speaker's intentions and distinguish between literal and implied meanings, which are core aspects of Theory of Mind. However, the moderate correlation coefficient observed between the tests suggests that sarcasm recognition extends beyond Theory of Mind, engaging additional facets of pragmatic language. This divergence highlights the multifaceted nature of pragmatic language, wherein sarcasm acts as a complex linguistic construct requiring the integration of Theory of Mind with broader cognitive-linguistic abilities. Furthermore, the associations observed with the MEC protocol subtests suggest that these tasks may tap more directly into the nuances of sarcasm processing, including its dependence on emotional recognition and advanced social reasoning. Thus, while the connection between Theory of Mind and sarcasm is significant, the findings underscore the importance of pragmatic language as a broader framework for understanding social cognition.

The NART test was added to the battery of instruments as a screening tool to ensure no deficits in premorbid verbal IQ interfered with the administration of the task. The lack of association between the sarcasm recognition task and the NART test is an expected and desired outcome, confirming that the task does not rely on high verbal ability for accurate performance. This independence is particularly significant for ensuring the accessibility and validity of the task across diverse populations. By designing the vignettes with simple and universally understandable expressions, the task remains effective for individuals with varied educational backgrounds and differing levels of premorbid verbal IQ, minimizing the risk of confounding results due to linguistic complexity (Nelson & Willison, 1991). This is crucial for practical research and clinical applications, as tasks overly reliant on verbal IQ may inadvertently exclude or misrepresent populations with lower educational attainment or neurodevelopmental conditions. Consequently, the sarcasm recognition task provides a more equitable measure of pragmatic and social cognitive skills, allowing for its broader applicability and relevance across demographic groups

Interestingly, no significant correlations were found between the sarcasm recognition task and the Metaphors 2 and Semantic Judgments subtests. The lack of correlation between the task to be validated and the responses mediated by options in the Metaphors 2 subtest can be associated with a ceiling effect in the MCE protocol subtest. The scores obtained by each subject in this half of the subtest were practically identical in all cases, with all participants achieving a perfect score or close to it. Due to this restricted score variability resulting from a ceiling effect, this half of the Metaphors subtest does not show significant correlations with any other test. The same was observed for the Semantic Judgments subtest, where the presence of a similar phenomenon is inferred, given that the restricted score variability again led to a lack of correlation with the sarcasm recognition task. The high accuracy percentages for these tests and the appearance of a ceiling effect can be attributed to the Montreal Communication Evaluation Protocol being developed to assess populations with right hemisphere damage (Ferreres et al., 2007). In contrast, our sample consisted of highly functional university students, as the sarcasm recognition task was designed to be applicable in healthy populations or those without neurological damage, thus prone to not present a ceiling effect.

Previous research on sarcasm recognition with functional magnetic resonance imaging (fMRI) has not been behaviorally validated, nor are there reports of validated tasks in Spanish or adapted to Latin American populations (Filik et al., 2019; Matsui et al., 2016; Nakamura et al., 2022; Valles-Capetillo et al., 2022). In summary, the sarcasm recognition task stands out as a novel instrument in this field.

It is worth noting that a relatively homogeneous sample was selected for our study, as the participants were highly educated (no subject with <12 years of formal education were included), mostly within the same age range (20 to 29 years), and they did not have a history of neurological or mental illness. Although this does not affect the conclusions reached, in future studies, it would be interesting to evaluate different groups to make comparisons considering that other studies have pointed out the importance of factors such as age and educational level in sarcasm comprehension. Implementation of the sarcasm recognition task on young participants samples could provide insights into the development of sarcasm recognition ability. This capacity is currently considered to develop around the age of 6 years old (Glenwright & Pexman, 2010), and our task could help to investigate the acquisition of sarcasm recognition in the developing brain.

Furthermore, considering the homogeneity of the current sample, the validation of the new instrument represents a critical step toward its application in diverse and clinically relevant populations. Once fully validated, the task could be instrumental in evaluating clinical populations with known deficits in social cognition, such as individuals with right hemisphere damage or neuropsychiatric conditions like autism spectrum disorders (ASD) or schizophrenia. These populations often exhibit impaired sarcasm recognition, which is closely tied to challenges in Theory of Mind and pragmatic language skills (Happé, 1993). By offering a standardized and accessible tool, this instrument could facilitate early detection and characterization of social cognitive impairments, providing valuable insights for tailoring therapeutic interventions. Additionally, its potential utility extends to monitoring the efficacy of rehabilitation programs aimed at improving social communication. Future research should prioritize testing the instrument in these clinical groups, ensuring its sensitivity, specificity, and cross-population adaptability to address the unique challenges faced by individuals with varying etiologies of social cognitive deficits.

