Indirect speech acts and Theory of Mind

The role of conventionalization and mind-reading abilities in understanding indirect speech acts

Eleonora Marocchini
XXXIV Cycle - Psychology and Cognitive Science

Introduction

Some of the things we say do not describe a state of things, nor the action of doing something: they actually perform that action. For example, by saying, "*Open the window, please*", I am performing an act: the speech act of requesting. Of course, there are many ways of performing the same speech act, and most of them would be indirect, like, for instance, "*Could you open the window?*", i.e., a question about the interlocutor's ability to open the window, or, even more indirect, "*It is hot in here*", i.e., an assertion about the temperature in the room.

Indirect requests have been widely investigated through empirical studies testing the main theoretical accounts about their processing. Nevertheless, only a few studies have investigated the cognitive functions underlying indirect speech acts comprehension and their possible correlation with the level of indirectness of the speech act. The aim of this project is to address these points, with particular regard to Theory of Mind, through a series of studies on children with Autism Spectrum Disorder, typically developing children and neurotypical adults, in the framework of behavioral Experimental Pragmatics.

Background

Since Searle's (1979) seminal work, theories on indirect requests (IRs) traditionally considered any request that is not typically performed through an imperative to be indirect. As a consequence, for years, the literature prominently focused on interrogative IRs of the same kind of the one's used by Searle as *the* example of an indirect speech act, i.e., an interrogative utterance, namely (1), which directly performs the illocutionary act of a question and indirectly conveys a request:

(1) Can you pass the salt?

IRs like (1) are said to be *conventionalized* IRs: along with a *convention of means* to refer to the ability of the speaker to perform the desired action, they follow a *convention of form* common enough to be defined idiomatic (Clark, 1979). However, requests can also be performed with highly indirect forms, such as:

(2) This soup is insipid.

These are known as highly IRs or hints. They usually follow a *convention of means* to state a fact about the world from which the hearer should infer that some action would be desired (Gibbs, 1981).

There are also "intermediate" forms, that follow the same *convention of means* of *conventionalized* IRs (i.e., to refer to the ability of the speaker to perform the desired action – or, to be more precise, to the possibility of the desired action to be performed) but not the same *convention of form*, such as:

(3) Is it possible to bring the salt?

These are considered *non-conventionalized* IRs, but they are not highly indirect.

To sum up, four main levels of (in)directness have been found in the literature on ISAs (Blum-Kulka, 1987; Ruytenbeek et al., 2017):

1. direct requests (DRs): *Open the window*;
2. conventionalized indirect requests (CIRs): *Can you open the window?*
3. non-conventionalized indirect requests (NCIRs): *Is it possible to open the window?*
4. highly indirect requests (HIRs): *It is very hot in here.*

The central prediction stemming from this scale is that "the more 'indirect' the mode of realization, the higher will be the interpretative demands on the hearer" (Blum-Kulka, 1987, p. 133). Traditionally, higher interpretative demands, i.e., higher needs of inferential processing, are thought to be reflected in higher cognitive costs (see Gibbs, 2002, for a review). This assumption has been shown to be true by some experimental studies on ISAs (Clark, 1979; Clark & Lucy, 1975; Ruytenbeek et al., 2017), though the picture on ISAs processing is yet to be completely defined.

Conventionalized IRs

Consistently with the theoretical debate, the early experimental literature particularly focused on conventionalized IRs (Clark & Lucy, 1975; Gibbs, 1983). When tested within the same context, *Can you…?* forms seem to show a difference in processing as compared to other non-conventionalized forms: (i) they are easier to process and/or interpret as a directive (Clark & Lucy, 1975; Deliens et al., 2018; Ruytenbeek et al., 2017); (ii) they seem to be biasing towards a directive interpretation even when used in their literal sense (Gibbs, 1983). These findings have been interpreted as evidence in favor of different processing procedures underlying conventionalized and non-conventionalized IRs.

What would the processing procedure for conventionalized IRs (CIRs) be? One account would be the *short-circuited implicature* (Morgan, 1978) or, more recently, *pragmatic routine* (Vega Moreno, 2017) account, holding that CIRs might be understood by compiling or short-circuiting the inferential steps involved in comprehension. The idea is that the frequency of use of a certain word or phrase with a specific function would make the hearer "derive roughly the same implications and enrich the explicit content in roughly the same ways", generating "highly activated assumptions", i.e., those which have been previously derived a reasonably high number of times (Vega-Moreno, 2007: 118). According to this view, the hearer would follow the inferential stages needed to interpret the utterance much faster (i.e., through a familiar inferential path) than they would do for creative, unfamiliar, highly indirect forms – but the inferential stages would stay the same per se, making the indirect interpretation of CIRs a *strong implicature*, but still an implicature (cfr. Ruytenbeek, 2019).

Conversely, Groefsema (1992, 1995) analysed the request assumption conveyed with *Can you…?* forms to be bound to the lexical meaning of the form: *Can* would have a "request" meaning as a direct development of its unitary meaning, guided by the communicative principle of relevance. This would make *Can you…?* forms a higher-order explicature, derived after enrichment of the logical form of the sentence uttered, just like it would happen with an imperative - and not an implicature. This approach suggests that interpreting CIRs would not need particular inferencing abilities – but we will get back to that.

Highly Indirect Requests

More recently, studies on adults looked at highly IRs, too (see Ruytenbeek, 2017, for a review). As these would need higher inferencing abilities, they are particularly interesting to study in development. However, when it comes to (a)typical development, the picture is still fragmented.

The few experimental works on the development of IRs suggest an understanding of IRs as early as 2;6 years of age (Reeder, 1980; Shatz, 1978), and a developmental pattern of different IRs forms between the age of 3 and 7 (Bernicot & Legros, 1987; Bucciarelli et al., 2003; Carrell, 1981). Interestingly, Kissine et al. (2015) showed that children between 2;7 and 3;6 years of age exhibit difficulties with IRs. According to the authors, understanding IRs might require theory of mind (ToM), and children's behaviour might be bound to a lack in this ability.

This last note partially suggests why the literature on IRs comprehension in Autism Spectrum Disorder (ASD) provides mixed evidence as well: ASDs' impairments in some domain of pragmatics have traditionally been attributed to a deficit in ToM, that is also known to be compromised in ASD, and which could impede inferring the others' intentions (Simon Baron-Cohen et al., 1985; Simon Baron-Cohen, 2000). In fact, there is evidence suggesting that ASDs have difficulties in understanding the intention of IRs (MacKay & Shaw, 2004). Counterevidence is also available. In fact, Kissine et al. (2012, 2015) tested IRs comprehension in ASD as well, and found that ASD children can comply with IRs at an above chance level. In the 2015 study, HFAs performed even better than the TD participants (who, as mentioned, were aged between 2;7 and 3;6).

However, Kissine et al. (2015) did not collect any ToM measures. More generally, despite an increasing interest in IRs and cognitive functions, no study to date on the development of IRs comprehension measured ToM, nor any other cognitive function that might play a role. Moreover, Kissine et al (2015) only tested highly IRs. Therefore, the conflicting results available in the literature might suggest that the level of conventionalization impacts IRs comprehension.

Overall, then, the experimental literature on IRs is still limited and mixed evidence has been provided. In particular, no study has directly addressed the issue of the different levels of conventionalization in IRs and the relationship with cognitive functions.