Moreover, this new task could serve to investigate constructs that overlap sarcasm processing, such as empathy or mentalization capacity as taking others' perspectives. Additionally, this research could lead to a fruitful program to develop a comprehensive battery for behavioral and functional magnetic resonance imaging evaluation of pragmatic language subdomains. The sarcasm recognition task is well-suited for use in fMRI experiments . It is brief, features visual stimuli that can be presented during scanning, and requires only a button box for responses, which is compatible with fMRI protocols (Vassolo et. al., 2025). Considering these characteristics, the task could be employed in future studies to gather more data on the role of each cerebral hemisphere in processing non-literal language functions. In addition to assessing brain lateralization of sarcasm processing, future studies using our task could potentially identify specific areas responsible for this domain.

6. Conclusions

The sarcasm recognition task has high reliability and construct validity. The composed vignettes can be distinguished as sarcastic or literal above chance level and exhibit positive Cronbach's alpha reliability coefficients.

Regarding construct validity, the sarcasm recognition task aligns with gold standard pragmatic language tests from the Montreal Communication Evaluation Protocol. The association between the sarcasm recognition task and other pragmatic language tests pinpointed overlapping subdomains within the social cognition construct.

The result of the study is a novel instrument, whose design characteristics make it suitable for behavioral testing and it could be used in functional magnetic resonance imaging protocols. Its level of difficulty allows for avoiding ceiling effects, even in high-performing populations, and its simple vocabulary design without cultural uses allows it to be applicable to individuals from various socioeconomic backgrounds and educational levels.

Data availability

The data supporting the findings of this study are openly available at the Open Science Framework (OSF) repository, accessible at https://osf.io/ghrjc/?view_only=0d5ddb6ed088478dbf6734f966cdd242.

Author contributions

All authors contributed substantially to the conception and design of the study, data collection, analysis, and interpretation. Pablo Ocampo drafted the initial version of the manuscript. Mariana Bendersky and Lucy Alba-Ferrara conceived and supervised the overall study design, provided guidance on data analysis and interpretation and critically revised all versions of the manuscript for important intellectual content. Nicolas Vassolo and Bautista Elizalde Acevedo provided critical revisions and intellectual input. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Pablo J. Ocampo: Writing – original draft. **Nicolás Vassolo:** Formal analysis, Data curation. **Bautista Elizalde Acevedo:** Investigation, Formal analysis. **Mariana Bendersky:** Supervision, Conceptualization. **Lucía Alba-Ferrara:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology, 11*(2), 231–239. https://doi.org/10.1016/s0959-4388(00)00202-6

Alba-Ferrara, L., Hausmann, M., Mitchell, R. L., & Weis, S. (2011). The neural correlates of emotional prosody comprehension: Disentangling simple from complex emotion. *PloS One, 6*(12), Article e28701. https://doi.org/10.1371/journal.pone.0028701

Bambini, V., Gentili, C., Ricciardi, E., et al. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bullet, 86*, 203–216. https://doi.org/10.1016/j.brainresbull.2011.07.015

Beaudoin, C., & Beauchamp, M. H. (2020). Social cognition. In *Handbook of clinical neurology* (1st ed.,, *173*. Elsevier B.V.. https://doi.org/10.1016/B978-0-444-64150-2.00022-8

Bendersky, M., Lomlomdjian, C., Abusamra, V., Elizalde Acevedo, B., Kochen, S., & Alba-Ferrara, L. (2021). Functional anatomy of idiomatic expressions. *Brain Topography*. https://doi.org/10.1007/s10548-021-00843-3

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*, 2669–2683. https://doi.org/10.1016/j.neuropsychologia.2012.07.021

Bosco, F. M., Tirassa, M., & Gabbatore, I. (2018). Why pragmatics and theory of mind do not (completely) overlap. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01453

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *PNAS, 106*(14). https://doi.org/10.1073/pnas.0811884106

Davis, C. L., Oishi, K., Faria, A. V., Hsu, J., Gomez, Y., Mori, S., & Hillis, A. E. (2016). White matter tracts critical for recognition of sarcasm. *Neurocase, 22*(1), 22–29. https://doi.org/10.1080/13554794.2015.1024137