IRs comprehension and the Linguistic- vs. Social-Pragmatics divide

A recent proposal by Andrés-Roqueta and Katsos (2017; 2020), suggests a distinction between Linguistic- and Social-Pragmatics. In their view, Linguistic-Pragmatics on one hand includes those pragmatic tasks whose comprehension relies on the hearer's egocentric point of view: linguistic abilities (e.g., lexical and morpho-syntactic competence) and basic knowledge of pragmatic norms would suffice to succeed such a pragmatic task (e.g. interpreting a scalar implicature such as *Some students passed the test* as meaning *Not all the students passed the test*).

Social-Pragmatics, on the other hand, includes those pragmatic tasks whose comprehension depends mainly on perspective-shifting skills: linguistic abilities and knowledge of the basic pragmatic norms seem not to be sufficient for ensuring comprehension of a pragmatic task requiring the comprehender to infer the speaker's mental state (e.g. irony).

The authors underline that this distinction is task-specific, rather than phenomenon-specific, and this is particularly crucial with regard to IRs: in fact, (1) conventionalized IRs might fall into the Linguistic-Pragmatics category, while at least some highly IRs (depending on the task) might fall into the Social-Pragmatics category; (2) collaborative experimental scenarios might facilitate default responses to IRs, regardless of their level of conventionalization (explaining part of the mixed results found in the literature); (3) individual differences might play a role in the kind of interpretive strategy for IRs comprehension depending on task-specific characteristics.

Aims

In this very short introduction I tried to summarize and underline the gaps in the IRs comprehension literature that I am trying to address with the studies that I had/have planned to conduct, namely: (i) whether IRs' comprehension is compromised in ASD; (ii) whether IRs comprehension in (a)typical development varies depending on the level of conventionalization of the request; (iii) whether cognitive functions – especially ToM abilities - play a role in IRs comprehension; (iv) whether different interpretive strategies could be at stake depending on the level of conventionalization of the IR. These last two questions will be addressed with studies on the adult population as well.

Reasons

Trying to participate in filling some the existing gaps in the literature on a specific phenomenon is a reasonable objective per se, but the existence of a gap does not always call for further investigation – and filling it does not always have potential consequences on the literature concerned with other phenomena.

In this case, ToM's role in understanding pragmatic phenomena is largely discussed in the Experimental Pragmatics framework (and in ASD research). Traditionally, from Sperber and Wilson's seminal work, pragmatics is viewed as a theory of the speaker's meaning in context, requiring the interlocutor to recognize and attribute intentions to the speaker: this would make a sub-module of ToM (Sperber & Wilson, 2002). More recently, other proposals mitigated this idea that pragmatics relies on ToM skills completely (Bosco et al., 2018; Domaneschi & Bambini, 2020). Andrés-Roqueta and Katsos (2017, 2020) Linguistic vs. Social Pragmatics distinction and the experimental evidence they provide from recent studies on pragmatic phenomena that do not seem to require advanced ToM skills constitute a step in this direction: observing whether the comprehension of the same phenomenon can rely on ToM or not depending on some specific characteristics of the instance of the phenomenon at stake (in this case, for

instance, conventionalization) or individual differences of the interlocutors, could contribute to the puzzle of ToM's relationship with pragmatics.

Timeline, methods and experimental design

Timeline

The timeline of the present project has undergone substantial changes due to the covid-19 pandemic. The new timeline can be seen in Figure 1, and the following sections will describe the originally planned and the actually conducted experiments.

In particular, while Experiment 1 was conducted during my first year and was therefore not affected by the covid-19 pandemic, I refer to the originally planned Experiment 2 as (Ex)Experiment 2, as it had been planned to be conducted in a kindergarten - which has obviously been closed on March 8th, 2020. At that time, I was still conducting the pilot study (which I will briefly present).

As a consequence, I have come up with an online experiment that I could conduct online during the pandemic, and that is what I refer to with Experiment 2. Experiment 3 will be a follow up to be conducted in person according to the present covid-19 restrictions and guidelines - if the sanitary conditions allow for it.

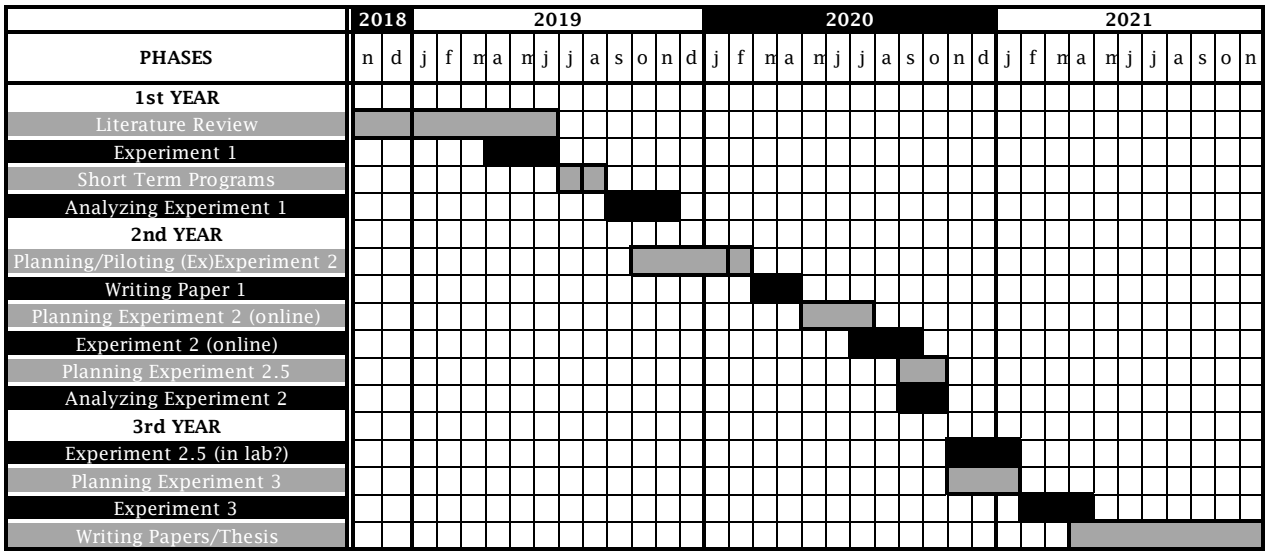

Figure 1: Timeline for the present project.

Experiment 1, as described in the following section, has been presented as a poster both at the XVI Italian Association of Cognitive Science (AISC) Annual Conference in Rome (December 11-13, 2019) and at the International Max Planck Research School (IMPRS) Conference that should have been hosted in Nijmegen (June 05-07, 2020), ultimately held online. During the lockdown, the paper has been written and carefully revised by all authors and subsequently submitted to a journal during the Italian lockdown. It is currently under review as Marocchini, E., Di Paola, S., Mazzaggio, G., and Domaneschi, F. *Understanding Indirect Requests in Children with High-Functioning Autism*.

Experiment 1

The first experiment of the project is a study focusing on HIRs in atypical development. As HIRs are understudied even in typical development, the experiment included two different kinds of HIRs, where one would need higher inferential abilities to be understood, as well as two groups of TD children of different age ranges.

Research Questions

1. Is the ability to comprehend IRs compromised in HFA? Are there any differences depending on the level of indirectness of the request (e.g., DRs, HIRs, even more indirect HIRs), i.e., on the amount of inferential work required by the request?
2. Does ToM play a role in understanding IRs, i.e., does the impairment in ToM affect HFAs' (un)preserved IRs comprehension?

Predictions

1. Since the earliest studies on IRs in adults (Clark and Lucy, 1975; Clark, 1979) and typical development (Shatz, 1978; Carrell, 1981) have shown that IRs are generally more difficult to understand than DRs, we expected both TDs and HFAs to have more difficulties understanding IRs (HIRS in particular) than DRs (i.e., effect of condition).
2. On the basis of the mixed results provided by previous studies on IRs comprehension in HFA, showing difficulties in understanding the use of ISAs in general (MacKay and Shaw, 2004), but no real problems with CIRs (Paul & Cohen, 1985), we expected a difference between TDs and HFAs in those cases (HIRs) involving higher inferential abilities.
3. Following the few studies on IRs comprehension in development targeting different age ranges (Carrell, 1981; Bernicot and Legros, 1987; Bucciarelli et al., 2003), we expected differences between the two TD groups, reflecting a developmental pattern: the younger TD group should have more difficulties in complying with IRs in general, particularly with those requiring higher inferential abilities, due to their less developed general cognitive functioning.
4. If HIRs belong to the Social Pragmatics type of pragmatic tasks (Andrés-Roqueta and Katsos, 2017), i.e., understanding IRs draws on mind-reading abilities, ToM might play a role for their comprehension. The greater the inferential work, the greater the involvement of ToM.

Method

Participants. The participants were 61 Italian monolingual children between 5;2 and 12 years of age, divided in 3 groups. 14 HFA children (2 F; age range: 9-12 years; mean age=10.6; SD=1.17), who received a diagnosis of HFA (according to the DSM-IV criteria) by a team of trained neuropsychologists; 28 age-matched TD children (11 F; age range: 9-12 years; mean age=11,03; SD=0.61); 19 younger TD children (6 F; age range: 5;2-6;3 years; mean age=5,35; SD=0.48). The participants were recruited through an ASD support center and two schools.

Experimental Design. The experiment had a 3x3 latin square within-subjects design. Requests were presented as DRs, HIRs (IND, in the graph), and more indirect HIRs (HIND, in the graph).

Task. Participants were asked to help the experimenter recreate a drawing of a farm. The experimenter would request their help through either DRs (e.g., *What colour are the pigs?*), HIRs (e.g., *I don't remember the colour of the pigs*) or even more indirect HIRs (e.g., *The colour of the pigs is hard to remember*). Accuracy was coded in terms of compliance to the directive interpretation (e.g., in this case, the only answer that would have been coded as accurate was saying that they were pink). Each child saw 4 items per condition and therefore had a 0-4 composite score per condition.

Other measures. Both language and ToM skills of the children participating in the study were tested through standardized tests, namely the BVL (Batteria per la Valutazione del Linguaggio) for their morphosyntactic and lexical skills (Marini et al., 2015) and two false belief tasks from The Theory of Mind Task Battery for their first and second order ToM skills (Hutchins & Prelock, 2014).

Outcomes

Children's accuracy in the IRs comprehension task is displayed in Figure 2 through their composite score. Children's scores in the BVL test and in the two ToM tasks, together with the composite (0-2) ToM scores, are reported in Table 1.

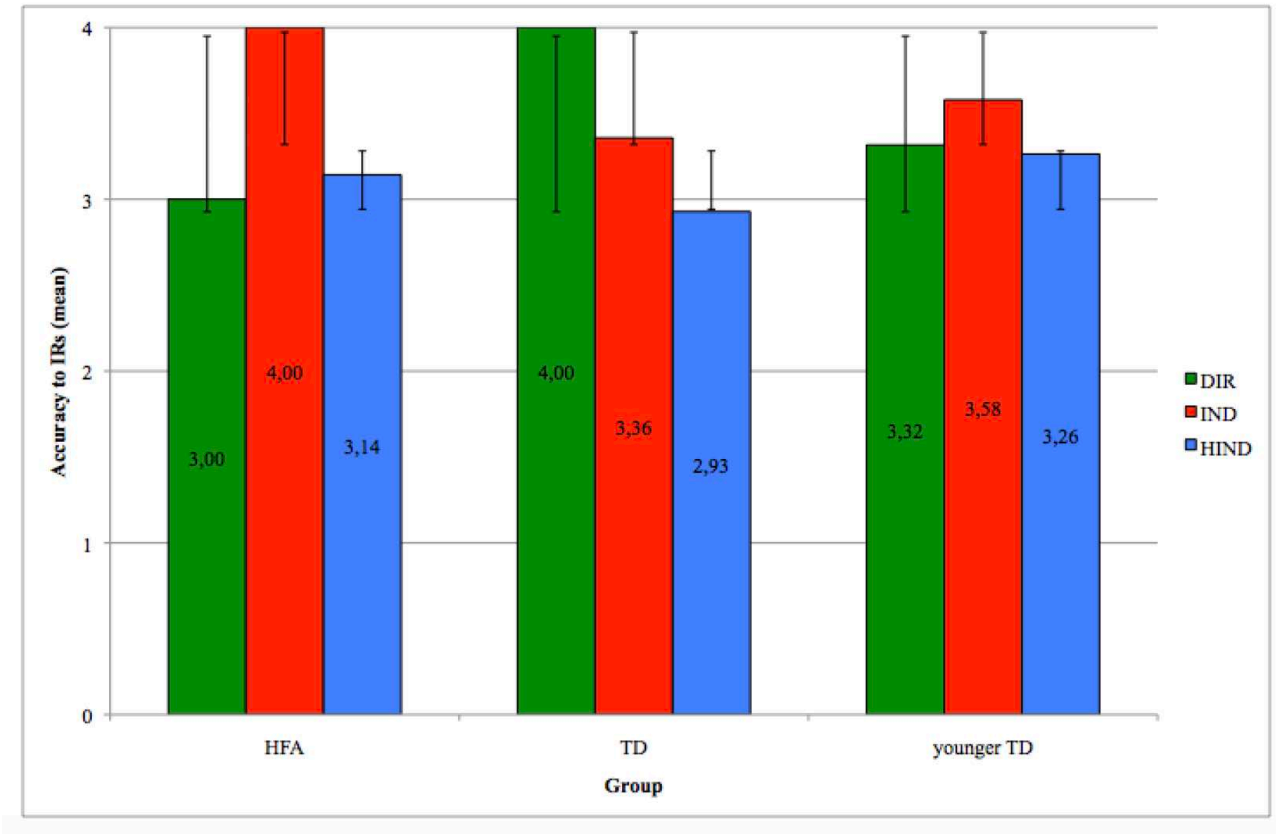

Figure 2. Mean composite score for the accuracy in the IRs task in each experimental group and condition. Error bars indicate the standard deviation.

	TDs	HFAs	Younger TDs
BVL score	37.3 (2.32)	30.1 (6.14)	29.2 (3.75)
First-order ToM	0.92 (0.26)	0.50(0.51)	0.57(0.50)
Second-order ToM	0.71(0.46)	0.35(0.49)	0.05(0.22)
ToM composite score	1.64 (0.62)	0.85 (0.66)	0.63 (0.59)

Table 1. Mean score (and SD) of each experimental group in the tasks assessing for linguistic (BVL) and mind-reading abilities (first- and second-order ToM; and ToM composite score).

As Figure 2 suggests, all children performed well in the task, but their performance varied depending on condition and group.