Del Ser, T., González-Montalvo, J. I., Martínez-Espinosa, S., Delgado-Villapalos, C., & Bermejo, F. (1997). Estimation of premorbid intelligence in Spanish people with the Word Accentuation Test and its application to the diagnosis of dementia. *Brain and Cognition, 33*(3), 343–356. https://doi.org/10.1006/brcg.1997.0877

Faust, M., & Mashal, N. (2007). The role of the right cerebral hemisphere in processing novel metaphoric expressions taken from poetry: A divided visual field study. *Neuropsychologia, 45*, 860–870. https://doi.org/10.1016/j.neuropsychologia.2006.08.010

Ferreres, A., Abusamra, V., Cuitiño, M., Côté, H., Ska, B., & Joanette, Y. (2007). Protocolo MEC. Protocolo para la evaluación de la comunicación de Montreal. *Buenos Aires: Neuropsi Ediciones*.

Filik, R., Ţurcan, A., Ralph-Nearman, C., & Pitiot, A. (2019). What is the difference between irony and sarcasm? An fMRI study. *Cortex; A Journal Devoted To The Study of The Nervous System And Behavior, 115*, 112–122. https://doi.org/10.1016/j.cortex.2019.01.025

Forgács, B. (2024). Meaning as mentalization. *Frontiers in Human Neuroscience, 18*, Article 1384116. https://doi.org/10.3389/fnhum.2024.1384116. Article.

Glenwright, M., & Pexman, P. M. (2010). Development of children's ability to distinguish sarcasm and verbal irony. *Journal of Child Language, 37*(2), 429–451. https://doi.org/10.1017/S0305000909009520

Hagoort, P., & Levinson, S. C. (2014). Neuropragmatics. In M. S. Gazzaniga, & G. R. Mangun (Eds.), *The cognitive neurosciences* (5th ed., pp. 667–674). Boston Review. https://doi.org/10.7551/mitpress/9504.003.0073.

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48*(2), 101–119. https://doi.org/10.1016/0010-0277(93)90026-R

Hauptman, M., Blank, I., & Fedorenko, E. (2023). Non-literal language processing is jointly supported by the language and theory of mind networks: Evidence from a novel meta-analytic fMRI approach. *Cortex; A Journal Devoted To The Study Of The Nervous System And Behavior, 162*, 96–114. https://doi.org/10.1016/j.cortex.2023.01.013

Hough, M. S. (1990). Narrative comprehension in adults with right and left hemisphere brain-damage: Theme organization. *Brain and Language, 38*, 253–277. https://doi.org/10.1016/0093-934X(90)90114-V

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

Laval, V., & Bert-Erboul, A. (2005). French-speaking children's understanding of sarcasm: The role of intonation and context. *Journal of Speech, Language, and Hearing Research, 48*(3), 610–620. https://doi.org/10.1044/1092-4388(2005/042

Matsui, T., Nakamura, T., Utsumi, A., Sasaki, A. T., Koike, T., Yoshida, Y., Harada, T., Tanabe, H. C., & Sadato, N. (2016). The role of prosody and context in sarcasm comprehension: Behavioral and fMRI evidence. *Neuropsychologia, 87*, 74–84. https://doi.org/10.1016/j.neuropsychologia.2016.04.031

Melton, L. M. (2016). If I don't laugh, I'll cry: Exploring humor coping in breast cancer. *Journal of Psychosocial Oncology, 34*(6), 530–541. https://doi.org/10.1080/07347332.2016.1233926

Monetta, L., Grindrod, C. M., & Pell, M. D. (2009). Irony comprehension and theory of mind deficits in patients with Parkinson's disease. *Cortex; A Journal Devoted To The Study of The Nervous System and Behavior, 45*(8), 972–981. https://doi.org/10.1016/j.cortex.2009.02.021

Nakamura, T., Matsui, T., Utsumi, A., Sumiya, M., Nakagawa, E., & Sadato, N. (2022). Context-prosody interaction in sarcasm comprehension: A functional magnetic resonance imaging study. *Neuropsychologia, 170*(March), Article 108213. https://doi.org/10.1016/j.neuropsychologia.2022.108213

Nelson, H. E., & Willison, J. (1991). *National adult reading test (NART)*. Windsor: NFER-Nelson.