The analyses were conducted through Linear-Mixed Models statistics (LMMs), using the packages lme4 (Bates et al., 2015), lmerTest (to provide F statistics with degrees of freedom), and emmeans (for Tukey contrasts for post-hoc comparisons) in R. The fixed effects structure of the LMM model included Group (HFA, TD, and younger TD) and Condition (DIR, IND, HIND), and their interactions. The random structure of the model included random intercepts for subjects.

The analyses had the following outcomes.

1. Children's accuracy significantly differed depending on the Condition only ($F(2, 116)=4.47$; $p=.01$). Pairwise comparisons for the main effect of condition revealed that children provided significantly more correct responses to the first type of HIRs (IND in Figure 2) than to the more indirect (HIND in Figure 2) HIRs ($b=-0.53$; $SE=0.18$; $DF=116$; $t=-2.92$; $p=.01$). No other significant differences between conditions emerged. The main effect of Group was not significant, either ($F(2, 58)=0.02$; $p=.97$).
2. Condition X Group emerged as a significant interaction ($F(4, 116)=4.39$; $p=.002$). Pairwise comparisons showed that older TDs exhibited higher accuracy with direct than both IND ($b=0.64$; $SE=0.25$; $DF=116$; $t=2.51$; $p=.03$) and HIND requests ($b=1.07$; $SE=0.25$; $DF=116$; $t=4.19$; $p=.0002$), while HFAs showed an opposite pattern: they gave significantly more correct responses to IND than both HIND ($b= -0.85$; $SE=0.36$; $DF=116$; $t=-2.37$; $p=.04$) and direct requests ($b =-1$; $SE=0.36$; $DF=116$; $t=-2.77$; $p=.01$).
   No significant differences emerged for the younger TDs.

The BVL and ToM scores were analyzed with the Kruskal-Wallis rank sum test statistics, with Dwass-Steel-Critchlow-Fligner contrasts for pairwise comparisons.

1. The BVL test showed significant group differences (KW $\chi 2(2)=35.5$; $p<.001$). Pairwise comparisons revealed that children in the older TDs group scored significantly higher than both HFAs ($W= 5.91$; $p<.001$) and the younger TDs ($W=-7.73$; $p<.001$). No differences emerged between HFAs and younger TDs.
2. Significant group differences emerged both in the 1st-order (KW $\chi^2(2)=11.1$; $p=.004$), and 2nd-order ToM task (KW $\chi^2(2)=20.3$; $p<.001$). Pairwise comparisons revealed that older TDs scored significantly higher than the younger TDs in both the 1st-order ($W=-4.02$; $p=.01$) and 2nd-order tasks, and significantly higher than HFAs in the 1st-order task ($W=4.45$; $p=.005$), while the difference only approached significance ($W=3.11$; $p=.07$) in the 2nd-order task, probably due to small sample sizes. No difference between HFAs and younger TDs was detected in the ToM tasks.
3. ToM composite scores revealed significant group differences (KW $\chi^2(2)=23.1$; $p<.001$). Pairwise comparisons confirmed that older TDs exhibited overall better ToM skills than both HFAs ($W=4.84$; $p=.002$) and the younger TDs ($W=-6.27$; $p<.001$), while there was no difference between HFAs' and younger TDs' scores.

LMM statistics was used for the analysis of predictors as well. Accuracy in the IRs task was the outcome variable and BVL and ToM scores were treated as predictors, along with Condition, Group, and the resulting interactions.

1. Children's composite scores for ToM marginally predicted their accuracy in the IRs task ($b=0.87$; $SE=0.50$; $DF=140.80$; $t=1.76$; $p=0.08$), and to the most highly indirect (HIND) requests in particular (Cond HIND:ToM: $b=-1.13$; $SE=0.61$; $DF=104$; $t=-1.83$; $p=.07$), while their BVL scores did not.
2. Accuracy to the HIND requests was significantly predicted by the older TDs composite ToM scores (Cond HIND X Group TD X ToM: $b=1.64$; $SE=0.74$; $DF=104$; $t=2.19$; $p=.03$).

Overall, then, these data suggest different patterns of IRs comprehension in the three groups of participants, and that ToM seems to play a role in IRs comprehension only for older TDs (as compared to HFAs and younger TDs) and for the most highly indirect requests.

With regard to the first research question, about the HFAs' ability to comprehend IRs, they performed well in either of the two IRs types presented. This result conflicts with our original prediction, though it fits with some of the previous studies on the topic (MacKay & Shaw, 2004; Kissine et al., 2012; Kissine et al., 2015; Deliens et al., 2018).

Nevertheless, the comprehension patterns of the HFA group differed from those of the other two. Older TDs interpreted direct requests at ceiling, while showing difficulties with both HIRs.

This is in line with our prediction that the higher inferential complexity of IRs makes it more difficult to understand them and with earlier studies on typical development (Bucciarelli et al., 2003; Bernicot et al., 2007). On the contrary, HFAs provided more correct responses to IRs (though not to the

most highly indirect ones) than to direct requests. There are two possible explanations for this: (i) the experimental scenario might have suggested a default response to IRs, as it was very cooperative, and a lack of attention during the session might have resulted in lower accuracy to direct requests; (ii) HFAs might have used an alternative strategy to interpret IRs that could rely on lexical cues rather than on an inferential path. For instance, in the IR *I don't remember the color of the grass,* the HFAs' response might have been triggered by the word *color* alone. This would be in line with the results of some previous studies like Ozonoff & Miller's (1996), where HFAs interpreted more often a request as indirect than direct even in contexts supporting the interpretation of the utterance as a question (e.g., *Can you water the lawn?* in a phone conversation about the lack of water). Others have suggested that HFAs could not rely on complex interpretive strategies for IRs comprehension (see Paul and Cohen, 1985; Kissine et al., 2012) – and this would not be. For the same reason, this might have worked well with one type of HIR, but not with the other, as HIRs such as *The color of the grass is hard to remember* are more obscure than HIRs such as *I don't remember the color of the grass*, due to their inferential load but also to their wording.

Younger TDs, instead, showed high accuracy rates but no differences among conditions. They seemed not to be sensitive to the manipulation. They might have used a similar simplified (lexical) comprehension strategy, facilitated by the cooperative nature of the game/experiment. This is in line with previous findings on TD preschoolers basing their interpretation on context, regardless of the linguistic form of the request (Bernicot & Legros, 1987).

In sum, different interpretive strategies might be at stake. School-aged TDs would follow a genuinely inferential strategy corresponding to the IR complexity; children with more limited cognitive functioning, either due to a neurodevelopmental disorder or because of their developmental phase would rely on simpler strategies based on linguistic cues.

With regard to ToM skills, they seem to enhance the HIRs comprehension for older TDs' only. This fits perfectly both with the idea that the different groups relied on different interpretive strategies to comprehend HIRs and with the idea that the most highly indirect requests were the most difficult to process.

Older TDs, though, were the only ones to show that they neatly perceived the manipulation, i.e., to show high sensitiveness to the different amount of inferential work needed to interpret the requests. For this reason, it makes sense to speculate that they genuinely used a mind-reading based interpretative strategy, while the other two groups probably did not. On this note, previous literature on HFAs pragmatic skills have shown that they can compute pragmatic inferences but they do so also in contexts in which they are not expected to (see Hochstein et al., 2017).

Overall, then, the idea that individual factors might influence the interpretative strategies at stake for the same pragmatic phenomenon seem to be plausible. In particular, it is possible that HFAs and younger TDs rely more on strategies associated with the Linguistic-Pragmatics realm to understand IRs, while older TDs are more prone to using strategies linked to Social-Pragmatics.

(Ex)Experiment 2

This last note allowed for further speculations on the possibility that different interpretive strategies might be at stake in understanding IRs, depending on individual differences and on the different levels of indirectness. Therefore, the second experiment of the project still focused on HIRs while taking NCIRs into account as well, and digging into similar research questions - on typical development only.

As mentioned earlier, this experiment could not be completed because of the covid-19 pandemic. However, a small pilot study has been conducted before the lockdown, along the following lines.

Research Questions

1. Does children comprehension vary depending on the level of conventionalization (CIRs vs. NCIRs vs. HIRs), with CRs being easier to comprehend than NCIRs and NCIRs being easier to understand than HIRs?
2. Is there a developmental path in IRs comprehension (i.e., group differences), with younger children having more difficulties than older children?

3. Does ToM play a role in IRs comprehension (i.e., will ToM skills be able to predict IRs comprehension)?

Predictions

1. If previous studies on CIRs in adults (Clark and Lucy, 1975; Ruytenbeek et al., 2017) and typical development (Carrell, 1981) have shown that NCIRs and HIRs are generally more difficult to understand than CIRs, then it is plausible that children will have more difficulties understanding HIRs than CIRs and NCIRs (i.e., effect of condition). However, this effect is expected to be more significant for HIRs than NCIRs, as conventionalization is probably more relevant for adults.
2. If the only study on IRs comprehension in development targeting different age ranges (Carrell, 1981) is confirmed, a developmental path will probably prove to be present, with HIRs being comprehended easily by the 5-6 year-old group and less by the other two groups, while CIRs and possibly NCIRs are expected to be comprehended by all groups, as they require lower inferential abilities (and, possibly, linguistic abilities only).
3. If HIRs belong to the Social Pragmatics type of pragmatic tasks (Andrés-Roqueta and Katsos, 2017) ToM might play a role for their comprehension, though language abilities will certainly be needed, confirming previous studies on other non-literal phenomena (Norbury, 2005, for metaphor). This would mean that both linguistic and ToM skills should be able to predict children's comprehension of HIRs, with children with higher linguistic and ToM skills having better understanding of HIRs.

Methods

Participants. The participants had to be 60 TD Italian children, divided in three groups: 3-4 year olders (20), 4-5 year olders (20), and 5-6 year olders (20), recruited in a kindergarten.

Experimental Design. The experiment had a 3x3 latin square within-subjects design, with 2 independent variables: age group and conventionalization of IRs. IRs were presented in 3 experimental conditions: CIRs, NCIRs, and HIRs. A control condition consisting in an assertion with no directive interpretation was also added, to ensure that compliance with requests was not due to a bias-to-action that younger children might present.

Task. The task was an act-out task. Children were shown 30 boards, one at a time, with 6 cards on each of them. The cards either showed animals, plants, or food (see Figure 3). Children were asked to help the experimenter removing some of the cards from each board and divide the removed cards into three groups, depending on their nature.

Figure 3. Example of a board with cards presented to the child.

The experimenter's instructions to remove the cards presented the manipulated variable (e.g., CIRs: *Can you remove the cats?*; NCIRs: *Is it possible to remove the cats?*; HIRs: *There are too many cats here*; control assertion: *There are no dogs here*). The target items (n=12) showed two identical cards. Removing the cards mentioned by the experimenter was considered a directive interpretation, while replying verbally was considered a non-directive interpretation, following previous studies on adults (Deliens et al., 2018; Ruytenbeek et al., 2017).

Accuracy was coded in terms of compliance to the directive interpretation. Any completed action was scored 1 and any attempted action or verbal answer was scored 0.

To prevent the creation of alternative strategies based on clearly identifiable patterns, there were cases where the requested cards were not on the board (n=8), filler questions about the presence of certain cards (n=4), and cases where the experimenter asked the child to give her one card and collaborated in the division of the cards (n=6).

After each trial, children were instructed to put the removed cards into three empty boards, surrounded by a recognizable pattern: the animals into the fence, the plants on the grass, and the food on the table. This division was created for two reasons: (i) to give the children a clear objective of the game and (ii) to collect a measure of attention.

Other measures. After the task, a battery of linguistic and neuropsychological tests was administered. The battery included: the BVL (Batteria per la Valutazione del Linguaggio) for their morphosyntactic and lexical skills (Marini et al., 2015); the Diverse Desire, Diverse Belief, Knowledge Access, and Contents False Belief tasks from the ToM scale by Wellman and Liu (2004), and two False Belief tasks from Hutchins and Prelock's (2014) battery for ToM; and two measures of executive functions, inhibition (Ponitz et al., 2008) and working memory (Hughes & Ensor, 2010).

Outcomes

The results of the pilot study on the 8 children (4 F; age range: 4;5-5;2 years; mean age=4.87; SD=0.23) I succeeded in testing before the lockdown in March were promising.
They suggested that an effect of condition would indeed have emerged, as Figure 4 seems to show.

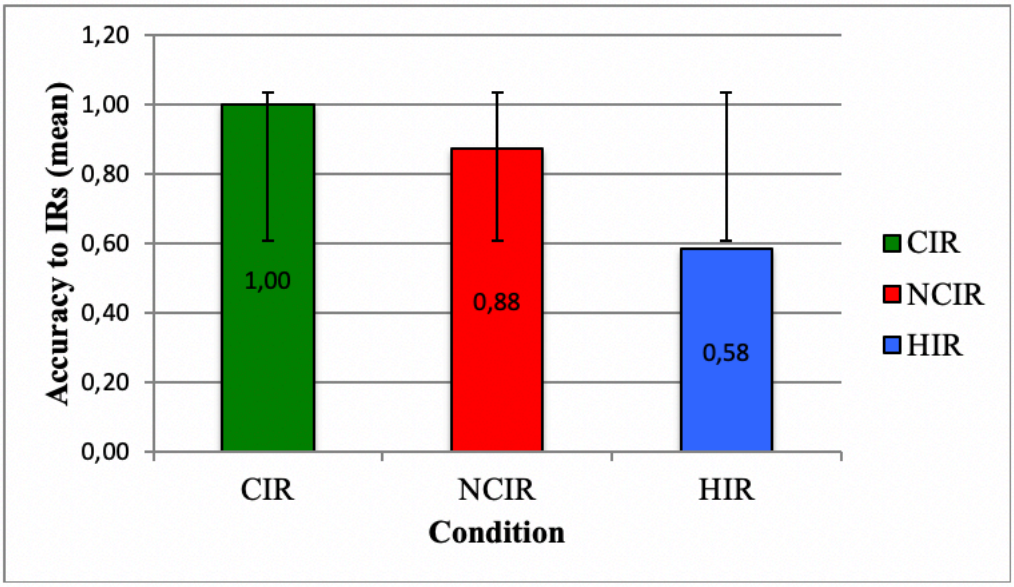

Figure 4. Mean score for the accuracy in the IRs task per condition. Error bars indicate the standard deviation.

Unfortunately, these data are not enough to even speculate on the foreseen effect, and the kindergarten was closed up until now. I therefore turned to the more reachable adult population for the actual Experiment 2, which I will present in the next section.