Ouerchefani, R., Ouerchefani, N., Ben Rejeb, M. R., & Le Gall, D. (2024). Pragmatic language comprehension: Role of theory of mind, executive functions, and the prefrontal cortex. *Neuropsychologia, 194*, Article 108756. https://doi.org/10.1016/j.neuropsychologia.2023.108756

Papagno, C., & Romero Lauro, L. J. (2010). The neural basis of idiom processing: Neuropsychological, neurophysiological and neuroimaging evidence. *Italian Journal of Linguistics, 22*, 21–40.

Premack, D., & Woodruff, G. (1978). Does the chimpanze have a theory of mind? *The Behavioral and Brain Sciences, 4*, 515–526.

Repetti, R., Taylor, S. E., & Saxbe, D. (2007). The influence of early socialization experiences on the development of biological systems. In J. E. Grusec, & P. D. Hastings (Eds.), *Handbook of socialization: Theory and research* (pp. 124–152). The Guilford Press.

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

Rothermich, K., Ogunlana, A., & Jaworska, N. (2021). Change in humor and sarcasm use based on anxiety and depression symptom severity during the COVID-19 pandemic. *Journal of Psychiatric Research, 140*(May), 95–100. https://doi.org/10.1016/j.jpsychires.2021.05.027

Schmidt, G. L., & Seger, C. A. (2009). Neural correlates of metaphor processing: The roles of figurativeness, familiarity and difficulty. *Brain and Cognition, 71*(3), 375–386. https://doi.org/10.1016/j.bandc.2009.06.001

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

Shamay-Tsoory, S. G., Tomer, R., & Aharon-Peretz, J. (2005). The neuroanatomical basis of understanding sarcasm and its relationship to social cognition. *Neuropsychology, 19*(3), 288–300. https://doi.org/10.1037/0894-4105.19.3.288

Sotillo, M., Carretié, L., Hinojosa, J. A., Tapia, M., Mercado, F., López-Martín, S., & Albert, J. (2005). Neural activity associated with metaphor comprehension: Spatial analysis. *Neuroscience Letters, 373*(1), 5–9. https://doi.org/10.1016/j.neulet.2004.09.071

Stringaris, A. K., Medford, N. C., Giampietro, V., Brammer, M. J., & David, A. S. (2007). Deriving meaning: Distinct neural mechanisms for metaphoric, literal, and non-meaningful sentences. *Brain and Language, 100*(2), 150–162. https://doi.org/10.1016/j.bandl.2005.08.001

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

Valles-Capetillo, E., Ibarra, C., Martinez, D., & Giordano, M. (2022). A novel task to evaluate irony comprehension and its essential elements in Spanish speakers. *Frontiers in Psychology, 13*, Article 963666. https://doi.org/10.3389/fpsyg.2022.963666, 2022 Nov 22.

Vassolo, N., Ocampo, P. J., Elizalde Acevedo, B., Bosch, S., Bendersky, M., & Alba-Ferrara, L. (2025). Understanding Sarcasm's neural correlates through a novel fMRI spanish paradigm. *Brain Topography, 38*(4), 46. https://doi.org/10.1007/s10548-025-01118-x

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: The role of prosody and context. *Brain : A Journal of Neurology, 129*(4), 932–943. https://doi.org/10.1093/brain/awl032

Wellenzohn, S., Proyer, R. T., & Ruch, W. (2016). Humor-based online positive psychology interventions: A randomized placebo-controlled long-term trial. *Journal of Positive Psychology, 11*(6), 584–594. https://doi.org/10.1080/17439760.2015.1137624

Weschler, D., & Stone, C. R. (1945). *Instruction manual for the weschler memory scale*. New York Psychol. Corps.

Williams, J. A., Burns, E. L., & Harmon, E. A. (2009). Insincere utterances and gaze: Eye contact during sarcastic statements. *Perceptual and Motor Skills, 108*(2), 565–572. https://doi.org/10.2466/PMS.108.2.565-572

Winner, E., Brownell, H., Happé, F., Blum, A., & Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and Language, 62*(1), 89–106. https://doi.org/10.1006/brln.1997.1889

Zempleni, M.-Z., Haverkort, M., Renken, R., & Stowe, L. A. (2007). Evidence for bilateral involvement in idiom comprehension: An fMRI study. *Neuroimage, 34*, 1280–1291. https://doi.org/10.1016/j.neuroimage.2006.09.049