Experiment 2

The actual second experiment of the project has involved neurotypical adults and, due to the restrictions that the covid-19 pandemic has imposed on research, has been conducted online. It investigated CIRs through reaction times, measured with OpenSesame's (Mathôt et al., 2012) online interface, OsWeb, hosted on JATOS' platform (Lange et al., 2015), and programmed in JavaScript – though mostly through OpenSesame's drag-and-drop python-based interface.

Research Questions

1. Is the comprehension strategy at stake in conventionalized IRs processing more lexically-based or more inferential?
2. Do ToM abilities play a role in conventionalized IRs processing?

Predictions.

1. If CIRs are comprehended on a more *lexically-based* strategy rooted in the *Can you…?* form, we could expect this to be visible (in terms of shorter reading times) right from the start of the expression, when the context is biasing towards a directive interpretation – as compared to their non-directive counterparts.
   Conversely, if understanding CIRs requires a more *inferential* strategy built on a path of implicatures (even though short-circuited), we should observe longer reading times (as compared to their non-directive counterparts) and, possibly, a wrap-up effect in the reading times at the end of the sentence.
2. If CIRs can be interpreted through a lexically-based strategy, ToM should not be needed to get to their directive interpretation; this would, in turn, mean that CIRs could belong to the Linguistic Pragmatics type of pragmatic tasks (Andrés-Roqueta and Katsos, 2017). Conversely, if additional inferential complexity is added, for instance through irony or sarcasm, ToM could probably play a role in understanding the same utterance as ironic/sarcastic (resulting in shorter reaction times to sarcastic items for adults with higher ToM skills).

Methods

Participants. The participants were 91 neurotypical adults, Italian native speakers, recruited online (59 F; age range: 22-59 years; mean age=35.85; SD=9.85). Informed consent was obtained from all participants at the beginning of the task.

Experimental Design. The experiment had a 1x3 latin square within-subjects design with context of the stories as an independent variable. The manipulation of the context created 3 experimental conditions: (i) stories with a context biasing towards an indirect, directive interpretation (i.e., as a request to the interlocutor to perform an action), (ii) stories with a context biasing towards a direct, non-directive interpretation (i.e., as a question about the interlocutor's ability to perform an action), and (iii) stories with a context biasing towards a sarcastic interpretation (i.e., as a joke about the interlocutor's ability or situation – the 'requested' action cannot be performed, or the interlocutor clearly does not want to perform it).

The 12 target sentences were independently tested and chosen among 25 questions through a between-subjects norming study. The norming study was also conducted online and the 106 participants (70 F; age range: 21-53 years; mean age=29.35; SD=6.20) were assigned to one of two groups: one group of participants was asked how probable it was (on a likert scale from 1 to 7) for the questions at stake to mean their directive interpretation, and the other was asked how probable it was for them to mean their non-directive interpretation. The 12 sentences selected for the task were the ones that received the highest and most similar rating in the two conditions (directive and non-directive).

Task. Participants were presented with 12 target stories (4 stories per condition) and 21 filler stories (7 ending with an imperative, 7 ending with a polar or a wh- question, and 7 ending with an assertion).

Each target story ended with a conventionalized IR (e.g., *Can you take notes?*). Participants read each story through self-paced reading: they had to press the space key on their keyboard to move on to the next part.

The story was divided into 3 parts: 2 context parts, and the target sentence (i.e., the conventionalized IR), as shown in Figure 5.

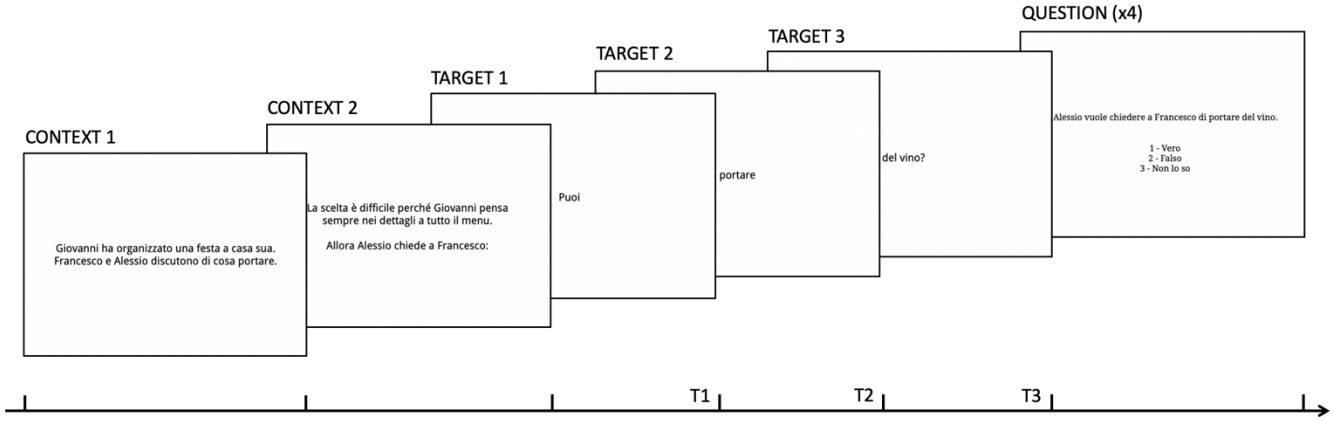

Figure 5. An example of screenshots from the IR task in the indirect, directive condition. The story screens lasted until participants pressed the space key; the (4) questions screens lasted until participants made their choice by pressing one of the available response keys (1; 2; or 3).

The last line of each story, i.e., the target sentence, was presented through self-paced region-by-region reading with noncumulative display (i.e., where only one region at a time is on screen: *Can*, then VP, then N). This choice is due to the particular importance of the initial analysis process with regard to the research questions at stake: there is experimental evidence available that participants self-pacing through a cumulative reading paradigm can develop the strategy of repeatedly pushing the button until they can see the whole sentence (Ferreira & Henderson, 1990; Just et al., 1982) – which results in a limited informativeness of the task in terms of on-line processes reveal. Conversely, the literature seems to suggest that self-paced region-by-region reading with noncumulative display can provide results that are comparable to eye-tracking paradigms, at least when the region is one-word only, i.e., the self-paced presentation is actually word-by-word (Ferreira & Clifton, 1986; Ferreira & Henderson, 1990, 1991).

Reading times were therefore calculated for each region (T1, T2, and T3) and for the whole utterance in all conditions. After reading it, participants were presented with four (true-false judgment) questions about the story, to make sure participants were paying attention to the story (and were aware of the main point introduced in the story by the experimental condition at stake). Participants gave their answers by pressing one of the available response keys (1 for *true*; 2 for *false*; or 3 for *I don't know*).

Other measures. ToM skills of the participants in the study were tested as well. ToM was tested through the Strange Stories (Happé, 1994) test in its Italian version (Mazzola & Camaioni, 2002) along with the Short version of Baron-Cohen's (2001) Autism Quotient test (Allison et al., 2012), in its Italian translation (Ruta et al., 2012). The Autism Quotient test has been used to measure autistic traits in the general, non-clinical population (see Ruzich et al., 2015, for a review), and it can therefore fit in the present study as a self-report measure along with the more objective measure provided by the Strange Stories score (which is also similar in structure to the IRs task).

Outcomes

A first exploratory analysis of the data provided interesting results, as shown in Figure 6.

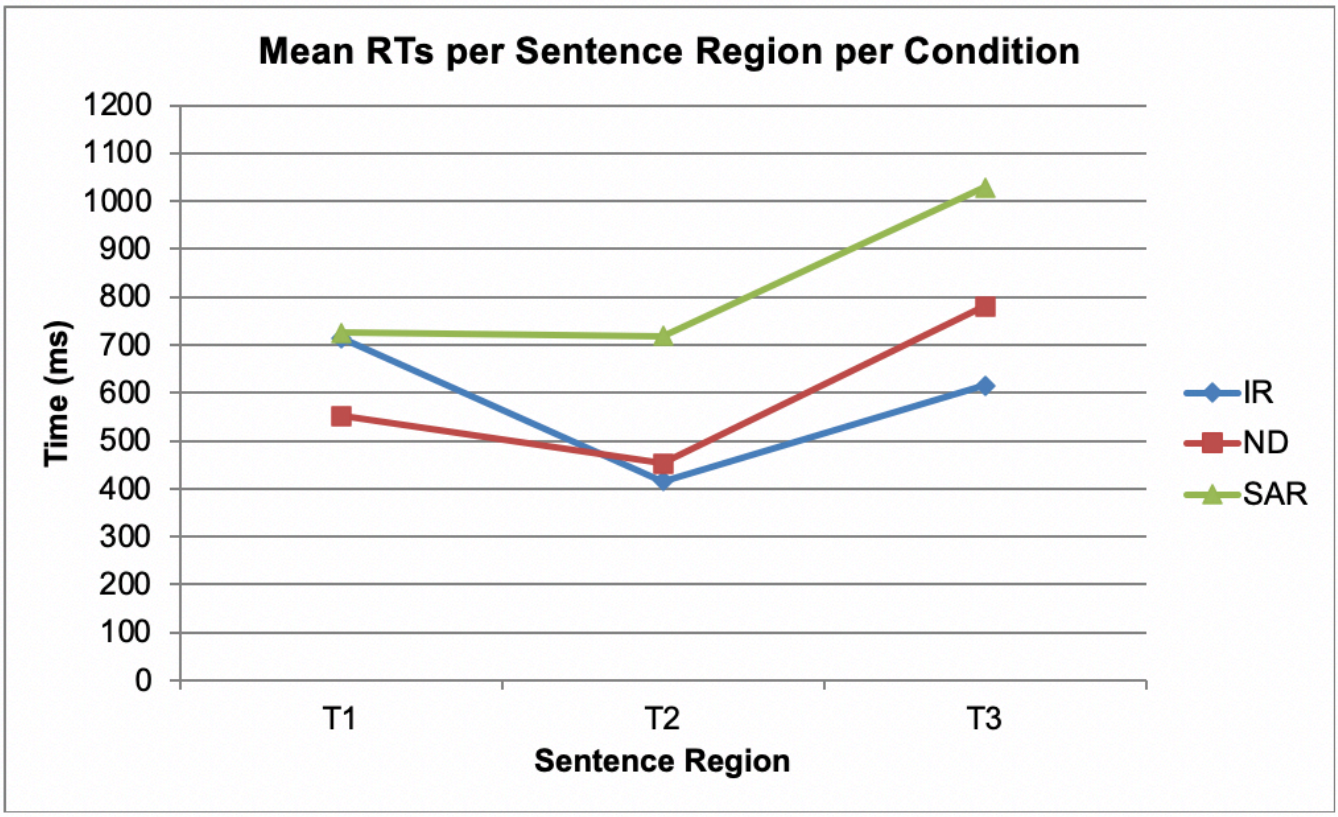

Figure 6. Mean reaction times (RTs) in milliseconds to the target sentence per condition and sentence region (T1, T2, and T3) in the IRs task.

Two types of analyses will be carried out through LMMs.
With regard to the primary object of our analysis and RQs, i.e., the interpretation strategy for conventionalized IRs, the first LMM will have participants' RTs to the target sentence in the IRs task as the outcome variable (both per region and as a whole), Condition (IR, Non-Directive, and Sarcastic) in its fixed effects structure, and random intercepts for subjects in its random structure.
This first analysis will aim at (i) studying whether there is a difference between conventionalized IRs and their non-directive counterparts, as signaled by faster overall reading times (ii) investigating whether such a difference is visible at the first- and second-region level and/or at the last-region level. Faster reading times, especially if observed at the first-region level, will be interpreted as potential evidence in favor of a more lexically-based interpretation strategy for conventionalized IRs.

A second LMM will be computed for the analysis of predictors, including participants' AQ scores and Strange Stories scores in the fixed effects structure. This will be done in order to investigate whether individual differences in ToM might have on influence on participants' behavior in the main task.

Future directions

As getting back to schools and kindergartens as researchers is highly improbable, Experiment 3 will likely target neurotypical adults again, in an attempt to dig deeper into the relationship between IRs and ToM.

It will be meant as a follow up to the actual Experiment 2 and it will add highly IRs to the picture (since Experiment 2 only focused on CIRs), investigating whether higher levels of indirectness do require higher inferencing abilities and – possibly – interpretive strategies relying on ToM.

The experimental design and methods for Experiment 3 will be defined in the following months and will strongly depend on covid-19's epidemiological curve and the government's subsequent countermeasures and restrictions. Ideally, Experiment 3 could make use of eye-tracking techniques – else, it will be programmed and administered through a web interface, much like Experiment 2.

References

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief "Red Flags" for autism screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist for Autism in toddlers in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(2), 202-212.e7. https://doi.org/10.1016/j.jaac.2011.11.003

Andrés-Roqueta, C., & Katsos, N. (2017). The Contribution of Grammar, Vocabulary and Theory of Mind in Pragmatic Language Competence in Children with Autistic Spectrum Disorders. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00996

Andrés-Roqueta Clara, & Katsos Napoleon. (2020). A Distinction Between Linguistic and Social Pragmatics Helps the Precise Characterization of Pragmatic Challenges in Children With Autism Spectrum Disorders and Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1494–1508. https://doi.org/10.1044/2020_JSLHR-19-00263

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

Baron-Cohen, Simon. (2000). Theory of mind and autism: A review. *International Review of Research in Mental Retardation*, *23*, 169–184. https://doi.org/10.1016/S0074-7750(00)80010-5

Baron-Cohen, Simon, Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Bates, D., Mächler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & al. (2015). *L. Package 'lme4'. 12*(1).

Bernicot, J., Laval, V., & Chaminaud, S. (2007). Nonliteral language forms in children: In what order are they acquired in pragmatics and metapragmatics? *Journal of Pragmatics*, *39*(12), 2115–2132. https://doi.org/10.1016/j.pragma.2007.05.009

Bernicot, J., & Legros, S. (1987). Direct and indirect directives: What do young children understand? *Journal of Experimental Child Psychology*, *43*(3), 346–358. https://doi.org/10.1016/0022-0965(87)90012-9

Blum-Kulka, S. (1987). Indirectness and politeness in requests: Same or different? *Journal of Pragmatics*, *11*(2), 131–146. https://doi.org/10.1016/0378-2166(87)90192-5

Bosco, F. M., Tirassa, M., & Gabbatore, I. (2018). Why Pragmatics and Theory of Mind Do Not (Completely) Overlap. *Frontiers in Psychology*, *9*(1453). https://doi.org/10.3389/fpsyg.2018.01453

Bucciarelli, M., Colle, L., & Bara, B. G. (2003). How children comprehend speech acts and communicative gestures. *Journal of Pragmatics*, *35*(2), 207–241. https://doi.org/10.1016/S0378-2166(02)00099-1

Carrell, P. L. (1981). Children's understanding of indirect requests: Comparing child and adult comprehension*. *Journal of Child Language*, *8*(2), 329–345. https://doi.org/10.1017/S0305000900003226

Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive Psychology*, *11*(4), 430–477. https://doi.org/10.1016/0010-0285(79)90020-3

Clark, H. H., & Lucy, P. (1975). Understanding what is meant from what is said: A study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior*, *14*(1), 56–72. https://doi.org/10.1016/S0022-5371(75)80006-5

Deliens, G., Papastamou, F., Ruytenbeek, N., Geelhand, P., & Kissine, M. (2018). Selective Pragmatic Impairment in Autism Spectrum Disorder: Indirect Requests Versus Irony. *Journal of Autism and Developmental Disorders*, *48*(9), 2938–2952. https://doi.org/10.1007/s10803-018-3561-6

Domaneschi, F., & Bambini, V. (2020). Pragmatic Competence. In E. Fridland & C. Pavese (A c. Di), *The Routledge Handbook of Philosophy of Skill and Expertise: Vol. Chap. 33*. Routledge.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368. https://doi.org/10.1016/0749-596X(86)90006-9

Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 555–568. https://doi.org/10.1037/0278-7393.16.4.555

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of*

Memory and Language, *30*(6), 725–745. https://doi.org/10.1016/0749-596X(91)90034-H
Gibbs, R. W. (1981). Your wish is my command: Convention and context in interpreting indirect requests. *Journal of Verbal Learning and Verbal Behavior*, *20*(4), 431–444. https://doi.org/10.1016/S0022-5371(81)90542-9
Gibbs, R. W. (1983). Do people always process the literal meanings of indirect requests? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(3), 524–533. https://doi.org/10.1037/0278-7393.9.3.524
Gibbs, R. W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, *34*(4), 457–486. https://doi.org/10.1016/S0378-2166(01)00046-7
Groefsema, M. (1992). 'Can you pass the salt?': A short-circuited implicature? *Lingua*, *87*(1–2), 103–135. https://doi.org/10.1016/0024-3841(92)90028-H
Groefsema, M. (1995). Can, may, must and should: A Relevance theoretic account1. *Journal of Linguistics*, *31*(1), 53–79. https://doi.org/10.1017/S0022226700000566
Happé, F. G. (1994). *An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults*. *24*, 129–154. https://doi.org/10.1007/ BF02172093
Hochstein, L., Bale, A., & Barner, D. (2017). Scalar Implicature in Absence of Epistemic Reasoning? The Case of Autism Spectrum Disorder. *Language Learning and Development*, *14*(3), 224–240. https://doi.org/10.1080/15475441.2017.1343670
Hughes, C., & Ensor, R. (2010). *Executive Function and Theory of Mind in 2 Year Olds: A Family Affair? 28*(2), 645–668.
Hutchins, T. L., Prelock, P. A., & Bonazinga-Bouyea, L. (2014). *Technical Manual for the Theory of Mind Inventory and Theory of Mind Task Battery*. ©.
Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228–238. https://doi.org/10.1037/0096-3445.111.2.228
Kissine, M., Cano-Chervel, J., Carlier, S., Brabanter, P. D., Ducenne, L., Pairon, M.-C., Deconinck, N., Delvenne, V., & Leybaert, J. (2015). Children with Autism Understand Indirect Speech Acts: Evidence from a Semi-Structured Act-Out Task. *PLOS ONE*, *10*(11), e0142191. https://doi.org/10.1371/journal.pone.0142191
Kissine, M., De Brabanter, P., & Leybaert, J. (2012). Compliance with requests by children with autism: The impact of sentence type. *Autism: The International Journal of Research and Practice*, *16*(5), 523–531. https://doi.org/10.1177/1362361311406296
Lange, K., Kühn, S., & Filevich, E. (2015). "Just Another Tool for Online Studies" (JATOS): An Easy Solution for Setup and Management of Web Servers Supporting Online Studies. *PLoS ONE*, *10*(6), e0130834.
MacKay, G., & Shaw, A. (2004). A comparative study of figurative language in children with autistic spectrum disorders: *Child Language Teaching and Therapy*, *20*(1), 13–32. https://doi.org/10.1191/0265659004ct261oa
Marini, A., Marotta, L., Bulgheroni, S., & Fabbro, F. (2015). *Batteria per la valutazione del linguaggio in bambini dai 4 ai 12 anni (BVL_4-12)*. Giunti.
Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324.
Mazzola, V., & Camaioni, L. (2002). Strane Storie: Versione italiana a cura di Mazzola e Camaioni. *Department of Dynamic and Clinical Psychology, Università "La Sapienza", Roma*.
Morgan, J. L. (1978). Two types of convention in indirect speech acts. In P. Cole (A c. Di), *Syntax and Semantics. Pragmatics* (Vol. 9, pag. 261—280). Academic Press.
Norbury, C. F. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology*, *23*(3), 383–399. https://doi.org/10.1348/026151005X26732
Ozonoff, S., & Miller, J. N. (1996). An Exploration of Right-Hemisphere Contributions to the Pragmatic Impairments of Autism. *Brain and Language*, *52*(3), 411–434. https://doi.org/10.1006/brln.1996.0022
Paul, R., & Cohen, D. J. (1985). Comprehension of Indirect Requests in Adults with Autistic Disorders and

Mental Retardation. *Journal of Speech, Language, and Hearing Research*, *28*(4), 475–479. https://doi.org/10.1044/jshr.2804.475

Ponitz, C. E. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly*, *23*(2), 141–158. https://doi.org/10.1016/j.ecresq.2007.01.004

Reeder, K. (1980). The emergence of illocutionary skills. *Journal of Child Language*, *7*(01), 13.

Ruta, L., Mazzone, D., Mazzone, L., Wheelwright, S., & Baron-Cohen, S. (2012). The autism-spectrum quotient - Italian version: A cross- cultural confirmation of the broader autism phenotype. *Journal of Autism and Developmental Disorders*, *42*(4), 625–633. https://doi.org/10.1007/s10803-011-1290-1

Ruytenbeek, N. (2019). Indirect requests, relevance, and politeness. *Journal of Pragmatics*, *142*, 78–89. https://doi.org/10.1016/j.pragma.2019.01.007

Ruytenbeek, N., Ostashchenko, E., & Kissine, M. (2017). Indirect request processing, sentence types and illocutionary forces. *Journal of Pragmatics*, *119*, 46–62. https://doi.org/10.1016/j.pragma.2017.07.011

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: A systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism*, *6*(1), 2. https://doi.org/10.1186/2040-2392-6-2

Searle, J. R. (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge University Press.

Shatz, M. (1978). Children's comprehension of their mothers' question-directives*. *Journal of Child Language*, *5*(1), 39–46. https://doi.org/10.1017/S0305000900001926

Sperber, D., & Wilson, D. (2002). *Pragmatics, Modularity and Mind-reading*. 21.

Vega Moreno, R. E. (2017). Creativity and Convention. In *Pbns.156*. John Benjamins Publishing Company.

Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x